

UNDERWATER NAVIGATION AND MAPPING WITH AN OMNIDIRECTIONAL OPTICAL SENSOR

Josep Bosch Alay

Per citar o enllaçar aquest document:
Para citar o enlazar este documento:
Use this url to cite or link to this publication:

<http://hdl.handle.net/10803/664260>



<http://creativecommons.org/licenses/by/4.0/deed.ca>

Aquesta obra està subjecta a una llicència Creative Commons Reconeixement

Esta obra está bajo una licencia Creative Commons Reconocimiento

This work is licensed under a Creative Commons Attribution licence



Doctoral Thesis

Underwater Navigation and Mapping
with an Omnidirectional Optical Sensor

JOSEP BOSCH ALAY

2018



Doctoral Thesis

**Underwater Navigation and Mapping
with an Omnidirectional Optical Sensor**

JOSEP BOSCH ALAY

2018

Doctoral Program in Technology

Supervised by:

NUNO GRACIAS
PERE RIDAO

Thesis submitted to University of Girona in fulfillment of the requirements for the degree
of

DOCTOR OF PHILOSOPHY

CERTIFICATE OF THESIS DIRECTION

Dr. Nuno Gracias and Dr. Pere Ridao, members of the Departament d'Arquitectura i Tecnologia de Computadors of Universitat de Girona,

DECLARE:

That the work entitled *Underwater Navigation and Mapping with an Omnidirectional Optical Sensor* presented by Josep Bosch Alay to obtain the degree in Doctor of Philosophy has been developed under our supervision.

Therefore, in order to certify the aforesaid statement, we sign this document.

Girona, August 2018

Dr. Nuno Gracias

Dr. Pere Ridao

ACKNOWLEDGEMENTS

This thesis would have not been possible without the help of many people who I want to address the lines below.

Com no podria ser d'altra manera m'agradaria començar amb un agraïment immens cap als meus pares. És gràcies a ells que he pogut estudiar i amb el seu exemple m'han ensenyat uns valors inestimables: la humilitat, la honradesa i l'esforç. També vull agrair als meus germans, cunyats i la resta de la família el seu interès i la seva ajuda sempre que ho he necessitat. Gràcies especialment a la Fàtima, amb qui aquesta tesi és només un petit projecte dels que ens queden compartir d'ara endavant.

Aquesta tesi no hagués estat possible si no hagués format part d'un laboratori de recerca tan fantàstic com el CIRS. He de començar donant les gràcies als meus directors de tesi: a en Nuno per la seva dedicació sense límit i a en Pere per buscar sempre solucions als problemes més variats. Gràcies a tots els sèniors: a en David que ha donat vida a les càmeres, a en Marc pels seus consells pragmàtics i les estones corrent, a en Narcís per estar sempre disposat a donar un cop de mà, a la Tali pels seus consells lingüístics i les tertúlies futbolístiques, a l'Aggelos pel seu enginy inesgotable, a en Rafa per la seva visió sempre encertada, i als Ricards i en Josep per la seva valuosa experiència i les reunions dels divendres. Thanks as well to all the colleagues which I shared a part of this journey to the PhD: Arnau, Patrick, Juan David (al final todo salió bien), Albert i Guillem, amb qui ja fa molts anys que compartim trajecte, Klemen, whose help was fundamental, Èric, Eduard, Dina, Khadija. També gràcies a en Carles i en Lluís, per dedicar les hores que facin falta per tenir els robots apunt, i a les Vicogirls Joseta, Mireia, Annes i Bego per estar sempre disposades a ajudar.

I would like also to thank Yvan Petillot and the members of the Ocean System lab, who introduced me to the underwater robotics field before starting my PhD in Girona.

Lastly, I would like to thank the anonymous reviewers that contributed to improve this thesis with their invaluable comments. Also the many reviewers that contributed to the different publications derived from this work.

This thesis was supported by "la Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya" and the National projects OMNIUS, ARCHROV and UDRONE, as well as the European projects MORPH and SUNRISE.

LIST OF PUBLICATIONS

Publications in the compendium

The presented thesis is a compendium of the following research articles:

- **J. Bosch**, N. Gracias, P. Ridaó, and D. Ribas. “Omnidirectional Underwater Camera Design and Calibration”. In: *Sensors* 15.3 (2015), pp. 6033–6065. ISSN: 1424-8220. DOI: 10.3390/s150306033
Quality index: [JCR2016 Instruments & Inastrumentation IF 2.677, Q1 (10/58)].
- **J. Bosch**, N. Gracias, P. Ridaó, K. Istenič, and D. Ribas. “Close-Range Tracking of Underwater Vehicles Using Light Beacons”. In: *Sensors* 16.4 (2016). ISSN: 1424-8220. DOI: 10.3390/s16040429
Quality index: [JCR2016 Instruments & Inastrumentation IF 2.677, Q1 (10/58)].
- **J. Bosch**, K. Istenič, N. Gracias, R. Garcia, and P. Ridaó. “Omnidirectional Multi-Camera Video Stitching using Depth Maps”
Submitted to IEEE Journal of Oceanic Engineering on June 2018
Quality index: [JCR2016 Ocean Engineering IF 2.297, Q1 (2/14)].

Publications derived from this thesis

The work developed in this thesis also led to the following publications:

- N. Palomeras, G. Vallicrosa, A. Mallios, **J. Bosch**, E. Vidal, N. Hurtos, M. Carreras, and P. Ridaó. “AUV homing and docking for remote operations”. In: *Ocean Engineering* 154 (2018), pp. 106–120. ISSN: 0029-8018. DOI: <https://doi.org/10.1016/j.oceaneng.2018.01.114>
- **J. Bosch**, N. Gracias, P. Ridaó, D. Ribas, K. Istenič, R. Garcia, and I. R. Rossi. “Immersive Touring for Marine Archaeology. Application of a New Compact Omnidirectional Camera to Mapping the Gnalič shipwreck with an AUV”. in: *ROBOT 2017: Third Iberian Robotics Conference*. Ed. by A. Ollero, A. Sanfeliu, L. Montano, N. Lau, and C. Cardeira. Cham: Springer International Publishing, 2018, pp. 183–195
- N. Hurtós, A. Mallios, N. Palomeras, **J. Bosch**, G. Vallicrosa, E. Vidal, D. Ribas, N. Gracias, M. Carreras, and P. Ridaó. “LOON-DOCK: AUV homing and docking

for high-bandwidth data transmission”. In: *OCEANS 2017 - Aberdeen*. June 2017, pp. 1–7. DOI: 10.1109/OCEANSE.2017.8084806

- G. Vallicrosa, **J. Bosch**, N. Palomeras, P. Ridaó, M. Carreras, and N. Gracias. “Autonomous homing and docking for AUVs using Range-Only Localization and Light Beacons”. In: 49.23 (2016). 10th IFAC Conference on Control Applications in Marine Systems CAMS 2016, pp. 54–60. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2016.10.321>
- G. Nuno, G. Rafel, R. Campos, R. Prados, **J. Bosch**, A. Elibol, T. Nicosevici, L. Neumann, and J. Quintana. “Omnidirectional Underwater Surveying and Telepresence”. In: *Proceedings of the 7th International Workshop on Marine Technology (MARTECH)*. 2016
- **J. Bosch**, P. Ridaó, D. Ribas, and N. Gracias. “Creating 360 degrees underwater virtual tours using an omnidirectional camera integrated in an AUV”. in: *OCEANS 2015 - Genova*. May 2015, pp. 1–7. DOI: 10.1109/OCEANS-Genova.2015.7271525
- N. Gracias, **J. Bosch**, and M. E. Karim. “Pose Estimation for Underwater Vehicles using Light Beacons”. In: 48.2 (2015). 4th IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles NGCUV 2015, pp. 70–75. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2015.06.012>

ACRONYMS

AUV Autonomous Underwater Vehicle.

CIRS Underwater Robotics and Vision Research Center.

DS Docking Station.

DVL Doppler Velocity Log.

FOV Field of View.

GPS Global Positioning System.

OMS Omnidirectional Multi-camera system.

ROV Remotely Operated Vehicle.

SfM Structure from Motion.

SLAM Simultaneous Localization and Mapping.

UdG Universitat de Girona.

VR Virtual Reality.

VT Virtual Tour.

LIST OF FIGURES

1.1	Different omnidirectional capturing systems.	9
5.1	Omnidirectional cameras designed during this thesis.	100
5.2	Location and sample image of a shipwreck virtual tour in Google Maps. . .	105
5.3	Results of the survey of the Gnalić.	105
5.4	Debug image of the optical tracking system based on light markers.	107
5.5	DS tracking performance during the docking command execution in one of the missions.	107

CONTENTS

Abstract	1
Resum	3
Resumen	5
1 Introduction	7
1.1 Motivation	8
1.2 Objectives	10
1.3 Context	11
1.4 Document Structure	12
2 Omnidirectional Underwater Camera Design and Calibration	13
3 Omnidirectional Multi-Camera Video Stitching using Depth Maps	47
4 Close-Range Tracking of Underwater Vehicles Using Light Beacons	71
5 Main Results and Discussion	99
5.1 Design of an Omnidirectional Multi-Camera System	100
5.2 Calibration of an Omnidirectional Multi-Camera System	101
5.3 Generation of panoramic imagery and video	103
5.4 Immersive Virtual Reality applications	104
5.5 Close Range Tracking of multiple vehicles	105
6 Conclusions and Future Work	109
6.1 Conclusions	110
6.2 Future work	111
Bibliography	113

ABSTRACT

Omnidirectional vision has received increasing interest during the last decade from the computer vision community. A large number of camera models have reached the market to meet the increasing demand for panoramic imagery. However, the use of omnidirectional cameras underwater is still very limited. In this thesis we propose a number of methods to create a reference resource for designing, calibrating and using underwater omnidirectional multi-camera systems.

The first problem we address is the design and calibration of omnidirectional cameras for the underwater domain. Among the different imaging system approaches to capturing omnidirectional imagery we chose the use of multi-cameras, due to the higher resolution and quality of the final images obtained. In order to assist the design and insure a proper view coverage, a field-of-view (FOV) simulator has been developed which takes into account the individual FOVs of the cameras, the position and orientation between them and the geometry and relative pose of the waterproof housing. The latter is especially relevant due to the strong image distortions caused by the refraction of the optical rays when travelling through the different media. Nonetheless, once the system is built, a very accurate calibration is required for any metrology or computer vision application. So, a full calibration method is presented for the estimation of both the intrinsic and extrinsic parameters of the cameras and the relative pose of the waterproof housing. This method is able to cope with wide-angle lenses and non-overlapping cameras simultaneously and applicable to both land or water Omnidirectional Multi-camera system (OMS).

Next, the topic of stitching strategies, to generate omnidirectional panoramas from the individual images, is studied in depth. Stitching strategies have the complex objective of joining the images in a way such that the viewer has the feeling the panoramas were captured from a single location. Conventional approaches either assume that the world is a simple sphere around the camera or use feature-based stitching techniques to align the individual images. However, this leads to artifacts and misalignments in the final panoramas due to parallax effects. This thesis presents a set of new stitching strategies, for both online and offline applications, aiming at processing the images according to available information of the multi-camera system and the environment.

Finally, we focus on potential underwater applications. We first explore the promising uses of omnidirectional cameras to create immersive virtual experiences. Then, we demonstrate the capabilities of omnidirectional cameras as complementary sensors for the navigation of underwater robots. Specifically, we present a new tracking system for autonomous underwater vehicles (AUVs) navigating in a close formation. The proposed system, which makes use of active light marker estimates the pose of a target vehicle at

short ranges, with high accuracy and execution speed.

In order to validate all presented algorithms, two custom omnidirectional cameras were built and several experiments with divers and underwater robots have been carried out to collect the necessary data.

RESUM

Durant l'última dècada s'ha despertat un interès creixent per a la visió omnidireccional en l'àmbit de la visió per computador. Podem trobar una gran varietat de càmeres omnidireccionals al mercat amb l'objectiu de cobrir l'acutal demanda d'imatges i vídeos panoràmics. Tot i això, l'ús de càmeres omnidireccionals submarines encara és molt limitat. En aquesta tesi doctoral hem creat un recurs de referència per al disseny, calibratge i ús de càmeres omnidireccionals submarines de tipus multi-càmera.

El primer problema que tractem és el disseny i el calibratge de càmeres omnidireccionals per al medi submarí. D'entre els diferents sistemes capaços de capturar imatges omnidireccionals, hem escollit els sistemes multi-càmera per l'alta resolució i qualitat de les imatges obtingudes. Hem dissenyat un simulador de camp de visió per tal d'ajudar en el disseny de sistemes multi-càmera i assegurar que són capaços de cobrir el camp de visió desitjat. Aquest simulador té en compte les característiques individuals de les càmeres, la posició i orientació entre elles i la posició relativa amb la carcassa submarina. És especialment rellevant tenir en compte aquesta última degut als importants canvis de direcció dels rajos de llum quan travessen diferents medis. Una vegada el sistema ha estat dissenyat i construït, és necessari un calibratge molt precís per utilitzar-lo en qualsevol aplicació de metrologia o de visió per computador. Per a resoldre aquest problema, presentem un sistema complet de calibratge que estima els paràmetres intrínsecs i extrínsecs de les càmeres, així com la posició relativa de la carcassa. Aquest mètode és capaç de tractar amb lents de tipus ull-de-peix o gran angular així com sistemes de càmeres sense solapament i és aplicable tant a càmeres submarines com terrestres.

Després ens centrem en l'estudi en profunditat de les estratègies de combinació d'imatges per tal de generar imatges panoràmiques partint de múltiples imatges capturades per diferents càmeres. Aquestes estratègies tenen com a objectiu la complexa tasca de crear una imatge panoràmica tal que l'observador tingui la sensació que la imatge ha sigut creada des d'un únic punt de vista. Les estratègies convencionals parteixen de l'assumpció que el món pot ser simplificat com una esfera centrada en el sistema de càmeres o fan servir tècniques de combinació basades en punts d'especial interès per tal d'alinejar correctament les imatges. Malauradament, aquestes estratègies resulten en panoràmiques finals amb alineaments dolents entre imatges i defectes visuals degut als efectes de parallax. Aquesta tesi presenta un conjunt d'estratègies noves per a ús en temps real o post-processat amb l'objectiu de millorar les panoràmiques finals fent ús de la informació disponible de l'entorn i del sistema multi-càmera.

Finalment, ens centrem en les aplicacions potencials de les càmeres omnidireccionals submarines. Primer, explorem els prometedors usos de les càmeres per tal de crear ex-

periències virtuals immersives. Després, demostrem les capacitats de les càmeres omnidireccionals com a sensors complementaris per a la navegació de robots. Concretament, presentem un sistema de seguiment de robots submarins autònoms (AUVs) que naveguin en formacions de distàncies curtes. El sistema proposat fa ús de marcadors de llum actius per tal d'estimar la posició relativa del vehicle amb una alta precisió i velocitat d'execució.

Per tal de validar tots els algoritmes presentats s'han dissenyat dues càmeres omnidireccionals i s'han realitzat múltiples experiments amb robots i submarinistes per a recollir les dades necessàries.

RESUMEN

Durante la última década se ha generado un interés creciente por la visión omnidireccional en el ámbito de la visión por computador. Podemos encontrar una gran variedad de cámaras omnidireccionales en el mercado para cubrir la actual demanda de imágenes y vídeos panorámicos. Aun así, el uso de cámaras omnidireccionales submarinas es todavía muy escaso. En esta tesis doctoral hemos creado un recurso de referencia para el diseño, calibración y aplicaciones de cámaras omnidireccionales submarinas de tipo multi-cámara.

El primer problema que tratamos es el diseño y calibración de cámaras omnidireccionales para el medio submarino. De entre todos los sistemas capaces de capturar imágenes omnidireccionales, escogimos los sistemas multi-cámara por la alta resolución y calidad de las imágenes obtenidas. Hemos diseñado un simulador de campo de visión con el objetivo de ayudar en el diseño de sistemas multi-cámara y asegurar que son capaces de cubrir el campo de visión deseado. Este simulador tiene en cuenta las características individuales de las cámaras, la posición y orientación entre ellas, y la posición relativa con la carcasa submarina. Es especialmente relevante tener en cuenta esta última debido a los importantes cambios de dirección que sufren los rayos de luz cuando atraviesan diferentes medios. Una vez el sistema ha sido diseñado y fabricado es necesaria una calibración muy precisa para usarlo para cualquier aplicación de metrología o visión por computador. Para resolver este problema, presentamos un sistema completo de calibración que estima los parámetros intrínsecos y extrínsecos de las cámaras, así como la posición relativa de la carcasa. Este método es capaz de tratar con lentes tipo ojo de pez y gran angular, así como sistemas sin solapamiento y es aplicable tanto para cámaras submarinas como terrestres.

Después, nos centramos en el estudio en profundidad de las estrategias de la combinación de imágenes con tal de generar imágenes panorámicas partiendo de múltiples imágenes capturadas por diferentes cámaras. Estas estrategias tienen como objetivo la compleja tarea de crear una imagen panorámica tal que el observador tenga la sensación de que la imagen fue capturada desde un único punto de vista. Las estrategias convencionales parten de la asunción que el mundo puede ser simplificado como una esfera centrada en el centro del sistema de cámaras o usan técnicas de combinación basadas en puntos de interés por tal de alinear correctamente las imágenes. Desafortunadamente estas estrategias resultan en panoramas finales con alineamientos malos entre imágenes y defectos visuales debido al parallax. Esta tesis presenta un conjunto de estrategias nuevas para su uso en tiempo real o durante un post-procesado con el objetivo de mejorar los panoramas obtenidos haciendo uso de la información disponible del entorno y del sistema multi-cámara.

Por último, nos centramos en las aplicaciones potenciales de las cámaras omnidireccionales submarinas. En primer lugar, exploramos los prometedores usos de las cámaras con tal de crear experiencias virtuales inmersivas. Después, demostramos las capacidades de las cámaras omnidireccionales como sensores complementarios para la navegación de robots. Concretamente, presentamos un sistema de seguimiento de robots submarinos autónomos (AUVs) que naveguen en formaciones de distancias cortas. El sistema propuesto hace uso de marcadores luminosos activos para estimar la posición relativa del robot con una alta precisión y velocidad de ejecución.

Con tal de validar todos los algoritmos presentados se han diseñado dos cámaras omnidireccionales y se han realizado múltiples experimentos con robots y submarinistas para recoger los datos necesarios.

1

INTRODUCTION

IN this chapter we present the main problems that have motivated this thesis: the use of omnidirectional camera systems for underwater navigation and mapping. The motivations behind this problem are introduced in Section 1.1, relating them to the requirements of real-world applications and the limitations of current approaches. Next, we state the objectives of the thesis in Section 1.2 and we briefly describe, in Section 1.3, the context in which this work has been carried out. Finally, the organization of the thesis document is presented in Section 1.4.

1.1 Motivation

Omnidirectional imagery has been a field of study for decades in the computer vision community [11, 12]. The main advantage on using omnidirectional cameras is their Field of View (FOV), which covers 360° horizontally and allows capturing a whole scene in a single action.

Omnidirectional cameras are used in many robotics and computer vision fields as diverse as augmented reality[13], object recognition and tracking [14, 15], obstacle avoidance [16], visual surveillance [17], navigation [18, 19], motion estimation [20] and Simultaneous Localization and Mapping (SLAM) [21, 22]. Moreover, they are becoming extraordinarily popular among the general public, to record personal video. The main appeal of the cameras comes from their ability to easily produce panoramic images and videos of a scene. The viewer is able to experience a unique immersive feeling when panoramic media are displayed in a spherical viewer or in a Virtual Reality (VR) headset. The rise in the demand for omnidirectional imagery has been supplied with the launch of many different cameras in the market over the last decade.

Multiple systems have been designed in order to capture omnidirectional images. One of the simplest and most widely used method consists in a conventional camera mounted in a rotating base which is aligned with the optical center of the camera (Fig. 1.1a). Through the rotation of the base it is possible to capture images of the whole scene and stitch them together in a later post-processing stage using one of the many commercial software available [23, 24, 25, 26]. This system is still very popular among professional photographers as it is possible to obtain very high quality panoramas by simply manually rotating the camera mounted on a tripod. However, this method is only applicable for static scenes as the images are not captured at the same time instant and, consequently, not suitable for recording omnidirectional video.

Another system that became very popular in the 90's and beginnings of 00's are the catadioptric systems, which consist of the combination of a conventional or wide-angle camera with parabolic or hyperbolic mirrors (Fig. 1.1b) [27, 28, 29, 30]. The setup is mounted with the camera pointing to the mirror, which allows a considerable increase in the FOV. In order to use the collected images for any computer vision application or to re-project them into another panoramic projection, the model of the camera along with the refractions caused by the mirror need to be estimated accurately [31, 32, 33]. The main drawback of the system lies in the angular resolution, which is not constant over the whole image.

More recently, advances in the resolution of optical sensors and optics manufacturing techniques have popularized wide-angle and fisheye cameras, which use special optics with extra-wide FOV able to open up to slightly more than 180° (Fig. 1.1c) [31, 34, 35]. As for the catadioptric systems, the distortions introduced by this kind of optics need to be modeled accurately before the images can be used. Many authors have developed models able to handle distortions of both catadioptric systems and fisheye-cameras [30, 36, 37, 38]. However, similarly to the catadioptric systems, their main disadvantage is that the angular resolution is not constant over the images.

Finally, since the 00's Omnidirectional Multi-camera Systems (OMS) are becoming more popular. They consist of clusters of cameras rigidly attached and positioned strategically to cover the whole scene (Fig. 1.1d). Using these systems it is possible to obtain omnidirectional images and videos with superlative resolution, when compared to the pre-

vious systems. Moreover, the panoramic images obtained have acceptably high uniformity in the angular resolution. With a calibrated OMS, the individual images collected can be used separately for vision and metrology applications, or they can be combined to create omnidirectional panoramas. However, for the latter case, they need to be stitched carefully to obtain a panorama without visible artifacts.



(a) Canon camera mounted in a RRS rotating head for capturing panoramas.



(b) VSN Mobil V.360 is a catadioptric system for capturing real time 360° panoramas.



(c) Kodak PIXPRO SP360 uses a fisheye optics to capture 360° panoramic images.



(d) Insta 360 Pro multi-camera system.

Figure 1.1: Different omnidirectional capturing systems.

The calibration of a OMS has two differentiated steps: the individual calibration of each camera (to obtain the intrinsic parameters) and the estimation of the geometrical relationship between the cameras (extrinsic parameters). While the first is a very well-studied problem for the computer vision community [39], the latter, although well-known, is still receiving research interest for the specific case of multi-camera systems. These systems usually have small or nonexistent overlap between neighboring cameras, which creates an important challenge to overcome [40, 41, 42].

The problem of combining the individual images to obtain panoramic images has been an intensely-studied research topic in recent years. Many authors have studied techniques to stitch individual images seamlessly [43, 44, 45]. However, research is still being carried out to develop algorithms that adequately address the issues related to parallax [46, 47, 48].

Although a large range of omnidirectional cameras are available on the market, their use in the underwater domain is still very limited due to the tough conditions of the medium. Underwater cameras suffer from problems related to color and contrast degradation, due to light absorption and scattering of the light in water [49, 50, 51]. Moreover, cameras need to be encapsulated inside a waterproof housing, which induce changes in the direction of the rays of light travelling through it. The distortion introduced by these direction changes needs to be modelled properly before the images can be used for any application. During the last decade, several authors have studied the distortions introduced by flat [52, 53, 54] and dome ports [55, 56].

Very few omnidirectional underwater cameras are available, either for commercial use or for research, and there are scarce applications taking advantage of their potential [57, 58]. Nonetheless there are many potential uses of omnidirectional cameras in the underwater domain. Panoramic content can be useful for pilots of Remotely Operated Vehicles (ROVs) to mitigate spatial awareness problems resulting from the lack of landmarks and the monotony of underwater scenes [5, 59]. Similarly, it could also help marine scientists on live inspection missions or during mission replays. The large FOV of the cameras can be very useful for object-tracking applications such as fish identification and tracking. Furthermore, panoramic content is a powerful tool for science dissemination as it attracts the attention of the public, especially among younger generations [60]. Moreover, it can also be used as a sensor for improving the localization capabilities of ROVs and Autonomous Underwater Vehicles (AUVs). Navigation and localization is a key aspect for the development of reliable underwater robots. Since absolute positioning systems such as Global Positioning System (GPS) are not available underwater, navigation relies on other localization systems such as inertial and Doppler Velocity Log (DVL) sensors. Optical cameras can be especially useful when used on confined environments or when navigating towards a known structure such as a docking station. They can be used for locating the robot using SLAM approaches or by comparison with a map known *a priori*. Using an omnidirectional camera can improve the performance of these localization approaches due to their wide FOV.

1.2 Objectives

Now that our motivations have been described, we can state the goal of this thesis:

To develop a complete reference resource for designing, calibrating and using an omnidirectional underwater multi-camera system and to prove its capabilities with practical uses for the marine robotics field.

This general goal can be broken into the following more specific objectives:

- **Design of an underwater OMS:** To propose a general guideline for designing a multi-camera system for the underwater medium, which takes into account the most common issues involved in designing and using an underwater camera. We aim to illustrate this objective with real custom-designed cameras.
- **Calibration of an OMS:** To propose a general calibration methodology for any multi-camera system, both for air and underwater cameras. We want to design a methodology that does not use any specialized calibration tool or pattern. Instead we prefer a methodology that could be used for any OMS user with minimal knowledge on computer vision.

- **Generation of panoramic imagery and video from an OMS:** To create the necessary tools to convert the individual images captured by the cameras into a single omnidirectional panorama or video. The proposed tools should be compatible with any OMS and easily configurable for users without deep knowledge on computer vision.
- **Exploration of the immersive Virtual Reality applications in the underwater context:** To explore all the potential uses of omnidirectional cameras in the underwater environment, especially the capability to offer unique immersive experiences. This includes uses for divers and underwater robots as well as potential applications for the scientific community.
- **Use of an OMS for aiding underwater navigation on an AUV:** To test the capabilities of omnidirectional cameras as underwater sensors for improving the navigation systems of robots. Specifically we want to evaluate them as sensors to detect close-range robots in cooperative navigation between AUVs.

1.3 Context

The work presented in this thesis has been supported by the FI 2015 grant from the *Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya* and has been developed at the Underwater Robotics and Vision Research Center (CIRS) research group of the Universitat de Girona (UdG), which is part of the VICOROB research institute. The group started researching in underwater vision and robotics in 1992 and it is currently formed by pre-doctoral researchers, engineers, technicians, postdoctoral fellows and permanent staff. The group is a leading team in the research and development of AUVs for accurate seafloor mapping and light intervention. It has participated in several European-funded and National-funded projects (of both basic and applied research) and it has also been involved in technology transfer projects and contracts with companies and institutions worldwide.

Among the large amount of articles published by the group, the most significant are regarding SLAM [61, 62, 63, 64], autonomous intervention [65, 66, 67, 68, 69] and path planning [70, 71, 72].

The group has developed several AUV prototypes and has currently two fully operative robots: Sparus II [73], a torpedo-shaped vehicle winner of multi-domain robotics competition Eurathlon 2014, 2015 and 2017, and Girona500 [74] a reconfigurable AUV for both survey and intervention. Both vehicles have been used during this thesis for the data collection and validation of the algorithms presented.

It is also especially relevant for this thesis the works of the Underwater Vision Lab, which is a subgroup of VICOROB, and covers topics such as 3D reconstruction [75], image dehazing [76], mosaic blending [77], multi-vehicle mapping [78] or fish detection [79].

The work presented in this thesis was developed in the scope of, and contributed to, the following National and European projects:

- FP7 EU Project MORPH: Marine robotic systems of self-organizing, logically linked physical nodes (FP7-ICT-2011-7-288704), funded by the European Commission.

- MINECO Project OMNIUS: Lightweight robot for OMNIdirectional Underwater Surveying and telepresence (CTM2013-46718-R), funded by the Spanish Ministry of Economy, Industry and Competitiveness.
- MINECO Project ARCHROV: marine ARChaeology through HROV/AUV cooperation (DPI2014-57746-C3-3-R), funded by the Spanish Ministry of Economy, Industry and Competitiveness.
- FP7 EU Project SUNRISE: Sensing, monitoring and actuating on the Underwater world through a federated Research InfraStructure Extending the Future Internet (FP7-ICT-611449), funded by the European Commission.
- MINECO Project UDRONE: Robot submarino inteligente para la exploración omnidireccional e inmersiva del bentos (CTM2017-83075-R), funded by the Spanish Ministry of Economy, Industry and Competitiveness.

1.4 Document Structure

The rest of the thesis is structured as follows:

Chapter 2: Omnidirectional Underwater Camera Design and Calibration addresses the problem of designing and calibrating an omnidirectional underwater camera and generating omnidirectional panoramas.

Chapter 3: Omnidirectional Multi-Camera Video Stitching using Depth Maps deals with the issue of mitigating the effect of parallax when generating omnidirectional images. We propose new strategies that take into account the structure of the scene for reducing the effects of parallax.

Chapter 4: Close-Range Tracking of Underwater Vehicles Using Light Beacons presents a practical application of underwater cameras. Specifically, in this chapter an omnidirectional camera is mounted as a sensor in a AUV and used to track other vehicles in a cooperative navigation mission.

Chapter 5: Main Results and Discussion lays out the results obtained during this thesis. The results are outlined according to the objectives already presented.

Chapter 6. Conclusions and Future Work summarizes the conclusions based on the contributions of this thesis. Based on these conclusions, proposed future work is also presented.

2

OMNIDIRECTIONAL UNDERWATER CAMERA DESIGN AND CALIBRATION

IN this chapter, we describe the design and calibration of an underwater omnidirectional camera. First we focus on the design of an OMS and we illustrate it with a underwater OMS based on an adaptation of a Point Grey Ladybug3 multi-camera. Afterwards, we focus on covering all calibration process for such a system. The process is validated with the calibration of the proposed system. Finally the problem of generating panoramic images from multiple individual cameras is addressed in order to appreciate the results obtained. All proposed work was described in detail and published in the following journal paper:

Title: Omnidirectional Underwater Camera Design and Calibration
Authors: J. Bosch , N. Gracias, P. Ridaó, and D. Ribas
Journal: Sensors
Volume: 15, Number: 3, Pages: 6033–6065, Published: 2015
DOI: 10.3390/s150306033
Quality index: JCR2016 Instruments & Instrumentation IF 2.677, Q1 (10/58)

Sensors **2015**, *15*, 6033–6065; doi:10.3390/s150306033

OPEN ACCESS
sensors

ISSN 1424-8220

www.mdpi.com/journal/sensors

Article

Omnidirectional Underwater Camera Design and Calibration

Josep Bosch *, Nuno Gracias, Pere Ridao and David Ribas

Computer Vision and Robotics Group, Centre d'Investigació en Robòtica Submarina,
Parc Científic i Tecnològic, Universitat de Girona, 17003 Girona, Spain;
E-Mails: ngracias@eia.udg.edu (N.G.); pere@eia.udg.edu (P.R.); dribas@udg.edu (D.R.)

* Author to whom correspondence should be addressed; E-Mail: j.bosch@udg.edu;
Tel.: +34-972-419-651.

Academic Editor: Vittorio M.N. Passaro

Received: 3 November 2014 / Accepted: 26 February 2015 / Published: 12 March 2015

Abstract: This paper presents the development of an underwater omnidirectional multi-camera system (OMS) based on a commercially available six-camera system, originally designed for land applications. A full calibration method is presented for the estimation of both the intrinsic and extrinsic parameters, which is able to cope with wide-angle lenses and non-overlapping cameras simultaneously. This method is valid for any OMS in both land or water applications. For underwater use, a customized housing is required, which often leads to strong image distortion due to refraction among the different media. This phenomena makes the basic pinhole camera model invalid for underwater cameras, especially when using wide-angle lenses, and requires the explicit modeling of the individual optical rays. To address this problem, a ray tracing approach has been adopted to create a field-of-view (FOV) simulator for underwater cameras. The simulator allows for the testing of different housing geometries and optics for the cameras to ensure a complete hemisphere coverage in underwater operation. This paper describes the design and testing of a compact custom housing for a commercial off-the-shelf OMS camera (Ladybug 3) and presents the first results of its use. A proposed three-stage calibration process allows for the estimation of all of the relevant camera parameters. Experimental results are presented, which illustrate the performance of the calibration method and validate the approach.

Keywords: omnidirectional; underwater; camera; calibration; housing; OMS; image stitching; image blending; panorama construction

1. Introduction

In the last few years, omnidirectional cameras have received increasing interest from the computer vision community in tasks such as augmented reality, visual surveillance, motion estimation and simultaneous localization and mapping (SLAM). The wide field of view (FOV) provides a comprehensive view of a scene. However, the use of these cameras underwater is still at a very early technological stage.

The use of omnidirectional cameras in underwater environments opens the door to several new technological applications in fields as diverse as underwater robotics, marine science, oil and gas industries, underwater archeology and science outreach. As an example, underwater panoramic images can be used to create virtual reality tours of zones of special interest, like shipwrecks or underwater nature reserves. In the first case, it would be an attractive and innovative tool to bring archeology closer to the general public, and in the latter, it can be an enticing way to promote awareness for the preservation of a specific region.

For underwater robotics, omnidirectional cameras are expected to have a large impact in both remotely-operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) [1]. It will allow ROVs to be piloted directly through the images captured by the omnidirectional cameras through virtual reality headsets. This immersive experience will extend the pilots spatial awareness and reduce the usual orientation problems during missions. For AUVs, the wide FOV of the cameras is very convenient for visual SLAM and mapping tasks, especially in confined or cluttered environments. In particular, the omnidirectional camera presented in this paper has been developed in the framework of the MORPH (Marine robotic systems of self-organizing, logically linked physical nodes) EU-FP7 project [2]. One of the tasks of the project involves cooperative navigation between AUVs. For this purpose, the camera will be integrated into the Girona500 AUV (Figure 1a) [3], enabling the tracking and pose estimation of other robots navigating close to it.

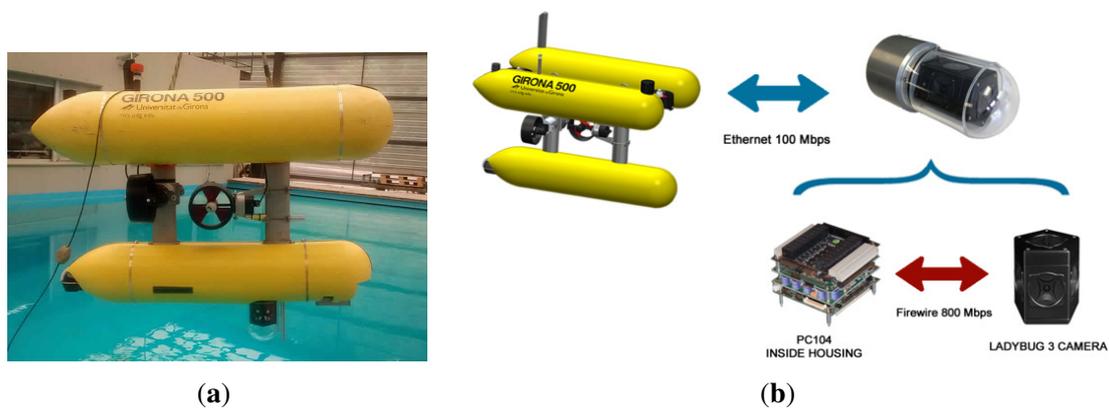


Figure 1. Integration of the omnidirectional camera in the Girona500 AUV. (a) The omnidirectional camera integrated with the Girona500 AUV in the CIRS (Underwater Robotics Research Centre) water tank; (b) scheme of the communications between the Girona500 AUV and the omnidirectional camera.

This paper presents an underwater omnidirectional multi-camera system (OMS) based on a Point Grey's Ladybug 3 [4] camera. The Ladybug 3 comprises six individual cameras and is designed for land applications. In order to be used underwater, structural changes are required in the camera itself, as well as the manufacturing of a custom waterproof housing. As a consequence, the factory-provided calibration is not valid, and a new custom calibration procedure for underwater OMS needs to be developed.

Calibration is a mandatory step for most camera applications. It enables the use of many computer vision tools for tasks, such as 3D reconstruction or motion estimation. Calibration for multi-camera systems typically covers two different sets of parameters: intrinsic parameters, concerning the image formation geometry for each individual camera, and extrinsic parameters, which describe the relative positions and orientations between cameras. In omnidirectional multi-camera systems, the calibration of the extrinsic parameters is an important challenge, due to the usual small overlap between neighboring cameras.

In this paper, the complete calibration is done in three different stages. The first consists of the estimation of the intrinsic parameters, which is done separately for each single camera in air and without the waterproof housing. The second stage consists of the estimation of the extrinsic parameters, also done in the same conditions as the first step. Finally, the last stage takes place underwater and estimates the camera pose with respect to the waterproof housing.

This calibration procedure is done in three stages rather than in a single combined step. The reason behind this is two-fold. Firstly, it allows for a smaller number of parameters to be estimated in each individual step, thus avoiding unwanted correlations among the parameter estimates. Secondly, it allows the use of image sets captured in air, for the estimation of the parameters that are not related with the underwater housing. This way, the intrinsic and extrinsic parameters are not affected by disturbances, such as the non-modeled geometric inaccuracies of the waterproof housing. Furthermore, it is significantly easier to find feature points in images captured in air than in water, due to light absorption and varying illumination conditions. The use of a larger number of well-spread feature points contributes to a higher calibration accuracy.

Different target applications of the OMS impose different requirements on the calibration accuracy. When used for science outreach and visualization purposes, the effort will be put in creating panoramas with the minimal amount of distracting visual artifacts. By contrast, when the OMS is used for object recognition or tracking, the effort will concentrate on achieving the best model possible for the image formation geometry.

1.1. Related Work

Camera calibration has been a topic of research since cameras have started being used for metrology [5,6]. There is a vast body of literature in the photogrammetry field [7–9] that focuses on modeling all of the relevant aspects of the image formation, towards obtaining high accuracy models. However, such a vast literature is almost completely devoted to aerial and land applications. Comparably fewer references can be found for underwater metrology, especially for the case of non-conventional camera systems, like OMS. Of particular relevance to this paper is the existing work on the modeling

of wide-angle and fish-eye lenses. Kannala and Brandt [10] presented a generic geometric model and a calibration method based on a planar calibration pattern suitable for both fish-eye lenses and conventional cameras. Scaramuzza *et al.* [11] propose another calibration method using a planar grid assuming that the image projection function can be described by a Taylor series expansion. Mei and Rives [12] propose a model based on the exact theoretical projection function and with the addition of parameters to model real-world errors.

Most omnidirectional cameras can be divided into two main groups [13]: central omnidirectional cameras, which strictly satisfy the single-viewpoint property, and non-central omnidirectional cameras. The first group is formed by all catadioptric systems, combinations of wide camera lenses and parabolic or hyperbolic mirrors. The later group, known as omnidirectional multi-camera systems (OMS) or polycameras, is formed by cameras composed of a cluster of individual cameras pointing to different directions in order to cover the maximum FOV possible. The first group of omnidirectional cameras is usually less expensive than an OMS, but their resolution tends to be lower and nonuniform. Typically, the resolution of catadioptric cameras is maximum in the center of the image and decreases significantly when approaching the corners. Furthermore, they are not as compact as an OMS, and its encapsulation to be used underwater is not trivial.

A few authors have analyzed the calibration of OMS with minimal overlapping between cameras. Kumar *et al.* [14] propose a calibration methodology for non-overlapping cameras using a mirror and a standard checker board. Ikeda *et al.* [15] propose a calibration method based on a calibration pattern and the use of a laser measurement system. Li *et al.* [16] presented a MATLAB toolbox for OMS calibration. This toolbox estimates both intrinsic and extrinsic parameters of the omnidirectional camera through the use of a custom descriptor-based calibration pattern rather than a standard pattern. The authors claim that the use of the custom pattern enables many more features of varying scales to be easily detected.

Regarding underwater cameras, very few works on omnidirectional underwater cameras can be found. Yamashita [17] proposed an omnidirectional underwater stereo sensor based on individual conventional video cameras and hyperboloid mirrors inside an acrylic cylindrical waterproof case. As mentioned before, this solution has not been adopted, as the use of an OMS allows one to capture panoramas in higher resolution and is more uniformly distributed. al waterproof case. As mentioned before, this solution has not been adopted, as the use of an OMS allows one to capture panoramas in higher resolution and is more uniformly distributed.

However, many authors have worked with underwater cameras and faced similar challenges due to the image distortion caused by the changes in the refractive indexes when the rays of light go through the waterproof housing. Kunz and Singh [18] examined the challenges that pressure housing interfaces introduce, focusing on hemispherical interfaces. They propose a camera calibration in two steps: a first traditional in-air calibration and the second step of adding the terms accounting for refraction. For hemispherical interfaces, there are three degrees of freedom due to the camera position inside the housing, apart from its radius and thickness, which can be measured physically. Through an optimization procedure and with the use of any standard calibration pattern, the three degrees of freedom are easily determinable. Sedlazeck and Koch [19] used mainly the same model presented by Kunz and Singh, but applied to flat ports. Similarly to [18], only three degrees of freedom are considered, corresponding to the plane orientation, plus a parameter d corresponding to the distance between the camera and the

interface. Both works use the same approach for the underwater housing calibration as the one used in this paper. However, the geometry of the housing developed in this paper and, therefore, its ray-tracing study are significantly more complex than the ones studied previously.

1.2. Contributions

The main contributions of this paper are:

1. A new calibration method applicable to multiple non-overlapping camera systems in both out of the water and underwater systems. This method has the following advantages:
 - (a) It overcomes the need of the cameras to see a calibration pattern entirely to compute accurately its intrinsic parameters.
 - (b) It is not required during the extrinsic parameters calibration that a calibration pattern must be seen entirely by different cameras at the same time. This allows one to calibrate cameras with non-overlapping FOV.
 - (c) It can handle the distortions introduced by a waterproof housing thanks to a ray tracing study.
2. The proposal and experimental validation of a compact underwater housing that does not block or limit the full FOV of the omnidirectional camera.
3. The development of an open source Linux driver [20] and a robot operating system (ROS) [21] package [22] for a Ladybug 3 camera or similar. This kind of driver was only available for Windows OS and was under the copyright of Point Grey Research Inc.

The rest of the paper is organized as follows. Section 2 presents the design of the camera and housing. Section 3 describes the single camera calibration problem. Section 4 presents the procedure used to calibrate the extrinsic parameters of all of the cameras. Section 5 introduces the challenges of the use of an omnidirectional camera underwater. In Section 6, the results of the calibration are presented. In the last section, we draw the conclusions of this work.

2. Camera Design

A custom housing has been designed for the Ladybug 3 camera to make it submersible to a water depth of 60 m. The housing is composed of a transparent poly-methyl methacrylate (PMMA) dome, which contains the camera, and a body made of aluminum alloy, which contains a small form factor computer, dedicated to processing the video feed (Figure 2).

The Ladybug 3 camera comprises six individual cameras (Figure 3). Five of these cameras, referred to as the lateral cameras (and numbered 0 to 4), have the optical centers on the same plane. Their optical axes also lie in the same plane with a 72° separation between neighboring cameras. The remaining camera, numbered as 5, points at the normal direction of the plane.

Each camera contains a 2 MPixel sensor, making a total of 12 MPixel for every frame captured. The images are acquired by a high-end, small form factor computer inside the housing. The role of this computer depends on the aim of the mission. When the aim of the mission is to record panoramic images of the sea, the computer will store the images without any further processing. By contrast, when

the camera is used as a real-time navigation sensor, it will perform all of the image processing and send only the valuable information to the host vehicle, such as, for example, the relative position of another robot. The communication between the camera and the robot is done through an Ethernet (100 Mb/s) connection (Figure 1b).



Figure 2. Final design of the omnidirectional underwater camera.



Figure 3. Arrangement of the cameras. The red overlay indicates the plane where the optic centers of the five lateral cameras are located (identified as Camera 0 to Camera 4). The last camera (Camera 5) has its optical axis approximately perpendicular to this plane.

One of the most important aspects to take into account when designing the housing is the presence of strong refraction effects on the optical rays due to the changes of media. A ray of light coming from the water changes its direction twice before reaching the sensor, as it must pass through two medium transitions (water-PMMA and PMMA-air). The change of direction is described by Snell's law [23], and it depends on two factors: the angle between the incident ray and the normal of the surface at the incidence point and the refraction indexes of the two media. Therefore, the refractions of the rays of light depend strongly on the geometry of the housing. These refractions affect the FOV of the individual cameras and the amount of overlapping between the images. Given the fact that the original camera lenses are designed for land applications, the use of an underwater housing may result in blind spots in the spherical view.

The two most typical geometries used when designing an underwater camera housing are flat and hemispherical interfaces. Flat interfaces are less expensive to manufacture and easy to mount, but they introduce important bending in the rays, which reduces the FOV of the cameras. By contrast, for an hemispherical interface with its center on the exact optical center of the camera, the incident angle and the normal of the surface are exactly the same for all of the rays, and no bending is produced during the transition. However, perfect hemispherical interfaces are difficult to produce and to mount at the exact desired position.

Due to the geometry of the Ladybug 3 camera, a transparent dome has been designed, composed of two pieces joined together with an adhesive. The first piece is a cylinder and covers the whole FOV of the five lateral cameras. A cylindrical interface has the advantage of allowing larger FOV (in one direction) for the five lateral cameras when compared with the option of having five individual flat ports (Figure 4) and is easier and less expensive to produce. The cylindrical housing only increases the FOV in the horizontal axis (perpendicular to its main axis). The FOV in the vertical axis is reduced in a similar way as a flat view-port. A hemispherical interface has the advantage of allowing larger FOV (in one direction) for the five lateral cameras when compared with the option of having five individual flat ports (Figure 4) and is easier and less expensive to produce. The cylindrical housing only increases the FOV in the horizontal axis (perpendicular to its main axis). The FOV in the vertical axis is reduced in a similar way as a flat view-port. A hemispherical interface has the advantage of allowing larger FOV (in one direction) for the five lateral cameras when compared with the option of having five individual flat ports (Figure 4) and is easier and less expensive to produce. The cylindrical housing only increases the FOV in the horizontal axis (perpendicular to its main axis). The FOV in the vertical axis is reduced in a similar way as a flat view-port.

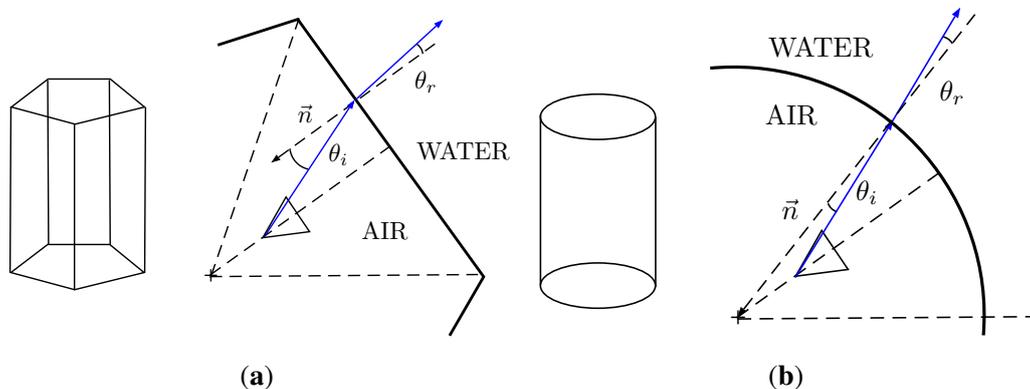


Figure 4. Comparison between a pentagonal prism and a cylinder as view-port options for the lateral cameras. When compared with the flat view-ports of the pentagonal prism, the cylinder has the advantage of being less affected by the refractions of the media transitions, along one of the directions. **(a)** Pentagonal prism shape; **(b)** cylindrical shape.

For the top-looking camera (down-looking in water), the final design was a hemispherical piece. The manufacturing of this piece was significantly more challenging than a flat port alternative, since it required a thermoforming process to obtain the intended shape. However, the flat port option had to be discarded due to the severe reduction of the FOV that it would cause.

In order to test possible scenarios for the choices of the shape of the view-ports, a FOV simulator was implemented. This simulator uses ray-tracing techniques that take into account the full camera projection model and Snell's law for the refraction effects of the possible shapes of the housing. The results of the FOV simulator can be seen in Figure 5, which presents a full-sphere representation of the FOV, with a horizontal view angle in the x -axis of the plot and a vertical view angle on the y -axis.

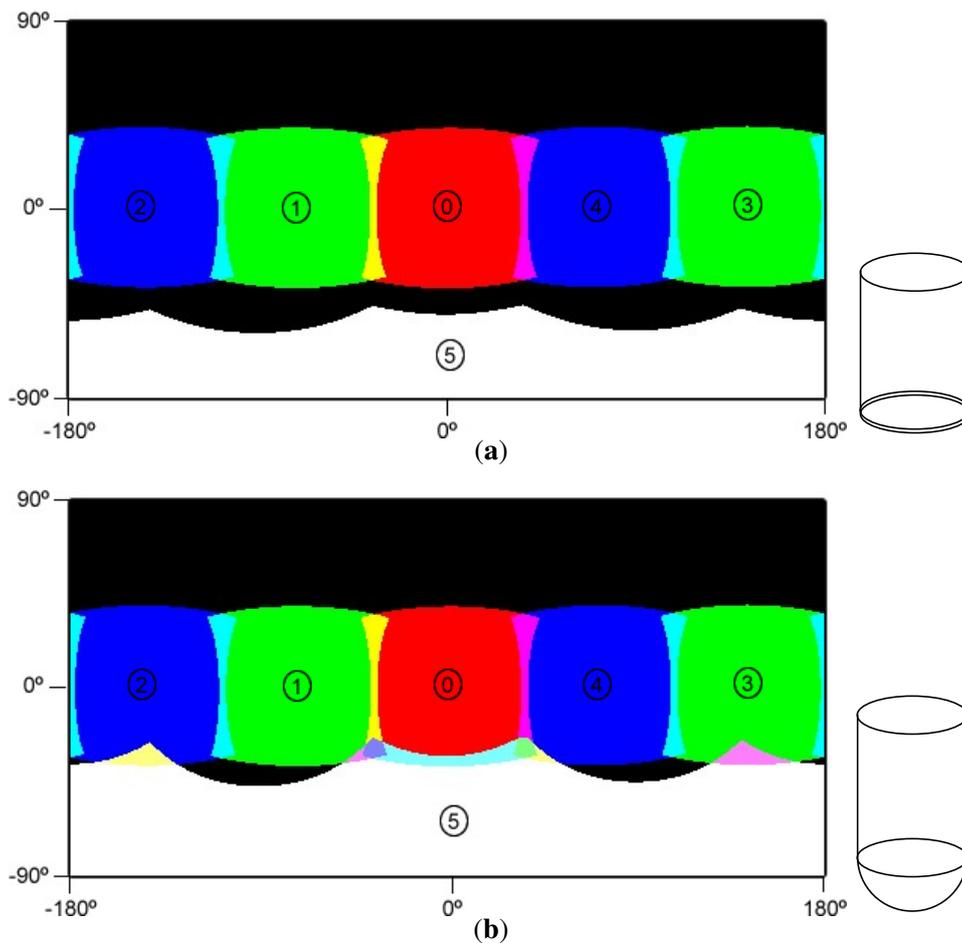


Figure 5. Equirectangular projection of the covered FOV at a 5-m distance in an underwater environment with different configurations. Each colored region represents the FOV of each camera (red, green, blue and white) and the areas of FOV intersection (other colors). (a) Projection of the covered FOV at a 5-m distance with a flat interface for the bottom camera; (b) projection of the covered FOV at a 5-m distance with a hemispherical interface for the bottom camera.

As illustrated in Figure 5a, using a flat view port for Camera 5, this camera would not have any overlap with any of the other cameras. By contrast, the hemispherical view port allows some overlap, but not complete coverage of the lower hemisphere. For this reason, the original optics (3.3 mm of focal length) of Cameras 1, 4 and 5 were replaced for others with wider FOVs (2.95 mm of focal length) to

achieve full coverage of the panoramic view (Figure 6). Further details on the computation of the FOV in underwater environments can be found in Sections 5 and 6.

During the design, it was important to ensure that the junction between the two parts of the dome was placed in a location that would not be visible by any camera, hence avoiding occlusions in the resulting image.

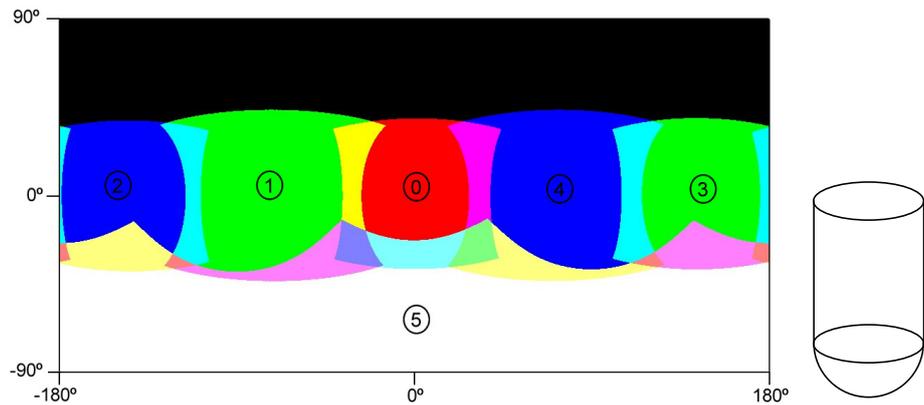


Figure 6. Equirectangular projection of the covered FOV at a 5-m distance with a hemispherical interface for the bottom camera and 2.95-mm focal length optics for Cameras 1, 4 and 5. Each colored region represents the FOV of each camera (red, green, blue and white) and the areas of FOV intersection (other colors).

3. Single Camera Calibration

In this section, we first present the camera model used and then the calibration procedure for each one of the six individual cameras.

3.1. Camera Model

The pinhole camera model [8,9,24] has been adopted for this work due to its compactness (Figure 7) and accurate results. This model will allow one to project any 3D world point $Q = (X, Y, Z)$ in the camera coordinate frame to a pixel position (u, v) in the image plane through Equations (1) and (2).

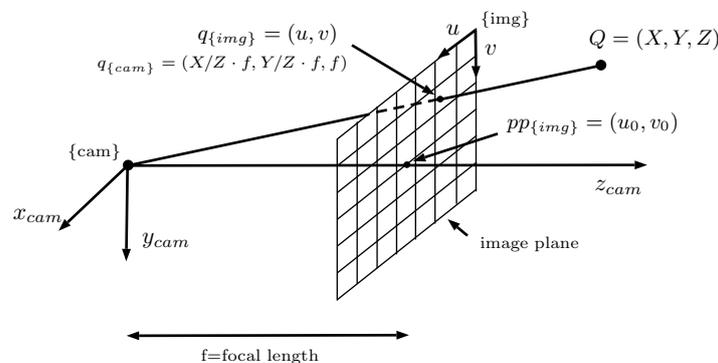


Figure 7. The pinhole camera model.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \quad (1)$$

$$x = \frac{X}{Z}, y = \frac{Y}{Z} \quad (2)$$

where:

- (u_0, v_0) is the location of the principal point in the image plane coordinates. This is the point where the camera optical axis intersects the image plane and is normally located near the center of the image.
- (f_x, f_y) are the focal lengths along the x and y directions, expressed in pixels. Most cameras have sensors with squared pixels, where $f_x = f_y$.

However, all lenses induce image distortions that are not modeled by the pinhole camera model. The most common one is the radial distortion, which is due mainly to the shape of the lenses and produces nonlinear distortions along the radial direction from the principal point. The further from the center of the image, the higher is the radial distortion. Radial distortion is very strong for wide-angle lenses. For example, in Figure 8a, the edges of the rectangular board look curved when, in reality, they are perfectly straight. It can also be noticed in the image that this distortion becomes more important in the regions further from the center: the bottom edge of the board looks straighter than the top one. Fortunately there are models that introduce corrections in the original images (distorted) to undistort them and create a new undistorted (ideal) image that follows the pinhole model (Figure 8b). In this work, we adopt the model proposed by Kannala [10].

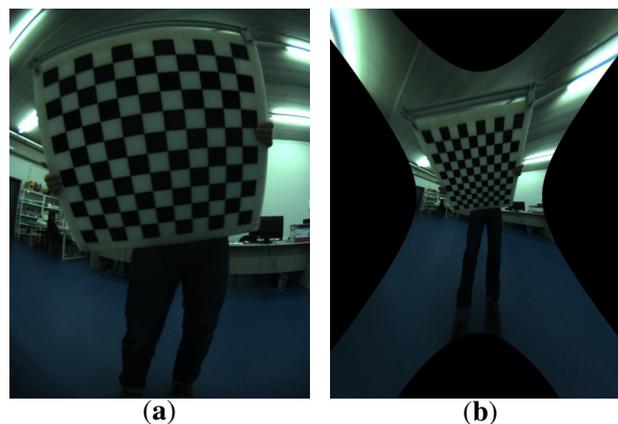


Figure 8. A sample image of a checker board captured by a wide-angle lens camera used for any standard calibration toolboxes, before (a) and after (b) the distortion correction. (a) Original image; (b) undistorted image.

Let (u, v) be the ideal (non-observable, distortion-free) pixel image coordinates corresponding to a point $Q = X, Y, Z$. These coordinates do not match with the real ones (u_d, v_d) due to the distortion effect. The relation between the real point x, y (expressed in the camera reference plane, $Z = 1$) and a

virtual point x_d, y_d that projects to the pixel coordinates (u_d, v_d) according to the pinhole model can be found through:

$$\theta_d = \theta (1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + k_4\theta^8) \quad (3)$$

$$x_d = x \left(\frac{\theta_d}{r} \right) \quad (4)$$

$$y_d = y \left(\frac{\theta_d}{r} \right) \quad (5)$$

where $r^2 = x^2 + y^2$, $\theta = \text{atan}(r)$ and k_1, k_2, \dots, k_n are distortion coefficients.

3.2. Calibration

A two-step method for the calibration has been devised and implemented. In the first step, a standard calibration toolbox is used in order to provide an initial estimate of the intrinsic parameters. In the second step, these values are refined in order to obtain better results.

3.2.1. Initialization

The OpenCV Library (Open Source Computer Vision Library) [25,26] methods for camera calibration and 3D reconstruction have been used to compute a first estimate of the intrinsic values of the six independent cameras. Given multiple shots of a planar grid (typically a checker board) acquired at different positions and orientations, the methods in this module allow one to compute the intrinsic camera parameters and extrinsic parameters for each of the views, based on the approaches presented in [10,24,27].

In most cameras, these methods are accurate enough for metrology applications, and there is no need for further refinement. However, for an OMS camera using wide-angle lenses, a refinement procedure can help to obtain better results, especially in the regions close to the borders of the image. In these regions is where the overlapping between camera images take place, and a very accurate calibration is required in order to avoid visible misalignments in the final panoramas.

3.2.2. Refinement

Due to the high distortion of the images (seen in Figure 8), it is not possible to place the checker board close to the corners of the image while seeing it entirely, as required by the standard calibration packages. This fact leads to calibration results that are inaccurate in the regions close to the image corners. For this reason, a different approach was implemented, which uses a more versatile pattern and allows results when only a portion of the pattern is visible in the images. An aerial image of the city of Girona (Figure 9a) has been used for this purpose, since it provides a large number of visual features at different scales.

Different shots of the poster in different positions and orientations have been taken, paying special attention to capturing parts of the poster rich in features in the corners of the images.

As a following step, scale-invariant feature transform (SIFT) [28] features are found in all of the selected shots and in the original poster image. Every feature has a keypoint and descriptor associated.

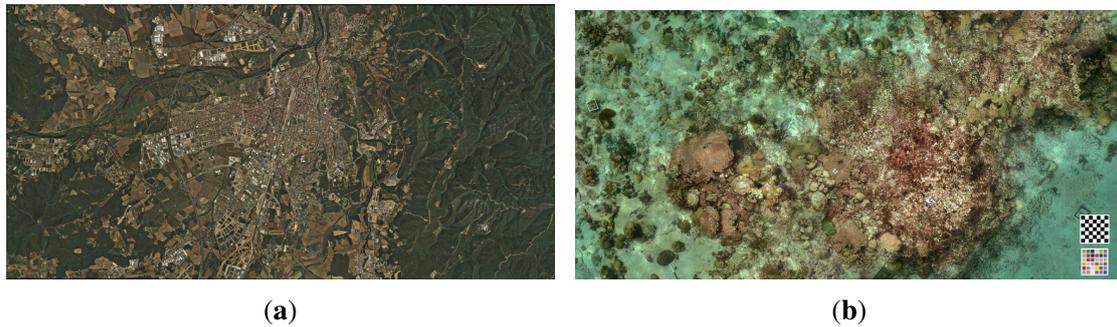


Figure 9. Posters used for the dry and underwater calibration, respectively. **(a)** Aerial image of the city of Girona used for both the intrinsic and extrinsic calibration procedures. The dimensions of the printed poster are 2.395×1.208 m; **(b)** Underwater image used for the optimization of the housing parameters. The printed poster measures 7.09×3.49 m and was placed in a flat area at the bottom of the test pool.

Afterward, putative matches between the features in the captured images and the original poster are found following Lowe's criterion [28]. Under this criterion, for every feature in the original image, the nearest neighbor in the captured image is defined as the keypoint with minimum Euclidean distance. In order to decide if this match is potentially the correct one, the second closest neighbor is found, and a ratio between its distance is computed as per Equation (6).

$$Ratio = \frac{Distance\ closest}{Distance\ second\ closest} \quad (6)$$

If the ratio is smaller than a threshold, the matching is considered valid. The threshold chosen in this work was 0.7, which, according to Lowe, eliminates about 95% of false matchings and discards about 8% of the correct matches. To find the closest and the second closest matches among all of the features, the Fast Library for Approximate Nearest Neighbors (FLANN) [29] has been used to speed up the process, instead of purely brute force.

If the number of matches is greater than a fixed minimum, e.g., 100 matches, then the image is accepted to be used to optimize the intrinsic parameters. Otherwise, the image is rejected.

For the accepted images, the poster pose that minimizes the re-projection error of all correct matches found is estimated making use of the initial intrinsic parameters estimated in the previous step and solving the perspective-n-point (PnP) problem [9]. The PnP problem is the problem of the determination of the position and orientation of a calibrated camera given a set of n correspondences between 3D points and their 2D projections. In this case, the 3D points belong to the poster, and the 2D projections belong to the camera images. We assume that the poster is totally flat and in the plane $z = 0$; hence, the 3D points used are the coordinates of the matches in the original poster image, in the real scale, *i.e.*, meters, with its coordinate z set to zero. The 2D projections are the coordinates of matches in the camera images, in pixels. The implementation in OpenCV, which uses an iterative method to minimize the re-projection error, has been used to solve the problem. The solution is robust to outliers, thanks to the use of random sample consensus (RANSAC) [30]. The features discarded by RANSAC will not be used further in the intrinsic parameter optimization.

From the set of all RANSAC inliers, a subset is chosen that contains image points that are well distributed on the image plane. The objective is to have approximately the same number of features in all regions of the image. A bucketing strategy [31] has been implemented to achieve this purpose. The image is divided into a number of disjoint regions of the same size. Then, every feature is associated with the region to which they pertain. Only one feature per region, selected randomly, will be used for the further calibration steps (Figure 10). The number of regions can be set arbitrarily, but must be in concordance with the size of the image and the number of features able to match in every shot, as it represents the maximum number of features selected. For this work, this parameter has been set to 500.

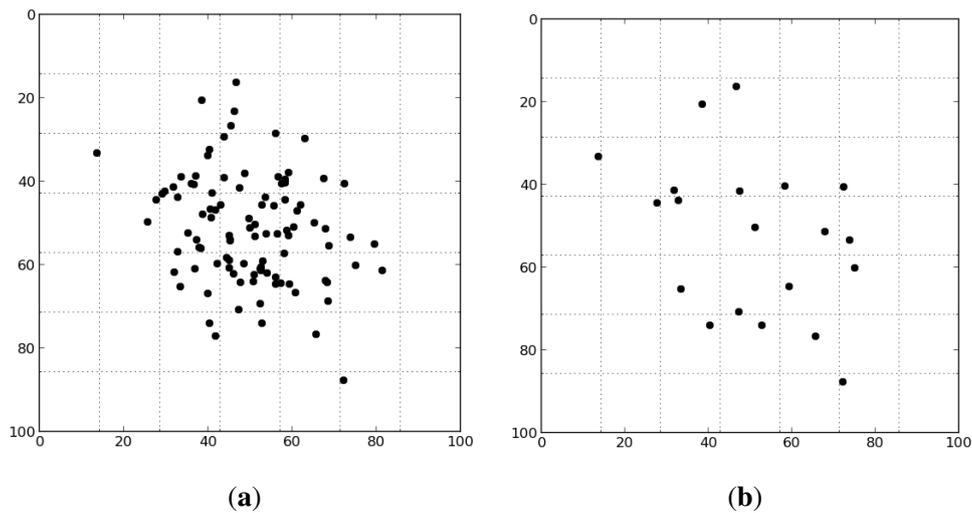


Figure 10. Selection of features with simulated data and 49 regions. (a) All feature matchings are associated with a region of the image; (b) only one feature per region is used for the optimization procedure.

Finally, we are ready to define the optimization problem that will find the refined intrinsic parameters. The number of variables to estimate will be the totality of the intrinsic parameters (focal length, principal point coordinates and distortion parameters) and the pose of the poster ($x, y, z, \alpha, \beta, \gamma$) in all of the selected images (Equation (7)). The initial values for the intrinsic parameters will be the ones found in the initialization step, while the initial values for the pose of the poster in every image will be the results of the PnP problem solved previously.

$$\theta = [f, u_0, v_0, k_1 \dots k_4, x_{img(0)}, y_{img(0)}, z_{img(0)}, \alpha_{img(0)}, \beta_{img(0)}, \gamma_{img(0)} \dots x_{img(n)}, y_{img(n)}, z_{img(n)}, \alpha_{img(n)}, \beta_{img(n)}, \gamma_{img(n)}] \quad (7)$$

The re-projection error for every feature p^p of the poster, which has a matching p^c in a captured image, can be defined as:

$$e_{reproj}(p^p, p^c, \theta) = \|\pi(p^p, \theta) - p^c\|^2 \quad (8)$$

where π is the projection function of the 3D point associated with p^p to the camera through Equations (1) and (3)–(5). For all of the matched features in an image, the total re-projection error can be expressed as:

$$\sum_k e_{reproj}(p_k^p, p_k^c, \theta) \quad (9)$$

Then, the cost function that includes all of the selected images, can be expressed as:

$$\sum_{img_i} \sum_k e_{reproj}(p_k^p, p_k^c, \theta) \quad (10)$$

This optimization problem is solved using a Levenberg–Marquardt algorithm [32,33] that minimizes the sum of all re-projection errors.

In order to quantify the uncertainty of the estimated parameters, a Monte Carlo analysis has been carried out. It consisted of repeating the same estimation procedure a significant number of times, but in this case, the set of paired features between the captured images and the original poster were not found using SIFT and RANSAC. For every feature in the original poster used for the estimation of the parameters, we find its projection in the captured image using the estimated intrinsic parameters and poster poses and add Gaussian noise. The Gaussian noise has zero mean, and the standard deviation was the value obtained as the standard deviation of all of the residues of the optimization. The study of the variability of the estimated parameters during the simulations provides valuable information about the uncertainty of the parameters. It also allows one to validate the adequacy of the input data in terms of the observability of the parameters being estimated. An example of this is checking that the used poster poses are sufficiently well spread to allow an accurate estimation of the parameters.

4. Extrinsic Calibration

The main problem when working with multiple cameras is the determination of the exact geometric relationship between the different camera frames. These rotations and translations are referred to as the extrinsic parameters. In this work, both these and the intrinsic parameters are assumed not to change in time. Each camera has its own independent coordinate system, but the local coordinate system of Camera 5 has been defined as a global frame that will be used to deal with the external world (Figure 11). Therefore, the global coordinate system follows the standard convention for underwater frames, where the z -axis points downwards into the sea floor, making it more intuitive when integrating the camera in a robotic platform.

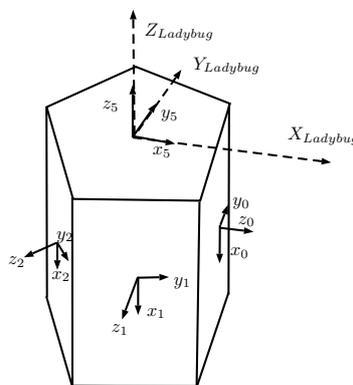


Figure 11. The relationship between cameras and the global coordinate system.

The procedure to estimate the extrinsic parameters will be very similar to the one used in the refinement of the intrinsic ones, in Section 3.2.2. However, in contrast, in this subsection, images from two or more cameras acquired at the same exact time frame are needed.

In order for the algorithm (Algorithm 1) to work properly, all cameras must acquire a recognizable section of the poster where one of the other cameras also acquires a different section of the poster simultaneously. The observation of different parts of the poster by two cameras at the same instant implicitly imposes constraints on the relative placement and relative orientation of the two cameras. These constraints are used in the cost function that is minimized. It should be noted that a standard checker calibration pattern cannot be used in this step, since it would be difficult to determine automatically which part of the pattern would be seen by each camera, due to the similarity of the squares in the grid. It is advantageous to have a wide range of images where the poster is visible from as many cameras as possible in different positions and orientations in order to estimate the unknown variables effectively.

In the first estimation, we will focus on estimating the rotations of each camera frame, and for that, we will assume that the sensors of the camera are placed ideally, *i.e.*, the optical centers of the five lateral cameras lie in the same plane and are placed in a perfect regular pentagon. The remaining camera is assumed to have its optical center placed along a perpendicular line passing through the center of the pentagon. With these assumptions, only three variables are needed to fully describe the location of the optical center of all of the cameras: d_1 represents the distance between the center of the pentagon and the optical center of the three lateral cameras with the original optics (Cameras 0, 2, 3); d_2 represents the same distance, but for the lateral cameras with the new 2.95-mm focal length optics (Cameras 1, 4); and d_3 represents the minimum distance from the plane containing the lateral cameras to the optical center of the remaining camera (Figure 12). For the optimization, the initial values of these three variables will be an approximated physical measurement.

Algorithm 1 Extrinsic parameter calibration.

```

Find SIFT features in original poster image
for  $frame = 0$  to  $frame = n$  do
  for  $cam = 0$  to  $cam = 5$  do
    Find SIFT features in the image
    Poster found?
  end for
  if Poster found in more than two cameras then
    Solve PnP problem for one of the cameras
    Find initial estimation of poster position in global coordinates
    Add to the list of frames and features to be used in the optimization
  end if
end for
Optimize parameters to minimize re-projection error sum

```

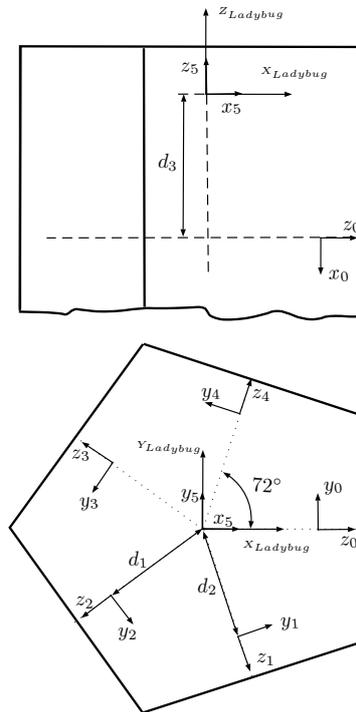


Figure 12. The geometrical unknowns during the first estimation of the extrinsic parameters of the cameras are: d_1 (for Cameras 0, 2 and 3), d_2 (for Cameras 1 and 4), d_3 (for Camera 5) and the exact orientation of each camera. Side view (top) and top view (bottom).

The orientation of the global and the Camera 5 frames will be fixed, while the orientation of frames of Cameras 0 to 4 will be estimated from the optimization procedure. Its initialization will be its ideal orientation (Figure 12). For Cameras 0 to 4, each camera is rotated 72° along the x -axis with respect to its neighbors, with the z - and y -axis of Camera 0 coinciding with the x - and y -axis of the global frame.

An initial pose of the poster for every different time frame will be needed to start the optimization procedure. This can be estimated by solving the PnP problem from any camera seeing the poster in that exact time frame and then converting it to global coordinates.

The vector containing the totality of parameters to estimate will be:

$$\theta = [d_1, d_2, d_3, \alpha_{c(1)}, \beta_{c(1)}, \gamma_{c(1)} \dots \alpha_{c(5)}, \beta_{c(5)}, \gamma_{c(5)}, x_{f(0)}, y_{f(0)}, z_{f(0)}, \alpha_{f(0)}, \beta_{f(0)}, \gamma_{f(0)} \dots x_{f(n)}, y_{f(n)}, z_{f(n)}, \alpha_{f(n)}, \beta_{f(n)}, \gamma_{f(n)}] \quad (11)$$

where $[\alpha_{c(i)}, \beta_{c(i)}, \gamma_{c(i)}]$ represent the orientation of camera i and $[x_{f(j)}, y_{f(j)}, z_{f(j)}, \alpha_{f(j)}, \beta_{f(j)}, \gamma_{f(j)}]$ represent the pose of the poster position in the global frame in time frame j .

Through Equations (8)–(10), the cost function that includes the re-projection errors of all of the features present in the images of the selected frames can be expressed as:

$$\sum_{frames_j} \sum_{cam_i} \sum_k e_{reproj}(p_k^p, p_k^c, \theta) \quad (12)$$

After this first estimation, a new optimization procedure will be performed, fixing this time the rotations of the camera frames and without constraints regarding its location. The cost function will remain the same, while the vector containing the totality of parameters to estimate will be:

$$\theta = [x_{c(1)}, y_{c(1)}, z_{c(1)} \dots x_{c(5)}, y_{c(5)}, z_{c(5)}, x_{f(0)}, y_{f(0)}, z_{f(0)}, \alpha_{f(0)}, \beta_{f(0)}, \gamma_{f(0)} \dots x_{f(n)}, y_{f(n)}, z_{f(n)}, \alpha_{f(n)}, \beta_{f(n)}, \gamma_{f(n)}] \quad (13)$$

where $[x_{c(i)}, y_{c(i)}, z_{c(i)}]$ represent the position of camera i .

The estimation of the relative rotations and translations of the cameras is done in two stages due to a unique step that led to physical non-sense values for the translations due to the sensitivity of the rotation parameters. A Monte Carlo analysis has been carried out to determine the uncertainty of the estimated parameters, similarly to the one described for the intrinsic parameters estimation.

5. Underwater Calibration

The direction of the rays of light changes in every medium transition found along the path from a point underwater to the imaging sensor inside the camera. In order to model accurately the distortion due to this effect, it becomes essential to explicitly model and simulate the intersection of each light ray with different media, as detailed next.

5.1. Ray Tracing

Once the intrinsic parameters of each camera are known, each pixel of an undistorted image can be associated with a 3D ray originated at the optical center of the camera and described in the 3D coordinate frame of the camera. The direction vector of this ray can be computed through Equation (14).

$$v_{0\{local\}} = \left(\frac{u - u_0}{f_x}, \frac{v - v_0}{f_y}, 1 \right) \quad (14)$$

The local vector can be transformed to the global frame pre-multiplying by the rotation matrix R that relates both coordinate systems:

$$v_0 = R \cdot v_{0\{local\}} \quad (15)$$

Let p_0 be the optical center of one of the cameras. The 3D ray can be described as:

$$s = p_0 + kv_0 \quad (16)$$

where k is a scalar.

When the camera is inside the waterproof housing, this ray will change its direction when transitioning to the PMMA interface, due to the refraction effect. The direction of the refracted ray, v_r , can be computed through Snell's law. In a 2D plane, as Figure 13, Snell's law can be expressed as:

$$\sin(\theta_a) \cdot n_{air} = \sin(\theta_g) \cdot n_{PMMA} \quad (17)$$

where θ_a is the angle between the incident ray and the normal vector of the surface in the intersection point, θ_g is the angle between the refracted ray and the normal vector and n_{air} , n_{PMMA} are the refractive indexes of air and PMMA, respectively.

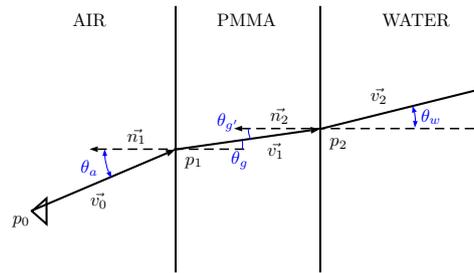


Figure 13. Ray tracing schematic of a single optical ray passing through air, PMMA and water.

Due to the complex geometry of the dome, it is better to work in the 3D space and use the following expression of Snell's law:

$$v_1 = \frac{n_{air}}{n_{PMMA}} (n_1 \times (-n_1 \times v_0)) - n_1 \sqrt{1 - \left(\frac{n_{air}}{n_{PMMA}}\right)^2 (n_1 \times v_0) \cdot (n_1 \times v_0)} \quad (18)$$

where n_1 is the normal vector of the surface in the intersection point between the ray and the surface and v_0, v_1 are the direction vectors of the incident and refracted ray.

In order to find both p_1 and n_1 , the waterproof housing must be geometrically modeled. It can be expressed as the union of a cylinder of radius r , direction vector v_c and origin p_c , with a hemisphere of radius r and center p_s .

The points q of the cylinder can be expressed as:

$$\|(q - p_c) - (v_c \cdot (q - p_c)) v_c\| = r \quad (19)$$

The points q of the hemisphere can be expressed as:

$$\|q - p_s\| = r \quad (20)$$

Knowing both the expression for the optical ray and the geometrical model of the housing, we can find the intersection point by replacing the expression of the ray by the surface point q , both in the case of the cylindrical part:

$$\|(p_0 + kv_0 - p_c) - (v_c \cdot (p_0 + kv_0 - p_c)) v_c\| = r \quad (21)$$

$$(p_0 - p_c + kv_0 - (v_c \cdot (p_0 - p_c + kv_0)) v_c)^2 - r^2 = 0 \quad (22)$$

and the hemispheric:

$$\|p_0 + kv_0 - p_s\| = r \quad (23)$$

$$(kv_0)^2 + 2k(v_0 \cdot (p_0 - p_s)) + (p_0 - p_s)^2 - r^2 = 0 \quad (24)$$

Solving the value of k from these equations, the intersection point can be found as:

$$p_1 = p_0 + kv_0 \quad (25)$$

Before applying Snell's law, the normal vector of the surfaces in the intersection point p_0 needs to be found. For the cylindrical part, the normal in a point q can be found through:

$$n = p_c - q + ((p_c - q) \cdot v_c) v_c \quad (26)$$

while the normal vector for the hemispheric part in a point q can be found through:

$$n = p_s - q \quad (27)$$

We can finally compute the refracted ray direction vector v_1 through the vectorial expression of Snell's law (Equation (18)). The refractive indexes of air and PMMA are assumed to be invariant for the conditions in which the camera will work. The assumed values are $n_{air} = 1$ [34] and $n_{PMMA} = 1.4914$ [35] (for wavelengths $\lambda = 0.589 \mu\text{m}$). The refracted ray can be expressed as:

$$s = p_1 + kv_1 \quad (28)$$

This ray will change once again its direction when moving from PMMA to water. The new refracted ray can be computed in an analog way to the transition detailed above. The refractive index of water for this transition has been assumed as $n_{water} = 1.333$ ($\lambda = 0.589 \mu\text{m}$). This is the refractive index for fresh water at 20° [36], which was approximately the water conditions of the water tank when the images for the calibration were taken. In the case of seawater, the refractive increases slightly with the salinity. The value of the refractive index can be easily tuned for further use after the calibration procedure if required.

Knowing both the intrinsic and extrinsic parameters of all of the cameras and through the use of the ray tracing approach, it is possible to project any 3D point underwater to any of the cameras composing the omnidirectional image. However, this projection is not straightforward. The distance d from the camera to the 3D point will be computed, and through the ray tracing study and an iterative method, it is possible to find the projection of the 3D point into the camera. An initial value for the projected pixel is required, and then iterating its position until the error between the 3D point and the point of the ray associated with the pixel at a distance d is negligible. The initial value of the pixel can be initialized with the value of the projection in air or simply to the center of the image. This iteration procedure can be solved through the Levenberg–Marquardt algorithm.

5.2. Housing Parameters Optimization

Due to the ray bending, any small variation in the assumed relative position of the housing can significantly affect the final direction of the rays and end up generating projection errors. In order to avoid this, the difference between the estimated relative position and the real one has to be minimized. There is no guarantee that the center of the global camera frame coincides exactly with the cylinder axis of the housing, nor is the alignment between the global coordinate system and the cylinder perfect. For this purpose, the relative position of the housing with respect to the camera will be estimated in a procedure almost identical to the one in Section 4, but using now images captured underwater and with a different poster specially prepared to be placed underwater (Figure 9b). The poster was placed at the bottom of a water tank of 5 m in depth, and different images were captured with the camera in different positions and orientations. It is important that the images are captured in the best lighting conditions possible, in order to minimize the light absorption caused by water, making difficult the recognition of features on the images.

The parameters to optimize (Equation (29)) are the location of the center of the cylinder (C_x, C_y), the orientation of the cylinder (R_x, R_y, R_z) and the position of the center of the hemisphere ($S_x,$

S_y, S_z), illustrated in Figure 14, apart from the poses of the poster for every different time frame i used $[x_{f(i)}, y_{f(i)}, z_{f(i)}, \alpha_{f(i)}, \beta_{f(i)}, \gamma_{f(i)}]$. All of the parameters are with reference to the global frame of the camera.

$$\theta = [C_x, C_y, R_x, R_y, S_x, S_y, S_z, x_{f(0)}, y_{f(0)}, z_{f(0)}, \alpha_{f(0)}, \beta_{f(0)}, \gamma_{f(0)} \dots x_{f(n)}, y_{f(n)}, z_{f(n)}, \alpha_{f(n)}, \beta_{f(n)}, \gamma_{f(n)}] \quad (29)$$

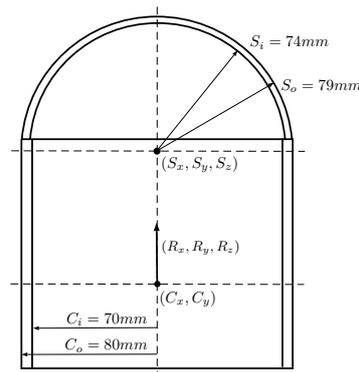


Figure 14. Cross-section representation of the PMMA waterproof housing.

The initial values for the target parameters are established from the knowledge of approximate geometry of the housing. For the poster pose initialization, for every different time frame, the poster pose can be estimated by solving the PnP problem from any camera seeing the poster without taking into account the distortion caused by the housing. It is preferably to compute this initial pose from Camera 5, as the distortion caused by the hemispherical port is less important than that caused by the cylindrical port; hence the initialization values will be more accurate. The estimation process is similar to the one described in Section 4. A cost function is defined on the residues of the re-projection of the points detected in the multiple images of the poster (Equation (12)). This cost function is parametrized by the unknowns described above and minimized using the Levenberg–Marquardt algorithm. A Monte Carlo method is also used to determine the uncertainty of the estimated values, as described in the last section.

6. Results

In this section, we present both the numerical and graphical results of all of the steps during the calibration procedure. In order to interpret correctly the reconstructed panoramas, a subsection is first presented explaining in detail the process behind the creation of the panoramas.

6.1. Panorama Composition

After a successful calibration of the omnidirectional camera, each pixel of any image can be associated with a 3D ray in space. The length of this ray depends on the distance to the objects in the scene. Except for the small area where there is image overlap, it is not possible to estimate the distance to

the objects from just a set of images acquired at a single location. For the overlapping parts, it is possible to estimate the distance to features seen from both cameras using the same method as in a conventional stereo camera. However, this would be very difficult in underwater environments where the overlapping is very small, the baseline is extremely short and there are very few features in most of the environments. Furthermore, this would be expensive computationally and would make it impossible to render panoramic images in real time with the existing hardware. For this reason, for visualization purposes, the world around the camera is assumed to be a sphere, where all of the points sensed by the camera are at a constant distance, pre-selected by the final user. Once the sphere radius is defined, a spheric point cloud is quick to compute, and it can be easily loaded in a computer 3D viewer or re-projected into a 2D image. For omnidirectional cameras, the equirectangular projection is the most commonly used [37].

Given a world point in Cartesian coordinates $Q = (X, Y, Z)$, it can be converted to spherical coordinates (Figure 15) $Q = (\theta, \phi, R)$ through Equations (30)–(32).

$$R = \sqrt{X^2 + Y^2 + Z^2} \quad (30)$$

$$\theta = \text{atan2}(Y, X), 0 \leq \theta \leq 2\pi \quad (31)$$

$$\phi = \text{acos}\left(\frac{Z}{R}\right), 0 \leq \phi \leq \pi \quad (32)$$

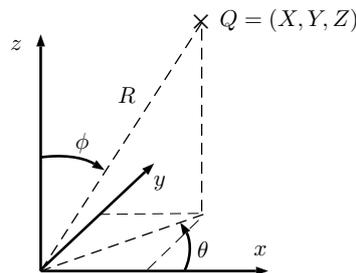


Figure 15. Conversion from Cartesian to spherical coordinates.

The equirectangular projection projects a given point q to a cylinder (Figure 16) through Equations (33) and (34):

$$u = \frac{\theta + \pi}{2\pi} \cdot W \quad (33)$$

$$v = \frac{\phi}{\pi} \cdot H \quad (34)$$

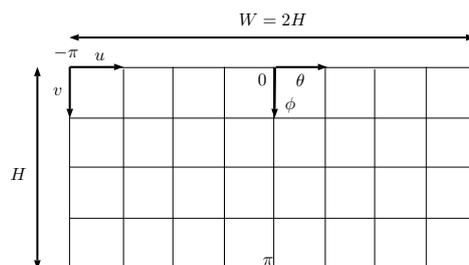


Figure 16. Equirectangular projection.

The inverse equations are:

$$\theta = \frac{u \cdot 2\pi}{W} - \pi \quad (35)$$

$$\phi = \frac{v \cdot \pi}{H} \quad (36)$$

The first step when composing a panorama (Algorithm 2) is choosing its parameters: projection type, projection distance and final size. For every pixel of the panorama, the 3D point it represents is computed according to the inverse equations of its projection (Equations (35) and (36)). This 3D point is then projected to each one of the six cameras according to its intrinsic and extrinsic parameters. For underwater panoramas, it will be necessary to do a numeric iteration to find this projection through the equations presented in Section 5. If the point is only in the FOV of one camera, we will give to the pixel of the panorama the same intensity values as the pixel corresponding to the projection of the 3D point into the camera. In the case of overlapping regions, a blending criterion [38] will be needed, to establish the value of the panorama pixel.

Algorithm 2 Panorama construction.

Choose projection

Set suitable sphere radius: r

Set panorama size

for all pixels in panorama **do**

 Compute 3D point according to pixels position, projection inverse equations and r

for $cam = 0$ to $cam = 5$ **do**

 Project 3D point to camera image.

if Point falls into camera FOV **then**

 Store color information and location of the pixel.

end if

end for

if The point is only seen for one camera **then**

 Give to the pixel the same value of the pixel of the camera that represents the point

else if The point is seen for more than one camera **then**

 Give to the pixel a value according to a selected blending criterion

else {The point is not in the FOV of any camera}

 Set the value of the pixel to black

end if

end for

Three different criteria are presented below. The first one does not do any smoothing on the transition between cameras. The second one is a basic, but fast smoothing approach, which can be executed online. The last one makes a smooth blending offline. A practical comparison between them can be found in Table 1.

Table 1. Transition in a panorama with different blending criteria applied and without or with individual gain correction. The color transition is more homogeneous when applying gain corrections, and the transition is smoother when moving from left to right using the blending criterion approach.

	Blending Criterion		
	Closest Camera	Weighted Mean	Gradient Blending
Original color			
Corrected color			

- Closest camera: From all of the pixels that represent the same 3D point:

$$p_{cam} = (u_{cam}, v_{cam}), \quad 0 \leq cam \leq 5 \quad (37)$$

the one with the minimum euclidean distance to its principal point is chosen:

$$d(p_{cam}, pp_{cam}) = \sqrt{(u_{cam} - u_{0,cam})^2 + (v_{cam} - v_{0,cam})^2} \quad (38)$$

The reason for using this criterion is that the distortion is lower for the points closer to the principal point of the image than for the ones further away. This will reduce the error in the panoramas related to the distortion of the lenses.

- Weighted mean: A weighted mean of all of the pixels representing the same 3D point is performed, giving more weight to the pixels closer to its principal point. A blending width threshold will decide where this criterion is applied. A bigger blending width will mean a smoother transition, but with a higher degree of blurriness. This method can be considered as a simpler version of Burt and Adelson's method [39].

The minimum distance between the pixels and their principal point is computed using Equation (38). The final value of the intensity for each one of the channels (R,G,B) is defined as:

$$I_{channel} = \frac{\sum_{cam} I_{p,cam,channel} \cdot (blending\ width - (d(p_{cam}, pp_{cam}) - d_{min}))}{\sum_{cam} blending\ width - (d(p_{cam}, pp_{cam}) - d_{min})} \quad (39)$$

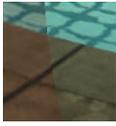
for $blending\ width \neq 0$.

- **Gradient blending:** Gradient blending methods are able to unify different exposures seamlessly and can lead implicitly to a high dynamic range from a set of low dynamic range images. However, they require solving large sparse systems of equations to recover the luminance from the gradient vectors. This method is based on the computation of the vertical and horizontal image gradients of the unblended image produced by the first criterion. The gradient field is modified to impose null gradients along the transition borders among images. This modified gradient field is no longer consistent, in the sense of allowing one to recreate an image that has this exact gradient field. Therefore, a least squares approximation to a consistent gradient field is performed. The final blended image is obtained from this approximated gradient field. More details for this technique can be found in [38,40].

To improve the final panoramic image and make the transitions between cameras softer, a gain compensation can be carried out. When in capture mode, the camera itself computes overall gain and shutter values to obtain visually pleasant images. All of the individual cameras are set with these values, but due to the fact that the lighting conditions are different for each one of the cameras and the replaced optics (Section 2) have slightly different aperture values than the original ones, there are luminosity differences between images from different cameras. To reduce these differences, the approach presented by Brown and Lowe [41] was implemented. This technique uses an individual gain for each image to minimize the intensity differences on the overlapping regions between images. The results of applying this technique, compared with the images without correction, can be seen in Table 1.

For a correct visualization of the results, it is very important that the estimated “radius” where the image is projected is properly set. It is important to notice, as well, that if all of the objects of the scene are not at the same distance, it will not be possible to align perfectly different images composing the panorama. This is not an error of the projection of the images, but rather an effect of the unknown distance of all of the objects of the scene, which is assumed to be constant and fixed. As can be seen in Table 2 for a scene with objects at a different distance, different distance projections will result in different objects aligning the overlapping regions of the panoramas.

Table 2. Details of the same scene projected at different distances. The details are ordered by increasing distance to the camera.

		Detail 1	Detail 2	Detail 3
Projection Distance	$R = 2$ m			
	$R = 5$ m			
	$R = 10$ m			

6.2. Single Camera Calibration

As explained in Section 2, for achieving a complete semi-spherical FOV without blind spots, two different type of lenses have been used. Three cameras were left with the original optics of a 3.3-mm focal length, while the other three were replaced with wider angle lenses of 2.95 mm in focal length. For the sake of simplicity, we will present here only the results for one of the optics, namely a 2.95-mm one, since it is more illustrative due to the presence of stronger distortions. The results for the other optics are similar, apart from minor numeric variations.

For the initialization step, where a standard calibration toolbox has been used, a few parameters need to be set before executing the toolbox, to obtain the best results possible. These parameters are the initial guesses for focal lengths and the principal point location. The initial guesses were set from the information provided by the lenses and camera manufacturers. The numeric index used to evaluate how accurate the calibration is is the root mean square (RMS) of re-projection error defined in Equation (40). The final RMS was 0.479 px for Camera 5 (2.95-mm focal length).

$$RMS(\text{reprojection error}) = \sqrt{\frac{\sum d(x_i, \hat{x}_i)^2}{n}} \quad (40)$$

Although the results in the first step appear to be very good, it is worth noting that these calibrations are only done with data from the image center, where all of the squares of the calibration grid are visible. Therefore, it is difficult to evaluate the calibration quality on the image corners, where the distortion effects are most prominent. For omnidirectional cameras, a very accurate single camera calibration is required. The refinement step will use a larger number of features and is better distributed, thus allowing one to obtain more accurate results.

The initial values for the optimization procedure are the values found by the standard calibration toolbox in the previous step. Table 3 presents the final results of the calibration. During the optimization, the RMS decreased from an initial value of 1.35 px to 1.09 px. The first column shows the result of the

standard calibration and the second column the results of the refinement step. Thanks to the refinement step, it has been possible to find matches in regions where it was not possible using only the checkerboard and the standard calibration methodology (Figure 17). This allowed one to estimate the intrinsic parameters much more accurately (Table 3). This is due to the fact that the dataset used in the refinement step had much better distributed features all around the image than the checkerboard dataset, which has most of its features in the central region of the image where the distortion is less prominent. As can be seen in Figure 18, the distribution of the errors closely follows the shape of a 2D Gaussian probability distribution. The result of the focal length after the refinement procedure is much more similar to the values provided by the lense manufacturers than after the initialization step.

Table 3. Initial and refined values of the intrinsic parameter optimization for Camera 5 (2.95-mm focal length) and standard deviation of the Monte-Carlo Simulation (MCS).

Parameter	Standard Calibration	Refinement	Std Deviation MCS (1000 Iterations, $\sigma = 1.086$ px)
Focal length (pixel)	682.47	674.84	0.29
Focal length (mm)	3.0	2.97	0.0013
Principal point (pixel)	[798.66, 617.97]	[799.38, 617.9]	[0.31, 0.33]
Distortion coefficients	$[-8.41 \times 10^{-4}, -1.82 \times 10^{-2}, 1.21 \times 10^{-2}, -3.7 \times 10^{-3}]$	$[-8.16 \times 10^{-4}, -1.1 \times 10^{-2}, 1.19 \times 10^{-2}, -5.3 \times 10^{-3}]$	$[2.69 \times 10^{-5}, 2.18 \times 10^{-4}, 2.17 \times 10^{-4}, 1.07 \times 10^{-4}]$
Number of images used	34	11	11
Number of checkerboard crosses/ matched points	3400	5268	N/A
Number of features used	3400	3494	3494

The last column of Table 3 shows the standard deviation of a Monte Carlo analysis with 1000 runs where it can be seen that the focal length, principal point and distortion coefficients are well determined.

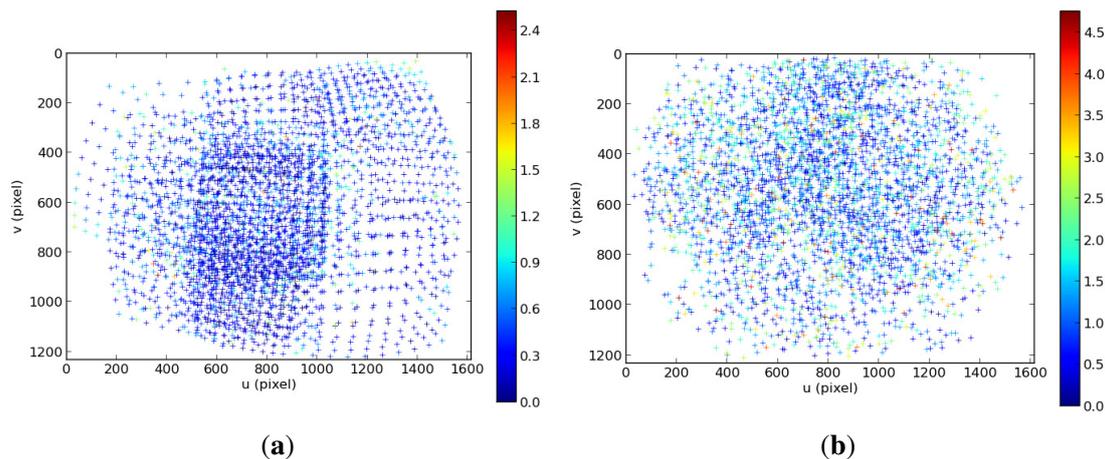


Figure 17. Comparison between the location of the features and its re-projection error in the initialization (a) and refinement step (b) for the intrinsic calibration of Camera 5 (2.95-mm focal length).

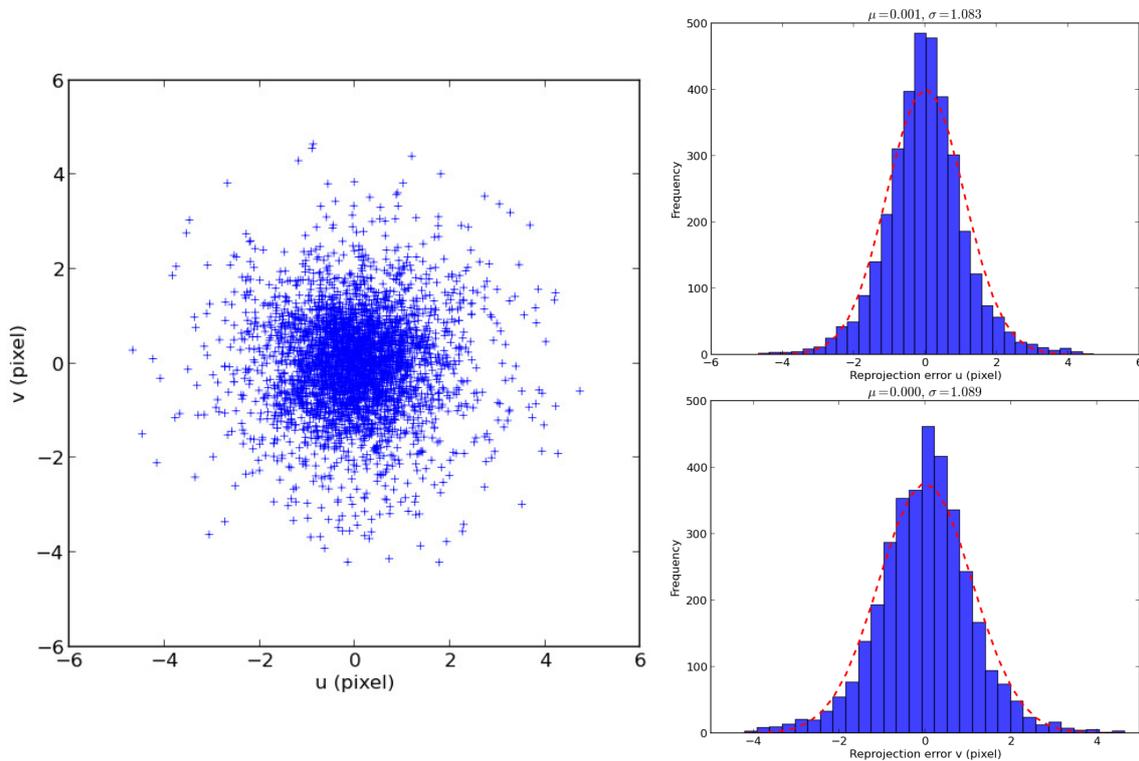


Figure 18. Re-projection error of the features after the refinement step of Camera 5.

6.3. Extrinsic Calibration

Once the intrinsic parameters of all of the cameras are known, the next step is the determination of the external geometric relationship, *i.e.*, translation and rotation, between them. Coarse physical measurements have been used for the initial values of d_1 , d_2 and d_3 (Figure 12). For the rotations, the ideal values have been used. The numeric results obtained during the optimization of the extrinsic parameters can be seen in Table 4. As in the intrinsic parameter calibration, the re-projection error of the features follows a Gaussian distribution and does not depend on its location on the image. The last column of Table 4 shows the standard deviation registered during the Monte Carlo analysis with 1000 runs. As can be seen, the estimated parameters are well determined and show low uncertainty.

Excellent graphic results have been obtained using this calibration parameter, as can be seen in Figure 19.

6.4. Underwater Housing Optimization

The last stage of the calibration consists of the estimation of the parameters associated with the geometry of the waterproof housing that cannot be measured directly. The starting values for these parameters in the optimization are given by approximated physical measurements. The numeric results obtained during the optimization can be seen in Table 5. The distribution of the errors approximately follows the shape of a 2D Gaussian probability distribution with a 3.45-pixel standard deviation.

Table 4. Initial and refined values of the extrinsic parameter optimization and results of the Monte-Carlo Simulation (MCS). Camera 5 is selected as the global reference frame and, therefore, not included in the table.

Parameter	Initial Values	After Optimization	Std Deviation MCS (1,000 iterations, $\sigma = 1.33$ px)
$[\alpha, \beta, \gamma]_{c(0)}$ (rad)	$[0, \frac{\pi}{2}, 0] = [0, 1.571, 0]$	$[1.26 \times 10^{-6}, 1.5719,$ $2.35 \times 10^{-8}]$	$[2 \times 10^{-20}, 9.22 \times 10^{-6},$ $2 \times 10^{-22}]$
$[\alpha, \beta, \gamma]_{c(1)}$ (rad)	$[\frac{2\pi}{5}, \frac{\pi}{2}, 0] = [1.257, 1.571, 0]$	$[1.2575, 1.5713,$ $-2.57 \times 10^{-5}]$	$[5.21 \times 10^{-6}, 8.06 \times 10^{-6},$ $1 \times 10^{-10}]$
$[\alpha, \beta, \gamma]_{c(2)}$ (rad)	$[2\frac{2\pi}{5}, \frac{\pi}{2}, 0] = [2.513, 1.571, 0]$	$[2.524, 1.5751,$ $9.88 \times 10^{-6}]$	$[9.6 \times 10^{-6}, 9.5 \times 10^{-6},$ $3 \times 10^{-21}]$
$[\alpha, \beta, \gamma]_{c(3)}$ (rad)	$[3\frac{2\pi}{5}, \frac{\pi}{2}, 0] = [3.77, 1.571, 0]$	$[3.7705, 1.568,$ $-4.59 \times 10^{-5}]$	$[2.2 \times 10^{-5}, 4.56 \times 10^{-6},$ $1 \times 10^{-18}]$
$[\alpha, \beta, \gamma]_{c(4)}$ (rad)	$[4\frac{2\pi}{5}, \frac{\pi}{2}, 0] = [5.027, 1.571, 0]$	$[5.0258, 1.5742,$ $1.45 \times 10^{-5}]$	$[1.42 \times 10^{-5}, 5.38 \times 10^{-6},$ $1 \times 10^{-19}]$
$[x, y, z]_{c(0)}$ (mm)	$[40, 0, -50]$	$[39.84, -3.37 \times 10^{-6},$ $-61.83]$	$[8.4 \times 10^{-5}, 4 \times 10^{-20},$ $5.5 \times 10^{-5}]$
$[x, y, z]_{c(1)}$ (mm)	$[12.36, -38, -50]$	$[12.55, -40.27, -61.28]$	$[9.9 \times 10^{-6}, 1.2 \times 10^{-4},$ $2.43 \times 10^{-3}]$
$[x, y, z]_{c(2)}$ (mm)	$[-32.4, -23.5, -50]$	$[-32.91, -24.84, -62.81]$	$[7.8 \times 10^{-5}, 2.5 \times 10^{-5},$ $4.9 \times 10^{-4}]$
$[x, y, z]_{c(3)}$ (mm)	$[-32.4, 23.5, -50]$	$[-32.21, 22.89, -61.71]$	$[3 \times 10^{-5}, 1.5 \times 10^{-5},$ $7.5 \times 10^{-4}]$
$[x, y, z]_{c(4)}$ (mm)	$[12.36, 38, -50]$	$[13.42, 39.83, -61.39]$	$[2.2 \times 10^{-4}, 3.5 \times 10^{-4},$ $4.9 \times 10^{-5}]$
Images used		23	
Different time frames		9	
Matched features	10,564		N/A
Features used		3212	
RMS error (pixel)	8.77	1.35	N/A



Figure 19. Equirectangular projection of the interior of the CIRS building, created with a re-projection distance of 10 m and using the closest blending method.

Table 5. Initial and refined values of the housing parameter optimization and results of the Monte-Carlo Simulation (MCS).

Parameter	Initial Values	After Optimization	Std Deviation MCS (500 iterations, $\sigma = 3.455$ px)
Cylinder center (mm)	[0, 0]	[0.514, -0.679]	[0.031, 0.0438]
Cylinder direction vector	[0, 0, 1]	$[-2.33 \times 10^{-3},$ $6.08 \times 10^{-4}, 1]$	$[1.5 \times 10^{-4}, 3.4 \times 10^{-5},$ $1.5 \times 10^{-4}]$
Hemisphere center (mm)	[0, 0, 15]	[0.328, -1.47, -2.6]	[0.155, 0.1876, 0.206]
Number of images used		15	
Number of different time frames		5	
Matched features		7286	N/A
Features used		560	
RMS of re-projection error (pixel)	11.24	3.81	N/A

Good visual results have been obtained using these calibration parameters (Figure 20). Misalignments between camera transitions are barely visible when the panoramas are rendered at the correct distance, even with the simplest blending criterion. Even though, the numeric results obtained are not as flawless as in the previous steps. The three main causes analyzed that could have a negative impact on the mean re-projection error obtained are as follows:

1. The hemispherical part of the housing cannot be modeled as a perfect hemisphere. Given the shape and dimensions of the camera system, it is not feasible to manufacture the PMMA cover as a single piece. For this reason, the hemispherical dome was thermoformed separately and then attached with adhesive to a prefabricated cylindrical body. Although this solution is simple and inexpensive, the hot-forming process is dimensionally inaccurate and induces changes in the thickness of the material along the body of the dome. During the ray tracing, this part has been considered as a perfect hemisphere, which may induce inaccuracies in both the intersection point between the ray and the hemisphere and the refracted ray direction. New manufacturing techniques and a new housing design are being studied to improve this condition.
2. The poster used during the estimation of the underwater parameters was placed at the bottom of a water tank. Due to the upward force that the water under the poster applies against it, it is not possible to guarantee the exact flatness of the poster, especially in its corners. This fact could have increased the final RMS due to the introduction of inaccuracies in the location of the features in the estimation procedure.
3. The refraction indexes used for the estimation of the housing parameters could not have been accurate enough according to the environmental conditions, leading the optimization algorithm to a higher residual re-projection error.

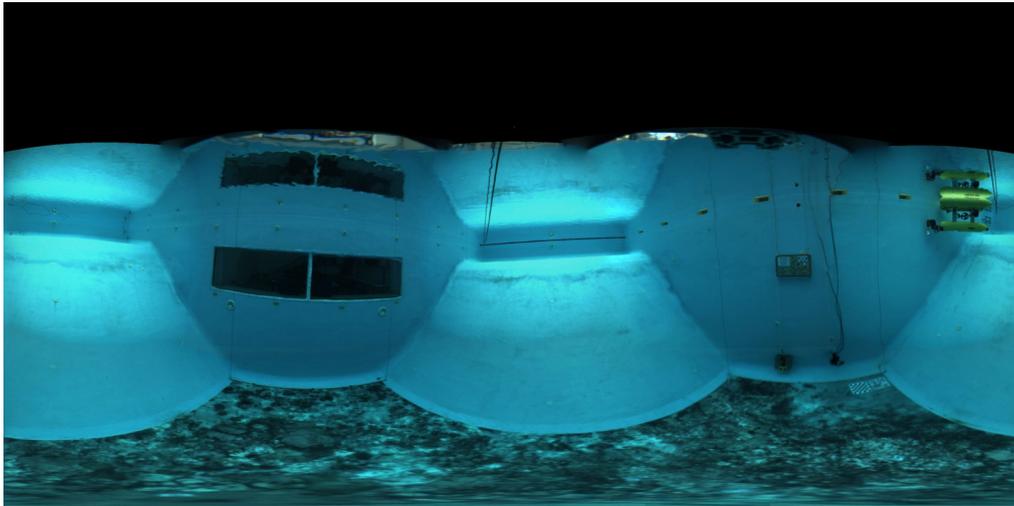


Figure 20. Equirectangular panorama of the CIRS water tank projected at a distance of 4 m with gradient blending.

7. Conclusions

In this paper, we have presented in detail a complete method to calibrate and model an underwater omnidirectional camera composed of multiple single cameras (OMS). This calibration has three different stages. The first one estimates the intrinsic parameters of the individual cameras composing the OMS and can handle strong distortions introduced by wide-angle lenses. The second stage estimates the geometrical relationship between all of the individual frames with respect to a global frame and can deal with cameras with small or no overlapping between cameras. The third stage consists of a ray tracing approach to model correctly the light rays when the camera operates inside a housing in underwater environments. This study can be easily modified to be adapted to different housing geometries. Using this approach, an FOV simulator was developed and used to determine a suitable housing shape and optics replacements to cover a complete hemisphere when operating underwater. The final solution for the Ladybug 3 camera required the replacement of three of the original optics for others with wider FOVs and a dome-shaped housing composed of a cylindrical and a hemispherical part. The results, both numerical and graphical, are very good for the dry calibration part. The parameters estimated have low uncertainty, and the RMS of the re-projection error is small (1.35 px). The final panoramas obtained have very good quality, and there are no visible misalignments when rendered at the correct distances. The results for underwater operation present good results, even though the RMS of the re-projection error is larger (3.85 px), mainly due to unmodeled imperfections from the manufacturing of the hemispherical section of the housing. As future work, other housing shapes and manufacturing techniques will be further analyzed in order to reduce the amount of distortion introduced by it and to improve the results. The graphical results obtained are very good, and misalignments not due to the rendering distance are barely visible.

The high quality of the overall results validate the approach and the methods proposed and pave the way for this OMS to be used both for visualization purposes as a means for popular science or dissemination or as an additional sensor in AUVs and ROVs for navigation, mapping and sea exploration.

Acknowledgments

This research was sponsored by the MORPH EU FP7-Project under Grant Agreement FP7-ICT-2011-7-288704, the Spanish National Project OMNIUS (Lightweight robot for OMNidirectional Underwater Surveying and telepresence) under Agreement CTM2013-46718-R, the Generalitat de Catalunya through the ACCIO (Agència per a la competitivitat de l'empresa de la Generalitat de Catalunya)/TecnioSpring program (TECSPR14-1-0050), with the support of “la Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya” and the University of Girona under a grant for the formation of researchers.

Author Contributions

The work presented in this paper has been done in collaboration of all authors. Josep Bosch and Nuno Gracias designed, implemented and tested the calibration method. David Ribas was in charge of the camera modifications and the housing design. Pere Ridao and Nuno Gracias were the project leaders and in charge of the direction and supervision. All authors discussed the obtained results together and reviewed the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Negahdaripour, S.; Zhang, H.; Firoozfam, P.; Oles, J. Utilizing Panoramic Views for Visually Guided Tasks in Underwater Robotics Applications. In Proceedings of the MTS/IEEE Conference and Exhibition (OCEANS 2001), Honolulu, HI, USA, 5–8 November 2001; Volume 4, pp. 2593–2600.
2. Kalwa, J.; Carreiro-Silva, M.; Tempera, F.; Fontes, J.; Santos, R.; Fabri, M.C.; Brignone, L.; Ridao, P.; Birk, A.; Glotzbach, T.; *et al.* The MORPH Concept and Its Application in Marine Research. In Proceedings of the 2013 MTS/IEEE OCEANS, Bergen, Norway, 10–14 June 2013; pp. 1–8.
3. Ribas, D.; Palomeras, N.; Ridao, P.; Carreras, M.; Mallios, A. Girona 500 AUV: From Survey to Intervention. *IEEE/ASME Trans. Mechatron.* **2012**, *17*, 46–53.
4. Point Grey Research, I. 360 Spherical—Ladybug3—Firewire Camera. Available online: <http://www.ptgrey.com/products/ladybug3/> (accessed on 1 September 2014).
5. Brown, D.C. Close-range camera calibration. *Photogramm. Eng.* **1971**, *37*, 855–866.
6. Clarke, T.A.; Fryer, J.G. The Development of Camera Calibration Methods and Models. *Photogramm. Rec.* **1998**, *16*, 51–66.
7. McGlone, J.; Mikhail, E.; Bethel, J. *Manual of Photogrammetry*; American Society for Photogrammetry and Remote Sensing: Bethesda, MD, USA, 2004.

8. Zhang, Z. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 1, pp. 666–673.
9. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2003.
10. Kannala, J.; Brandt, S.S. A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1335–1340.
11. Scaramuzza, D.; Martinelli, A.; Siegwart, R. A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. In Proceedings of the IEEE International Conference on Computer Vision Systems (ICVS 2006), New York, NY, USA, 4–7 January 2006; p. 45.
12. Mei, C.; Rives, P. Single View Point Omnidirectional Camera Calibration from Planar Grids. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, 10–14 April 2007; pp. 3945–3950.
13. Puig, L.; Bermúdez, J.; Sturm, P.; Guerrero, J. Calibration of omnidirectional cameras in practice: A comparison of methods. *Comput. Vis. Image Underst.* **2012**, *116*, 120–137.
14. Kumar, R.; Ilie, A.; Frahm, J.M.; Pollefeys, M. Simple Calibration of Non-Overlapping Cameras with a Mirror. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK, USA, 23–28 June 2008; pp. 1–7.
15. Ikeda, S.; Sato, T.; Yokoya, N. A Calibration method for an omnidirectional multicamera system. *Proc. SPIE* **2003**, *5006*, 499–507.
16. Li, B.; Heng, L.; Koser, K.; Pollefeys, M. A Multiple-Camera System Calibration Toolbox Using a Feature Descriptor-Based Calibration Pattern. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 1301–1307.
17. Yamashita, A. Underwater Sensing with Omni-Directional Stereo Camera. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, 6–13 November 2011; pp. 304–311.
18. Kunz, C.; Singh, H. Hemispherical Refraction and Camera Calibration in Underwater Vision. In Proceedings of the OCEANS 2008, Quebec City, QC, Canada, 15–18 September 2008; pp. 1–7.
19. Sedlazeck, A.; Koch, R. Calibration of Housing Parameters for Underwater Stereo-Camera Rigs. In Proceedings of the British Machine Vision Conference, Dundee, UK, 29 August–2 September 2011; Hoey, J., Ed.; BMVA Press: Dundee, UK, 2011.
20. CIRS Lab, Universitat de Girona Ladybug Linux Driver, 2015. Available online: https://bitbucket.org/udg_cirs/ladybug_driver (accessed on 1 September 2014).
21. Martinez, A.; Fernández, E. *Learning ROS for Robotics Programming*; Packt Publishing: Birmingham, UK, 2013.
22. CIRS Lab, Universitat de Girona Ladybug ROS Package, 2015. Available online: https://bitbucket.org/udg_cirs/ladybug3_ros_pkg (accessed on 1 September 2014).
23. Ghatak, A. *OPTICS*; Tata McGraw-Hill Publishing Company Limited: New York, NY, USA, 1977.
24. Bouguet, J. Matlab Camera Calibration Toolbox, 2000. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/ (accessed on 1 September 2014).

25. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *25*, 122–125.
26. Bradski, D.G.R.; Kaehler, A. *Learning Opencv*, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2008.
27. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334.
28. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
29. Muja, M.; Lowe, D.G. Fast Matching of Binary Features. In Proceedings of the Computer and Robot Vision (CRV 2012), Toronto, ON, Canada, 28–30 May 2012; pp. 404–410.
30. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395.
31. Zhang, Z.; Deriche, R.; Faugeras, O.; Luong, Q.T. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.* **1995**, *78*, 87–119.
32. Levenberg, K. A method for the solution of certain problems in least squares. *Q. J. Appl. Math.* **1944**, *2*, 164–168.
33. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **1963**, *11*, 431–441.
34. Ciddor, P.E. Refractive index of air: New equations for the visible and near infrared. *Appl. Opt.* **1996**, *35*, 1566–1573.
35. Sultanova, N.; Kasarova, S.; Nikolov, I. Dispersion Properties of Optical Polymers. In Proceedings of the International School and Conference on Photonics (PHOTONICA 2009), Belgrade, Serbia, 24–28 August 2009; Volume 116, pp. 317–324.
36. Daimon, M.; Masumura, A. Measurement of the refractive index of distilled water from the near-infrared region to the ultraviolet region. *Appl. Opt.* **2007**, *46*, 3811–3820.
37. Wang, Y.; Gong, X.; Lin, Y.; Liu, J. Stereo Calibration and Rectification for Omnidirectional Multi-camera Systems. *Int. J. Adv. Robot. Syst.* **2012**, *9*, 143.
38. Prados, R.; Garcia, R.; Gracias, N.; Escartin, J.; Neumann, L. A Novel Blending Technique for Underwater Giga-Mosaicing. *IEEE J. Ocean. Eng.* **2012**, *37*, 626–644.
39. Burt, P.J.; Adelson, E.H. A Multiresolution Spline with Application to Image Mosaics. *ACM Trans. Graph.* **1983**, *2*, 217–236.
40. Gracias, N.; Gleason, A.; Negahdaripour, S.; Mahoor, M. Fast Image Blending Using Watersheds and Graph Cuts. In Proceedings of the British Machine Vision Conference (BMVC 2006), Edinburgh, UK, 4–7 September 2006.
41. Brown, M.; Lowe, G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *74*, 59–73.

3

OMNIDIRECTIONAL MULTI-CAMERA VIDEO STITCHING USING DEPTH MAPS

Following the work presented in the previous chapter, we analyze in depth the state of the art for generating omnidirectional images and video for OMS. Afterwards we propose new methods to improve the quality of the panoramas obtained. This methods take into account the environment where the cameras is located to mitigate the parallax effects. The methodologies are applied and compared in simulation and with a real dataset. This work has recently been submitted to the following journal and is currently under review:

Title: Omnidirectional Multi-Camera Video Stitching using Depth Maps
Authors: **J. Bosch**, K. Istenič, N. Gracias, R. Garcia, and P. Ridao
Submitted to: IEEE Journal of Oceanic Engineering
Quality index: JCR2016 Ocean Engineering IF 2.297, Q1 (2/14)

J. Bosch, K. Istenic, N. Gracias, R. Garcia, and P. Ridao. “Omnidirectional Multi-Camera Video Stitching using Depth Maps”. *IEEE Journal of Oceanic Engineering*
Manuscript submitted for publication

Abstract

Omnidirectional vision has recently captured plenty of attention within the computer vision community. The popularity of cameras able to capture 360° has increased in the last few years. A significant number of these cameras are composed of multiple individual cameras that capture images or video, which are stitched together at a later post-process stage. Stitching strategies have the complex objective of seamlessly joining the images, so that the viewer has the feeling the panorama was captured from a single location. Conventional approaches either assume that the world is a simple sphere around the camera, which leads to visible mis-alignments on the final panoramas, or use feature-based stitching techniques that do not exploit the rigidity of multi-camera systems.

In this paper we propose new stitching strategies for both online and offline applications. The techniques aim to stitch the images according to available information on the multi-camera system and the environment. Exploiting the spatial information of the scene helps to achieve significantly better results. While for the online case this information is obtained from a SLAM process, for the offline case, it is estimated from a 3D-reconstruction of the scene. The information available is represented in depth maps, which provide all information in a condensed form and allow easy representation of complex shapes. The strategies proposed are compared, both visually and numerically, against conventional approaches, using a real dataset. The dataset was collected in a challenging underwater scene with a custom-designed multi-camera system. The results obtained surpass those of conventional approaches.

4

CLOSE-RANGE TRACKING OF UNDERWATER VEHICLES USING LIGHT BEACONS

IN this chapter, we focus on the application of omnidirectional cameras for aiding the navigation of underwater robots. We present a new tracking system for AUVs navigating in a close formation, based on computer vision and the use of active light markers. The proposed system allows the estimation of the pose of a target vehicle at short ranges, with high accuracy and fast execution. The use of an omnidirectional camera provides a full coverage of the lower hemisphere and enables the concurrent tracking of multiple vehicles in different positions. The system was evaluated in real sea conditions by tracking vehicles in mapping missions, where it demonstrated robust operation during extended periods of time.

The proposed method has been published in the following article:

Title: Close-Range Tracking of Underwater Vehicles Using Light Beacons
Authors: **J. Bosch**, N. Gracias, P. Ridaó, K. Istenič, and D. Ribas
Journal: Sensors
Volume: 16, Number: 4, Published: 2016
DOI: 10.3390/s16040429



Article

Close-Range Tracking of Underwater Vehicles Using Light Beacons

Josep Bosch *, Nuno Gracias, Pere Ridao, Klemen Istenič and David Ribas

Computer Vision and Robotics Group, Centre d'Investigació en Robòtica Submarina, Parc Científic i Tecnològic, Universitat de Girona, 17003 Girona, Spain; ngracias@eia.udg.edu (N.G.); pere@eia.udg.edu (P.R.); klemen.istenic@udg.edu (K.I.); dribas@udg.edu (D.R.)

* Correspondence: j.bosch@udg.edu; Tel.: +34-972-419-651

Academic Editor: Jaime Lloret Mauri

Received: 15 January 2016; Accepted: 18 March 2016; Published: 25 March 2016

Abstract: This paper presents a new tracking system for autonomous underwater vehicles (AUVs) navigating in a close formation, based on computer vision and the use of active light markers. While acoustic localization can be very effective from medium to long distances, it is not so advantageous in short distances when the safety of the vehicles requires higher accuracy and update rates. The proposed system allows the estimation of the pose of a target vehicle at short ranges, with high accuracy and execution speed. To extend the field of view, an omnidirectional camera is used. This camera provides a full coverage of the lower hemisphere and enables the concurrent tracking of multiple vehicles in different positions. The system was evaluated in real sea conditions by tracking vehicles in mapping missions, where it demonstrated robust operation during extended periods of time.

Keywords: tracking system; AUV; relative navigation; pose estimation; light beacons; active markers

1. Introduction

Oceanographic exploration and research are still today challenging tasks due to the demanding conditions underwater. The use of remotely-operated vehicles (ROV) and autonomous underwater vehicles (AUV), especially in deep-water operation, is essential for applications as varied as environmental surveying, geology, archeology, cable inspection and several others relating to industry and the military. However, the existing technology is still immature for close-range surveying of rugged terrain, such as caves, narrow passages or overhangs, due to limitations on the terrain sensing and on the navigation accuracy.

The use of a team of robots navigating in a close formation has the potential to significantly expand the coverage swath in mapping missions that require close proximity to the seafloor, such as optical or electromagnetic surveying. In areas of high topography, rigid arrays of sensors cannot be used safely, whereas AUV formations can provide the required degree of terrain compliance. The present work has been developed within the framework of the MORPH (Marine robotic system of self-organizing, logically linked physical nodes) EU-FP7 project (2012–2015) described in [1]. This project proposes a novel concept of an underwater robotic system that emerges out of using different mobile robot modules with distinct and complementary resources. These mobile robots navigate at a very close range as a group and have the ability to adapt the formation to changes in the terrain. The most relevant concept with respect to this paper is that an underwater vehicle equipped with a multibeam sonar profiler advances at the forefront of the formation, flying at a “safe” altitude from the sea-floor, while two other vehicles fly behind, very close to the bottom, acquiring images. As can be deduced, precise knowledge of the poses of all robots during the missions is fundamental for both safe navigation and an accurate reconstruction of the optical and acoustic maps. The relative localization

between vehicles is done through acoustic ranging. There has been space, though, for experimenting with the use of a vision-based method as an alternative for relative localization at short distances, where acoustics cannot provide updates with enough precision and frequency to ensure safety.

Under adequate visibility conditions, optical cameras can be very effective for computing precise position estimates, including full inter-vehicle poses. The effects of absorption and scattering often preclude the use of standard feature detectors [2] as a solution to the problem of vision-based formation sensing. To improve the chances of detecting point features and to identify individual vehicles, this paper proposes to endow the AUVs with light beacons, namely a set of active light markers blinking with distinctive patterns to facilitate their recognition. With this system, it is possible to track vehicles with full information about their relative pose with high accuracy and rapid update rates. In order to have a sensor with the widest possible field of view, an omnidirectional underwater camera was used to provide full vision of the lower hemisphere during the experiments (Figure 1).

This paper presents all of the aspects related to the system: the components and methodologies used, as well as the experiments performed and the results obtained.

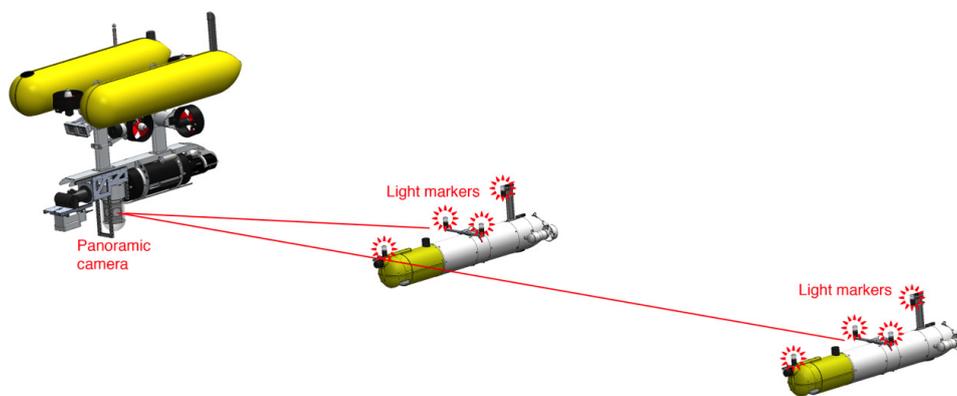


Figure 1. Representation of the localization of two target vehicles using active light markers. The leading vehicle, endowed with a wide field of view camera, localizes two target vehicles that are equipped with a set of active light markers.

1.1. Related Work

Navigation and localization are two of the most important topics for underwater robotics. While navigation in land and air robotics is mainly based on the use of GPS and inertial sensors, the inability to receive GPS updates underwater makes the task of navigating precisely more challenging [3]. Most AUVs rely on the use of inertial sensors combined with a Doppler velocity log (DVL) [4], an acoustic-based instrument that measures relative velocities with respect to the water or ground. However, this navigation technique is subject to drift over time. To avoid the unbounded growth of the navigation error in long missions, the system must restart the navigation periodically either by surfacing and receiving a GPS update [5] or by determining its relative position from an external reference point. It is at this point that acoustics are highly relevant. With acoustic ranging, it is possible to determine the relative position between an AUV and one or multiple beacons placed in a known underwater position. Between all existing acoustic technologies, the most widely used are long baseline systems (LBL) and ultra-short baseline systems (USBL) [6,7].

An LBL system [8,9] comprises two or more geo-located beacons, which are usually attached to the seafloor. Whenever these beacons receive an acoustic signal from an AUV, they reply to it after a short known delay. With the knowledge of the two-way time-of-flight time of the signals, the position of the beacons and the speed of sound, it is possible to precisely localize an underwater vehicle. An USBL system [7,10], instead, consists of a single acoustic beacon, which is localized by an

array of transceivers able to estimate both the range and angles of the incoming signal and, hence, the relative position of the beacon. This system can be used for tracking an AUV from a vessel (where the transceiver array is placed) or for improving the navigation system of the AUV, placing the array in the vehicle and the beacon in a geo-located position.

A solution to the underwater relative position measurement for multiple AUVs, was developed within the framework of the European GREX (Coordination and control of cooperating unmanned systems in uncertain environments) project (2006–2009) [11,12]. The navigation systems of each vehicle were combined with acoustic ranging from modems, to keep formation while following a predefined path. This achievement paved the way for underwater applications featuring multiple AUVs. Formation flying was limited to areas with approximately flat seafloors due to constraints on the vehicles pre-planned formation and to ensure reliable use of DVLs. Over rough terrain, DVLs tend to be highly inaccurate, and are of limited use for ensuring vehicle safety in tight formations. In the TRIDENT (Marine Robots and Dexterous Manipulation for Enabling Autonomous Underwater Multipurpose Intervention Missions) project (2010–2013) [13], a homing and docking solution using a USBL was tested. An intervention AUV (I-AUV) [14] was placed in charge of a survey and intervention mission, while an autonomous surface craft (ASC) was employed at the surface for communications purposes. Once the intervention mission was finished, the I-AUV (fitted with the USBL transceiver array) started a homing and docking procedure in order to dock in a special structure in the ASC, where the USBL beacon was located.

The use of easily-identifiable light sources for pose estimation has gained momentum in recent years in applications of land and aerial robotics. Recent examples are the work of Censi *et al.* [15] and Faessler *et al.* [16], where favorable visibility conditions allow the use of fast cameras and infrared LEDs to provide very fast pose updates. However, in underwater applications, where the detection and identification of the light sources is far more challenging, few attempts have been made.

Krupinski *et al.* presented [17] a docking panel equipped with active light markers as an alternative to acoustic localization for close ranges. Li *et al.* [18] applied this concept in a docking station for underwater vehicles. Four green LEDs were placed along a large funnel to make it visible to underwater vehicles using a stereo camera. Nevertheless, as there was no necessity of estimating the orientation nor the motion of the docking station, all of the green LEDs were permanently lit, and no identification was necessary. The fact that in our case the markers are placed on a mobile target makes it essential to distinguish and identify each one of the beacons to predict the motion accurately.

1.2. Contributions

The main contributions of this paper are:

1. A new method for pose sensing and relative navigation for multiple AUVs in short ranges based on the use of active light markers. This method has the following advantages in comparison with traditional systems:
 - (a) **High rate pose estimation:** The update rate depends on the frame rate of the camera and the capacity of the computer in charge of processing the images. It is expected then that the update rate can be higher in the near future with the rapid evolution of computer technologies.
 - (b) **High precision:** The minimum number of markers to retrieve the pose of a target vehicle is three. When using only three markers, the accuracy of the estimated pose depends strongly on the position of the markers on the vehicle and its location relative to the observer. The use of extra markers drastically reduces the uncertainty of the poses obtained, making the technique suitable for applications where very good accuracy is needed, such as cooperative underwater manipulation. A second source of uncertainty has to do with the location of the beacons in the image. The rapid evolution of underwater cameras in terms of resolution and sensitivity will lead to further improvements in the pose accuracy.

- (c) **Relative orientation data:** The most frequently-used acoustic localization systems, such as USBL or LBL, provide information about the relative position of a target, but they cannot provide information about its orientation. The light beacon system is able to provide this information with little uncertainty.
 - (d) **Low-cost:** Another handicap of acoustic-based systems is their cost. Acoustic modems and arrays are significantly more expensive when compared to optical cameras. The decreasing price of cameras makes the approach described in this paper extremely competitive against traditional systems.
2. Real experiments and results of the method presented: The system was tested at sea over several missions with results that support the advantages listed above. In the experiments presented in this paper, the filtered pose estimates were updated at approximately 16 Hz, with a standard deviation lower than 0.2 m in the distance uncertainty between vehicles, at distances between 6 and 12 m.

The rest of the paper is organized as follows. Section 2 describes the different components of the system. Section 3 presents the approach followed for the tracking process. In Section 4, the results of the experiments for testing the capacities and reliability of the whole system are presented. In the last Section 5, we present some conclusions.

2. System Description

The objective of the method is the real-time localization of underwater vehicles for distances less than 10 m and to obtain both position and orientation information with high update rates. The resulting system must also be robust to short temporal occlusions of the direct line of vision to the target markers.

The proposed solution consists of the placement of a set of light beacons, or active markers, on the target vehicles, which are optically tracked by a wide field of view camera placed in a camera vehicle. The tracking of these markers allows estimating the 3D pose of the target vehicles. Tracking of multiple target vehicles is possible by using different blinking pattern frequencies. The underlying assumptions are that the camera field of view covers the areas where the vehicles operate and that the visibility conditions are not severe for the intended inter-vehicle distances.

The light beacons and the camera system are the two main hardware components and are detailed in the following section.

2.1. Light Beacons

Each set of beacons consists of four markers connected through electric cables to a control board inside a watertight housing. This housing is placed in the payload area of the vehicles and is powered by the batteries of the robot. This setup makes it easy to install the markers on different vehicles and to distribute them in different geometries according to the vehicle design (Figure 2b). It is essential to have a precise measurement of the location of the markers with respect to the navigation origin of the vehicle for correct operation of the system. Each individual light beacon consists of five high-intensity LEDs oriented strategically to create a homogeneous omnidirectional lighting effect inside a cylindrical waterproof housing (Figure 2a). The system is operated at 24 V, and the maximum power consumption is 22 W, when all markers are lit.

The number of beacons used and how they are placed on a target vehicle are two factors that directly influence the precision of the estimated pose. A comprehensive study of such factors can be found in [19].

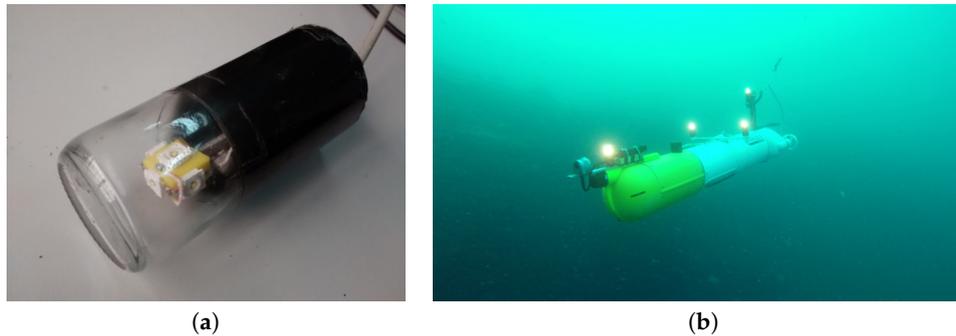


Figure 2. Detail of a light beacon and demonstration of a light beacon set assembly in a real AUV. They can be easily installed and adapted to any vehicle geometry. (a) Details of an active marker. Each one consists of five high-intensity LEDs; (b) The Sparus II AUV [20] equipped with a set of four active markers during a mission.

Various possibilities of differentiating each individual marker of the set to allow its identification were studied. One of the most popular and simple techniques used in land robotics is the use of colors to distinguish the different beacons composing the set, but this option was discarded, due to the difficulty in consistently discerning colors for light sources at distances larger than a few meters. The use of different blinking frequencies for each individual marker was discarded due to the use of this strategy for identifying different sets of light markers, and hence, being able to estimate the pose of different vehicles simultaneously. Instead, different blinking patterns, illustrated in Figure 3, are used to allow the identification of the different markers in each set.

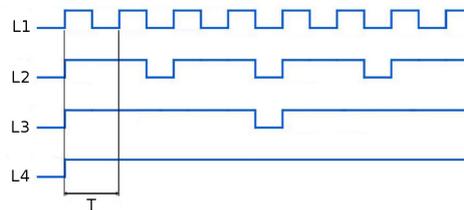


Figure 3. Blinking pattern of each marker of the light beacon set. Each beacon has a distinctive pattern allowing the global identification in the minimum time possible while there are at least three beacons lit more than 50% of the time.

Two objectives were considered in designing the patterns. On the one hand, it should allow the identification of all lights in the minimum time possible, and on the other, it should maximize the time when all lights are on, thus allowing the pose estimation algorithm to compute a large number of poses per second, which facilitates tracking.

Two different sets of active markers with different minimum cycle periods, T , were manufactured to allow the simultaneous tracking of two target vehicles on the same mission. These periods must be chosen according to the camera frame rate, f_{ps} , and the design requirements, as they implicitly define the minimum time necessary for the identification of the light beacons. The minimum cycle period, T , contains for the fastest marker (L1) one *on* and one *off* period, and we need to be sure that the camera captures at least one frame and preferably two to improve robustness: one where the beacon is lit and another where the beacon is off. Thus, $\frac{T}{2} > \frac{1}{f_{ps}}$.

For the experimental setup used in this paper, the camera has a frame rate of 16 fps, and so, $T > 0.125$ s. The values used for the first and second sets of light beacons were $T_1 = 0.25$ s, and $T_2 = 0.7$ s, respectively.

The minimum time needed for the identification of n markers, $T_{ident}(n)$, is proportional to the minimum cycle period. For the pattern presented in Figure 3, this time is $T_{ident}(n) = 2^{n-1} T$. Thus, the minimum time necessary for the identification of three of the markers, $T_{ident}(3)$, which is the minimum necessary to compute a first pose estimation, is: $T_{ident}(3) = 4 T$.

2.2. Panoramic Camera

The camera used for the localization must have a wide field of view (FOV) in order to keep track of the target vehicle in a wide range of positions and orientations of both the target and the leading vehicle. For this reason, in the experiments presented in this paper, a panoramic (or omnidirectional) camera was used instead of a conventional one.

2.2.1. Model and Assembly

The camera used is an omnidirectional multi-camera system (OMS), based on a Point Grey's Ladybug 3 [21]. The Ladybug 3 comprises six individual cameras and is designed for land-based applications. A custom housing was designed to make it submersible up to 60 m (Figure 4a). The housing is composed of a transparent poly-methyl methacrylate (PMMA) dome, which contains the camera, and an aluminum alloy body, which contains a small form factor computer dedicated to processing the video feed. The computer is connected directly to the Ladybug 3 through a FireWire 800 Mbps connection. The housing has a single external Ethernet cable used for both power and communications.

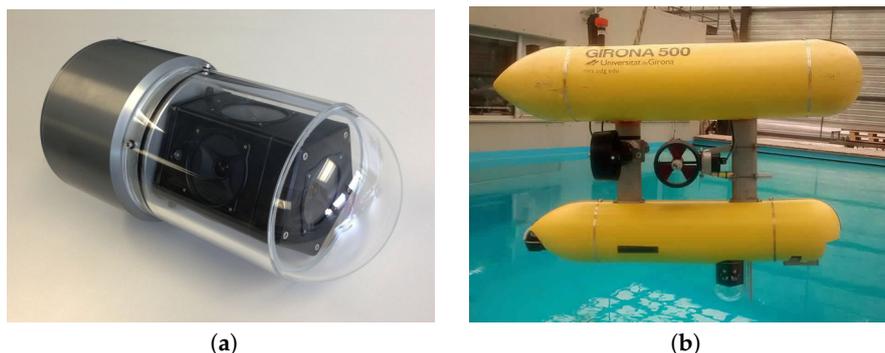


Figure 4. Omnidirectional camera and assembly in Girona500 AUV. The camera was installed under the bottom cylinder to have a clear view of the bottom hemisphere and is protected by two aluminum bars installed in front of the camera. (a) Omnidirectional underwater camera based on a Point Grey's Ladybug camera and a custom waterproof housing; (b) the Girona500 AUV equipped with the omnidirectional camera.

For the experiments presented in Section 4, the omnidirectional camera was mounted in the bottom part of Girona500 AUV [22], as shown in Figure 4b. To protect the camera from any damage in the unlikely event of a collision, two aluminum bars were placed in front of the camera. These bars have no impact on the performance of the tracking system, as the target vehicles were always behind Girona500 in the formation employed during the missions. The Girona500 is able to power up or down the camera through a digital output and communicates through Ethernet with the computer embedded inside the housing, which provides the estimated poses of the target vehicles.

2.2.2. Camera Calibration

The camera outputs six separate images that can later be combined to create a hemispherical panorama or treated separately as individual images according to the mission objectives. In both cases, the camera must first be calibrated to ensure proper use of the images collected. The calibration takes

into account all of the distortions introduced by both the lenses of the camera and the waterproof housing, as well as the relative positioning between the individual cameras. The calibration of such a complex camera was divided into three different steps: intrinsic, extrinsic and underwater calibration.

The intrinsic parameters of all single cameras are necessary to project a 3D point in space onto the 2D image plane. They depend on specific geometry properties of each camera, as well as lens properties, such as focal length (f_l), principal point (c) or distortion coefficients. The pinhole camera is the most used camera model due to its compactness and freedom from distortions [23]. However, all lenses introduce image distortions that are not explicitly included by this model. The most common one is radial distortion, which is due mainly to the shape of the lenses and produces nonlinear distortions along the radial direction from the principal point. The calibration of the intrinsic parameters is done separately for each single camera in air and without the waterproof housing, making use of a standard calibration toolbox. It is important to note that due to the high distortion of the lenses used in the Ladybug3 camera, a fisheye distortion model was used to properly correct the radial distortion.

The calibration of the extrinsic parameters consists of the determination of the exact geometric relationship between the different camera frames composing the OMS. For this calibration, a specific procedure was developed. This procedure was based on a bundle adjustment of multiple features observed from different images, similar to the calibration of a stereo camera. The data necessary for this calibration were collected in air and without the waterproof housing.

The underwater calibration consisted of determining the exact position and orientation of the waterproof housing with respect to the camera. It is worth noticing that the direction of the rays changes at every medium transition found along the path from the imaging sensor inside the camera to a point underwater (Figure 5). A small error in the relative position of the housing can lead to a big inaccuracy in the direction of the final ray.

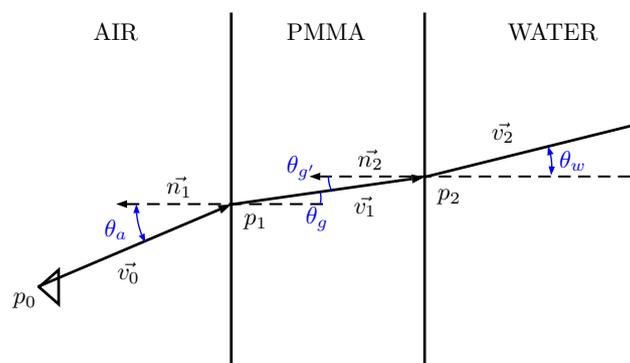


Figure 5. Ray tracing schematic of a single optical ray passing through air, PMMA and water.

Once the multi-camera system has been calibrated, it is possible to obtain the projection function f , which projects a 3D point into a 2D location in the image sensor of a chosen camera, and its inverse f^{-1} , which projects a 2D pixel of an image sensor onto a 3D ray.

For a conventional pinhole camera, the function f^{-1} for projecting a 2D pixel, $u = [u_x, u_y]$, from a non-distorted image onto a 3D ray with center at the origin of the camera and direction vector $v = [v_x, v_y, v_z]$, is straightforward:

$$v = \left[\frac{u_x - c_x}{f_l}, \frac{u_y - c_y}{f_l}, 1 \right] \quad (1)$$

where $c = [c_x, c_y]$ is the location of the principal point in the imaging sensor and f_l is the focal length.

For underwater cameras, the fact that the direction of the ray changes in every medium transition makes the process more laborious, as for each transition, the intersection point and the direction of the

rays must be computed according to the laws of physics. For the sake of simplicity, the details are not described here, but can be found in [24].

For the case of the projection function f that projects a 3D point $p = [p_x, p_y, p_z]$ into a 2D location in the image sensor, it is equally simply to find an expression for pinhole cameras:

$$u = \left[f_l \frac{p_x}{p_z} + c_x, f_l \frac{p_y}{p_z} + c_y \right] \quad (2)$$

In contrast, it is not possible to find such an expression for projecting an underwater 3D point. To solve this problem, an iterative process is run instead. This process goes along the pixels of the sensor and selects the one whose associated 3D ray passes closer to the desired 3D point.

3. Approach

The tracking of the target vehicle is divided into two stages (Figure 6). The first stage consists of an initialization step, where the pose of the vehicle is unknown and there is not enough information available for its estimation. The second stage begins when there is enough information for estimating the pose of the target vehicle, and it lasts until the tracking of the vehicle is lost, where the system returns to the initial stage.

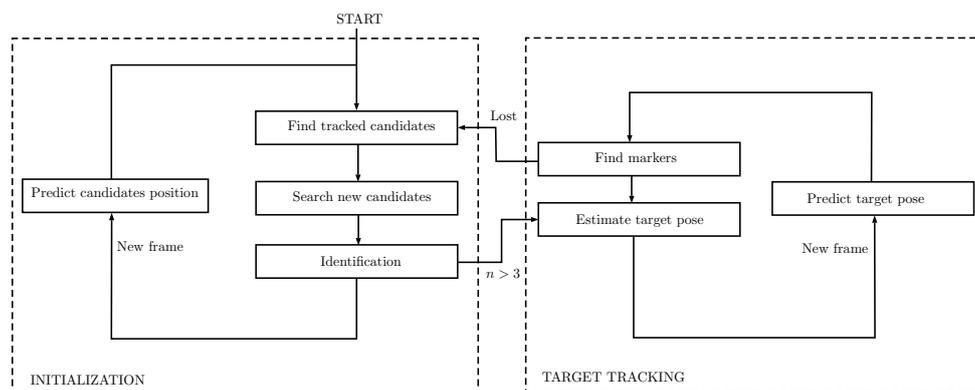


Figure 6. Block diagram of the light beacon tracking system.

During the initialization stage, the tracking consists of an independent tracking process for each of the lights. In contrast, the second stage, named *target tracking*, consists of the global tracking of the target vehicle.

All software programs have been implemented in C++ to achieve the best temporal performance possible and to make the tracking system able to work in real time. Different programming libraries have been used for the implementation of the full system, with special relevance on OpenCV [25] for the treatment of digital images and Ceres-Solver [26] for solving the non-linear least squares problems.

3.1. Initialization

During the initialization stage, there are three main tasks: (1) searching for new light candidates; (2) tracking previous candidates; and (3) deciding if they correspond to one of the beacons. When at least three candidate lights have been identified, the system moves to the second stage. During this stage, we will make use of acoustic ranges as extra information for making the prediction of the lights in future frames more precise, but this could be replaced by the assumed distance between the camera and the target vehicle for each mission.

3.1.1.1. New Candidates

The system starts the identification process over the bright spots of the image. With this purpose, the gradient image is computed from the grayscale image using the Sobel operator. The Sobel operator applies two 3×3 convolution kernels to obtain two images that contain an approximation to the horizontal, G_x , and vertical, G_y , derivative images:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A \quad (3)$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * A \quad (4)$$

where $*$ denotes the 2D convolution operation and A is the grayscale image.

The two derivative images are combined in order to obtain a gradient image:

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (5)$$

A mask can be applied to the resulting image with the aim of avoiding the further inspection of bright spots directly related to the body of the vehicle carrying the camera (Figure 7).

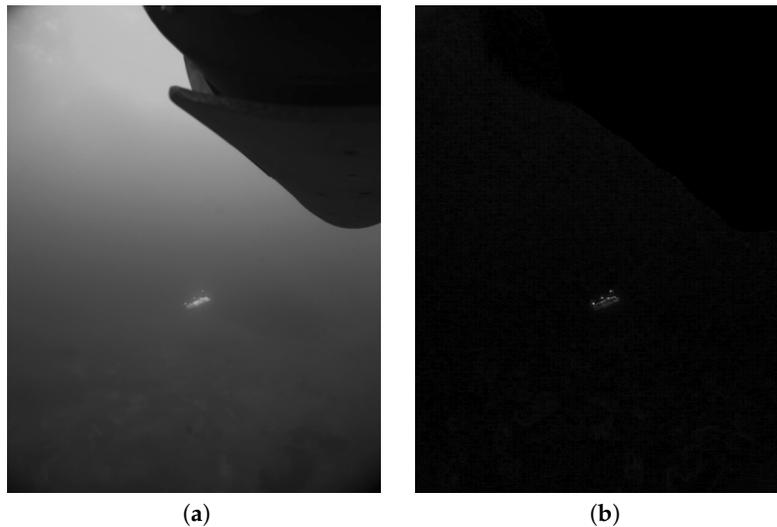


Figure 7. Sample image and its corresponding gradient image. (a) Original grayscale image; (b) gradient image with a mask for the part corresponding to the body of the AUV carrying the camera.

The next step consists of selecting from the gradient image the n brightest spots with a minimum distance d between them to analyse them in detail. For each one, we select a window in the original grayscale image, with size w centered on the point found previously, and we search for the local maximum closest to the center of the window, as the location may vary from the gradient image. Once the window has been re-centered and we are sure the candidate spot is in the center, we will check the following different conditions before accepting it definitely as a candidate light.

- **Intensity:** A minimum intensity value is required to accept a bright spot as a candidate light. This minimum depends on the existence of candidates in the previously-processed images. In the

case that no previous candidates exist, the value depends both on the last acoustic range received and on an extra parameter reflecting the visibility (and sun conditions in case the mission is performed in shallow waters). In cases where there were candidates present in previous frames, the value of intensity required is slightly smaller than the minimum intensity of the existing candidates.

- **Size and shape:** For accepting the pre-candidate, its shape must be similar to a 2D-Gaussian distribution, and its area cannot be greater or lower than certain selected values. The first step required is determining the size of the spot analyzed. For this purpose, a technique very similar to the radial contrast function (RcF) method is used [27]. This algorithm was developed for source detection in astronomical images, but is flexible enough to be applicable to the images processed by our system. It operates by choosing the brightest pixel and analyzing the mean of the neighbor pixels at incrementally larger radial distances. The size of the light is determined when the intensity profile obtained stops decreasing and remains constant. Once the size is determined, we must ensure that its value is reasonable. As in the previous case, the minimum and maximum size values depend on the existence of candidates in the previous frames. In the positive case, the minimum and maximum values are determined from the features of the existing candidates, while in the other case, they depend on the last distance estimate available and the visibility conditions.

3.1.2. Tracking of Candidates

Every time the system analyzes a new image, we look for lights that could correspond to previous candidates. To estimate the position of a previous candidate in a new image, we consider that the lights are still in the 3D space, and we take into account only the movement of the camera. It is important to note that any small rotation of the camera results in a significant displacement in the image; thus, assuming that the candidate lights are still in the image plane would result in a less efficient tracking.

The distance between the camera and the marker must be assumed in order to project the 2D marker position from the last image frame, ${}^{k-1}u = [u_x \ u_y]$, to a 3D point, ${}^{k-1}p$. In our case, that distance was assumed to be the last acoustic range, r , between the vehicle carrying the camera and the target vehicle,

$${}^{k-1}p = f^{-1}(\text{cam}, r, {}^{k-1}u) \quad (6)$$

where f is the projection equation according to the camera calibration, as detailed in Section 2.2.2 and cam is the number of cameras.

Once the 3D point corresponding to the last frame, $k - 1$, has been computed, it can be rotated and translated according to the transformation matrix, ${}^kT_{k-1}$, which transforms a point from the $k - 1$ coordinate system to the k coordinate system and is computed according to the navigation system of the camera vehicle:

$${}^k p = {}^k T_{k-1} \cdot {}^{k-1} p \quad (7)$$

Finally, the rotated 3D point can be projected back to the image plane through the iterative method described in Section 2.2.2 to obtain the predicted position of the marker in the new image, ${}^k u$:

$${}^k u = f(\text{cam}, r, {}^k p) \quad (8)$$

In the case that the predicted candidate position is close to the limit of the image, an image from one of the adjacent cameras will be selected accordingly before any further processing.

For each candidate light, we select a patch around its predicted position. The size of the patch, w , depends on the previous light size and the assumed distance r . In the patch, we search for the closest local maximum, and we again use the RcF technique described in the previous subsection. We check if there exist similarities between the candidate light and the local maximum found. Particularly, they must have similar intensity, size and intensity gradient, or otherwise, it is assumed that the maximum found does not correspond to the tracked light. Depending on the results obtained, we tag

the candidate light as *on* or *off*. In the case that a candidate light has been tagged as *off* for a number of frames that is significantly larger than the expected number according to the beacons' pattern, we remove it from the candidates list.

3.1.3. Beacon Identification

After the tracking of the candidate lights, an identification method checks if any of the candidates can be associated with the beacons. The method described in this section has been proven very effective for the pattern described in Section 2, but may need modifications in the case of a different blinking pattern.

Each one of the tracked lights has a record of its full *on-off* state history. Especially important are the mean of the periods when the light was *off* (not detected) and the mean of the periods when the light was *on* (detected). A score matrix is computed to evaluate every possible association, containing as many rows as candidates and as many columns as markers. The matrix is initialized with a negative value for all cells, and two conditions must be met for computing a score value.

- An *off period* must have been detected for the candidate light, that is a light cannot be associated with a marker, if it has not been absent for at least one frame and detected again. Furthermore, the duration of this *off period* must be very close to the expected cycle time within a tolerance t .
- Additionally, for comparing a candidate light i with a marker j , the light must have been tracked for at least the duration of the marker full period; otherwise, we could not be sure the association is correct:

$$\text{light tracking duration}(i) > \text{marker period}(j) \quad (9)$$

If these two conditions are met, a score is computed for every possible association using:

$$\text{score}(i, j) = 1 - \frac{\text{expected time on}(j) - \text{mean time on}(i)}{\text{expected time on}(j)} \quad (10)$$

Once all of the cell values are computed, we find the maximum value of the matrix. In the case that this value is greater than a certain threshold, 0.8 for the results presented in this paper, we identify the association as valid. The column and the row where the maximum occurs are removed, and a maximum is searched for again. If this maximum is greater than the threshold, the association is identified, and the corresponding row and column are removed. This process is repeated until all values of the matrix are lower than the threshold.

In the case that we have identified at least three lights, it is then possible to estimate the pose of the target vehicle. If the system was in the first stage of processing, it moves to the second one.

3.2. Target Tracking

Once a first pose estimation of the target vehicle has been computed, the system starts a tracking process over the target vehicle. The procedure is as follows:

1. Each time there is a new image acquisition, the elapsed time between the previous processed image and the new one is computed. A prediction of the movement of the target vehicle S with respect to the camera in the elapsed time is computed, taking into account both the motion of the camera and the dynamics of the target vehicle.
2. According to the predicted 3D pose of the target vehicle and the known position of the active markers along the vehicle's body, each one of the markers is projected onto the image plane to obtain its 2D predicted position through the use of the projection equation (Equation (8)).
3. Each one of the markers is searched in the images according to its predicted 2D location using an identical process to the one described in Section 3.1.2. If at least three markers are detected, a new pose estimation is computed. Otherwise, the predicted pose is assumed to be the real one.

In order to reduce the noise in the estimated poses and obtain the smoothest possible dynamic model of the target vehicle, we make use of an extended Kalman filter (EKF). It has been found possible to reduce the noise for each one of the estimates and, thus, to produce a better result.

Details of the implementation of the EKF and the pose estimation are presented in the following subsections.

3.2.1. Temporal Filtering

The use of an EKF filter proved very useful to reduce the noise of the estimated 3D poses of the target vehicle. At the same time, it allowed a better prediction of the 2D position of the markers in the images and significantly improves the performance of the system.

State vector:

The EKF state vector has two different parts, $x_k = [p \ v_s]^T$. The term p contains the six degrees of freedom defining the current position, p_1 , and orientation, p_2 , of the target vehicle S represented in the camera frame C at time k (see Figure 8):

$$p = [p_1^T \ p_2^T]^T = [x \ y \ z \ \phi \ \theta \ \psi]^T \quad (11)$$

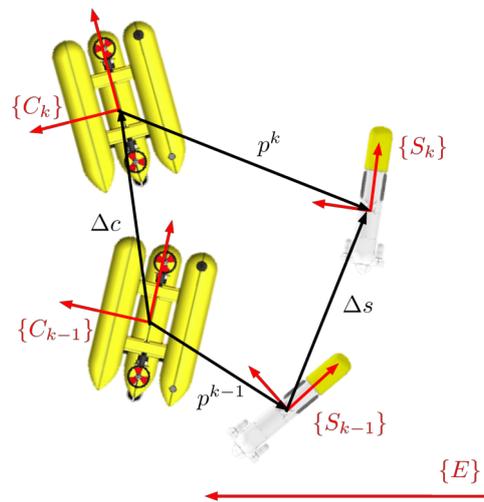


Figure 8. Representation of the target vehicle (S) and the camera vehicle (C) at instants $k-1$ and k .

The term v_s contains the six degrees of freedom defining the linear, $v_{1,s}$, and angular, $v_{2,s}$, velocities of the target vehicle with respect to the inertial frame E represented in the tracked vehicle frame S_k at time k :

$$v_s = [v_{1,s}^T \ v_{2,s}^T]^T \quad (12)$$

Prediction:

Our model is governed by a non-linear function f :

$$x_k = f(x_{k-1}, u_k, n_k) \quad (13)$$

which relates the state at a time k , x_k , given the state at a time $k-1$, x_{k-1} , a control input u_k and a non-additive noise $n_k = [n_1^T \ n_2^T]^T$ that follows a Gaussian distribution with zero mean and covariance Q_k .

According to the notation used in Figure 8, and assuming that the target vehicle follows a constant velocity model, f can be expressed as:

$$x_k = \begin{bmatrix} p^k \\ v_s^k \end{bmatrix} = \begin{bmatrix} (\ominus \Delta c) \oplus^A \left(p^{k-1} \oplus^B \Delta s \right) \\ v_s^{k-1} + n_2 \Delta t \end{bmatrix} \quad (14)$$

where operators \oplus and \ominus denote the commonly-used six degrees of freedom inversion and compounding operations [28], the term Δt denotes the time elapsed between time $k-1$ and k , the term Δc denotes the variation of the pose of the camera vehicle in the elapsed time Δt and is part of the control input u_k , the term Δs corresponds to the variation of the pose of the target vehicle in the camera frame, C_k , and can be computed as:

$$\Delta s = \begin{bmatrix} v_{1,s}^{k-1} \Delta t + \frac{1}{2} n_1 \Delta t^2 \\ J_\omega(p_2^{k-1}) \left(v_{2,s}^{k-1} + n_2 \Delta t - R^T(p_2^{k-1}) v_{2,g}^{k-1} \right) \Delta t \end{bmatrix} \quad (15)$$

where $v_{2,g}^{k-1}$ is the angular velocity of the camera vehicle at the instant $k-1$ and is part of the control input u_k , $J_\omega(p_2^{k-1})$ is the Jacobian that transforms the angular velocity of the target vehicle (S) with respect to camera vehicle (C) to $p_2^k = [\phi \ \theta \ \psi]^T$ and is given by:

$$J_\omega(\phi, \theta, \psi) = \begin{bmatrix} 1 & \sin \phi \tan \theta & \cos \phi \tan \theta \\ 0 & \cos \phi & -\sin \phi \\ 0 & \frac{\sin \phi}{\cos \theta} & \frac{\cos \phi}{\cos \theta} \end{bmatrix} \quad (16)$$

and $R^T(p_2^{k-1})$ is the rotation matrix that transforms a point expressed in the S coordinate system to the G coordinate system which depends on their relative attitude p_2^{k-1} .

The prediction of the state \hat{x}_k^- and its associated covariance P_k^- are given by:

$$\begin{aligned} \hat{x}_k^- &= f(\hat{x}_{k-1}, u_k, n_k = 0) \\ P_k^- &= A_k P_{k-1} A_k^T + W_k Q_k W_k^T \end{aligned} \quad (17)$$

where:

$$A_k = \frac{\partial f(x_{k-1}, u_k, n_k)}{\partial x_{k-1}} \Big|_{(\hat{x}_{k-1}, u_k, n_k=0)} \quad (18)$$

and

$$W_k = \frac{\partial f(x_{k-1}, u_k, n_k)}{\partial w_k} \Big|_{(\hat{x}_{k-1}, u_k, n_k=0)} \quad (19)$$

According to Equations (14) and (15), the Jacobians A_k , W_k are:

$$A_k = \begin{bmatrix} J_{2\oplus}^A \left(J_{1\oplus}^B + J_{2\oplus}^B \begin{bmatrix} 0_3 & 0_3 \\ 0_3 & J^* \end{bmatrix} \right) & J_{2\oplus}^A J_{2\oplus}^B \begin{bmatrix} I_3 \Delta t & 0_3 \\ 0_3 & J_\omega \Delta t \end{bmatrix} \\ 0_6 & I_6 \end{bmatrix} \quad (20)$$

$$W_k = \begin{bmatrix} J_{2\oplus}^A J_{2\oplus}^B \begin{bmatrix} \frac{1}{2} \Delta t^2 I_3 & 0_3 \\ 0_3 & J_\omega(p_2^{k-1}) \Delta t^2 \end{bmatrix} \\ I_6 \cdot \Delta t \end{bmatrix} \quad (21)$$

where J^* can be obtained by symbolic differentiation of the following expression:

$$J^* = \frac{\partial \left(J_\omega(p_2^{k-1}) \left(v_{2,s}^{k-1} - R^T(p_2^{k-1}) v_{2,g}^{k-1} \right) \Delta t \right)}{\partial (p_2^{k-1})} \quad (22)$$

Measurement model:

The general model for measurements is:

$$z_k = H_k x_k + m_k \quad (23)$$

where z_k is the measurement vector and corresponds to the relative pose of the target vehicle with respect to the camera:

$$z_k = \begin{bmatrix} x_s^c & y_s^c & z_s^c & \phi_s^c & \theta_s^c & \psi_s^c \end{bmatrix}^T \quad (24)$$

H_k is the observation matrix and has the form:

$$H_k = \begin{bmatrix} I_6 & 0_6 \end{bmatrix} \quad (25)$$

the term m_k is a vector of white Gaussian noise with zero mean and covariance $R_k = \Sigma_{z_k}$. This covariance is provided by the pose estimation module (Section 3.2.2).

Update:

Updates happen when a new measurement is computed from the pose estimation module. Although the tracking system has been proven to be reliable, there are still situations where a misidentification or mismatching could cause the wrong calculation of the pose estimation and, thus, negatively affect the pose estimate of the filter. To avoid taking into account outlying observations, we check if the observation is consistent with the current prediction computing the innovation term, r_k , and its covariance, S_k :

$$r_k = z_k - H_k \hat{x}_k^- \quad (26)$$

$$S_k = H_k P_k^- H_k^T + R_k \quad (27)$$

Then, in order to determine the compatibility of the measurement, an individual compatibility (IC) test is performed. With this aim, the Mahalanobis distance [29] is computed as:

$$D_k^2 = r_k^T S_k^{-1} r_k \quad (28)$$

The observation is considered acceptable if $D_k^2 < \chi_{d,\alpha}^2$, where the threshold $\chi_{d,\alpha}^2$ is given by the chi-squared distribution with d degrees of freedom and a cumulative value of $1 - \alpha$. For the implementation in this paper, values of $d = 6$ and $\alpha = 0.05$ were used.

Since the measurement model is linear, the classical KF equations can be used for the update; where the Kalman gain, K_k , is computed as:

$$K_k = P_k^- H_k^T S_k^{-1} \quad (29)$$

and the estimate of the current state \hat{x}_k and its covariance P_k according to the Joseph form are:

$$\hat{x}_k = \hat{x}_k^- + K_k r_k \quad (30)$$

$$P_k = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \quad (31)$$

3.2.2. Pose Estimation

The pose x^* that best fits the observation of the markers in the image, u , is found using non-linear least squares minimization, by searching for the values of the variable x that minimize the sum of $f(x)$:

$$\min_x \frac{1}{2} \sum_i \|f_i(x_{i_1}, \dots, x_{i_k})\|^2 \quad (32)$$

The variable x contains the complete pose of the target vehicle with respect to the camera $p = [x_s^c \ y_s^c \ z_s^c \ \phi_s^c \ \theta_s^c \ \psi_s^c]^T$. The function f computes the re-projection error for each one of the markers; that is, the difference between the real observation and the projection of the marker according to the variable x and the calibration parameters of the camera. The problem is solved with a Levenberg-Marquardt [30] iterative algorithm available in the Ceres library [26]. As with all iterative methods, it needs an initial guess of the variables, which in our case is the predicted relative pose according to the EKF filter described previously.

In addition to the pose estimate, it is essential to have an estimate of the associated uncertainty, so that the pose information can be adequately used in a navigation filter. A first-order approximation of the pose covariance Σ_{x^*} can be computed from the assumed covariance Σ_u of the pixel location of the beacons in the image and the Jacobian $J(x^*) = \frac{\partial u}{\partial x}(x^*)$ that relates small changes in the pose parameter with small changes in the observations. The Levenberg-Marquardt implementation used computes and provides this Jacobian at the end of the minimization. The pose covariance estimate is given by:

$$\Sigma_{x^*} = (J(x^*)^T \Sigma_u^{-1} J(x^*))^{-1} \quad (33)$$

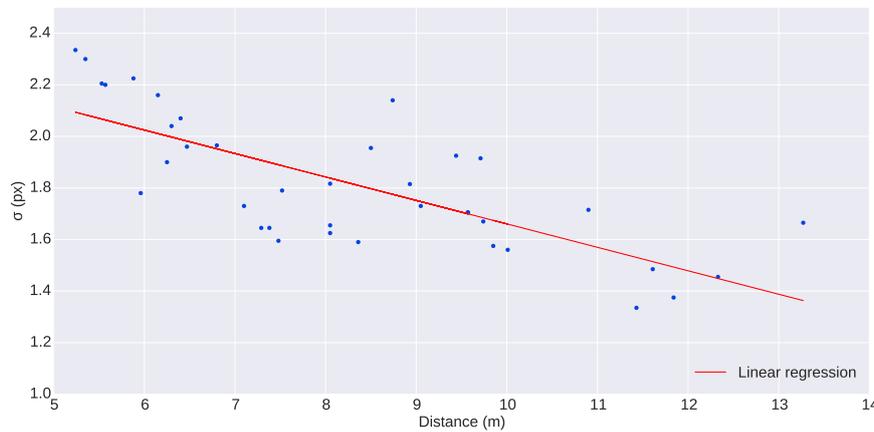


Figure 9. Evolution of the standard deviation of the Gaussian distributions fitted to the lights along distance.

The uncertainty in the localization of the lights in the image is inversely dependent on the distance of the beacon from the camera. The closer the beacon is to the camera, the larger the projected light disk will be in the image, thus leading to higher location uncertainty than far away beacons that appear in the image as small disks. In order to have an approximate value of this uncertainty, the size of the lights was analyzed from a set of selected images of the experiment presented in this paper. The beacons in the images were fitted to a 2D Gaussian distribution centered at u , with standard deviation σ , amplitude A and an offset c_0 : $f(u, \sigma, A, c_0)$. The standard deviation of the 2D Gaussian distributions, σ , found can be used as an indicative value of the uncertainty of the location procedure. The experimental evidence from the mission presented in this paper (Figure 9) showed that the variation of σ is small enough to be considered constant within the range of distances of the experiments (5 to 12 m). A conservative mean value of $\sigma = 2$ pixels was chosen for the standard deviation of both horizontal and vertical pixel uncertainties. The covariance used was:

$$\Sigma_u = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad (34)$$

4. Results

The performance of the light beacon tracking and pose estimation method was tested in several trials during the MORPH EU FP7-Project. The most relevant field testing took place in Sant Feliu de Guíxols, Spain, in March 2015 and in Horta, Azores Islands, in September 2015. This section presents the results obtained in one of the longest and most successful missions (Azores 15). The mission was performed by a total of five vehicles (four submerged and one at the surface) with the objective of mapping a vertical wall and the sea floor next to it. The Girona500 (G500) carried the omnidirectional camera and performed the role of the leading vehicle, while the Sparus II and Seacat AUVs [31] were the optical mapping vehicles, in charge of capturing high resolution imagery of the seabed (Figure 10).

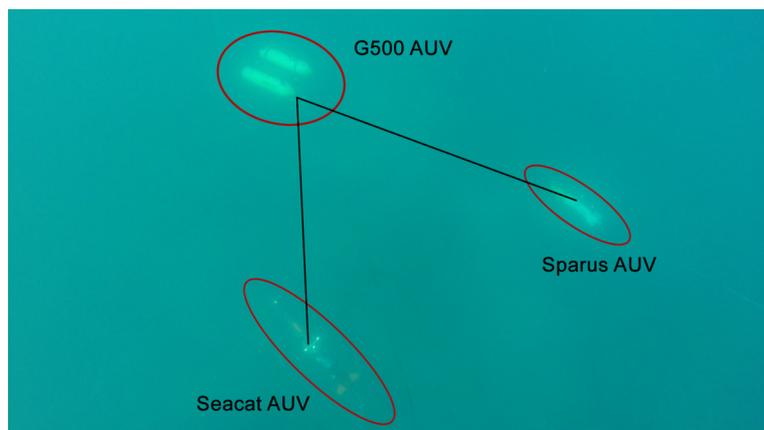


Figure 10. Image of G500, Sparus II and Seacat AUVs captured by a diver during a mission in Horta, Azores Islands, September 2015.

Precise navigation data for all vehicles are fundamental for a good reconstruction of the multi-vehicle optical mosaic [32]. The formation control was performed with acoustic ranging following the MORPH guidelines [33,34]. The light beacons were used as an experimental technology being field tested at the time. As such, the tracking was not used in the formation control loop, but served as an important tool for online mission monitoring and to assist the data post-processing.

Figure 11 illustrates the optical tracking results. The system capabilities allow the reconstruction of the scene in a 3D viewer with a high update rate of both the position and orientation of the target vehicles, making it very easy for an operator to understand the development of the mission in real time or during replay.

The intended positions of the target vehicles with respect to the leading vehicle varied considerably during the mission. During most of the survey, the target vehicles were surveying the horizontal floor, and the desired positions were 5 m behind the G500 AUV and 5 m to both the left and right sides. However, when surveying the vertical wall, the two vehicles were commanded to move to the same side of the G500 and explore the wall with different altitudes (Figure 12). For this reason, along with the normal oscillation of the relative positioning due to the control system, the distance between the G500 and the target vehicles was not constant and varied between 4 and 18 m. The performance of the optical tracking system depends strongly on the distance between the camera and the target vehicles, the geometry of the relative positions of the projections of the light beacons on the image [19] and on the visibility conditions in the water, principally the turbidity and the presence of strong veiling light. The complete characterization of the environmental conditions under which the system can operate is undergoing assessment and is outside the scope of this paper. However, the tests reported in this paper were done in realistic conditions of visibility at sea. Although not measured precisely, the visibility was estimated by local divers to be in the 20–25 m range.

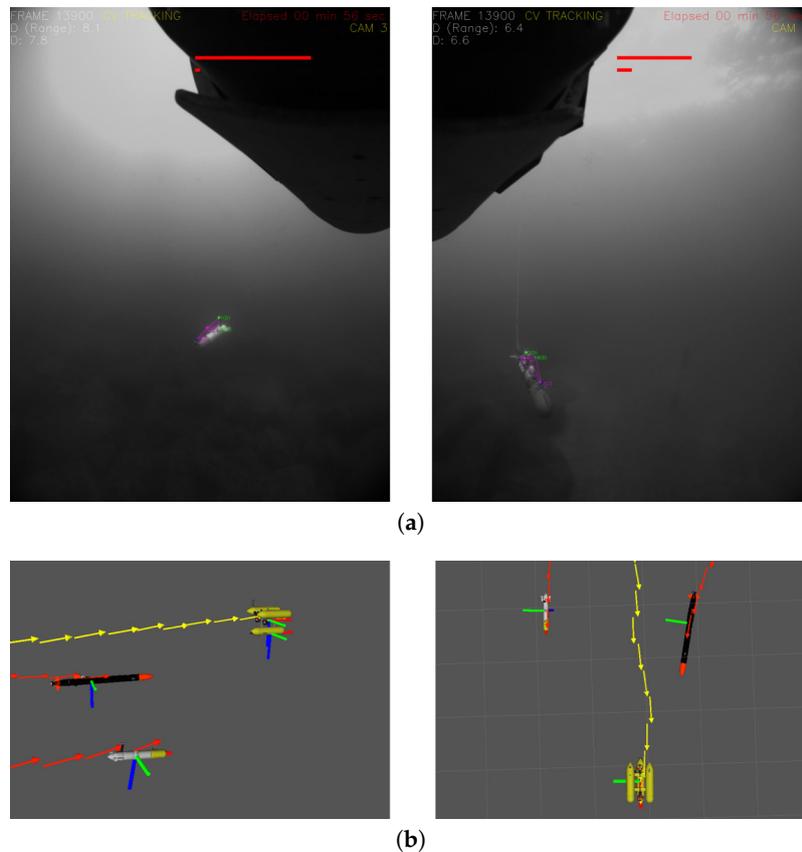


Figure 11. Snapshots of the system while tracking simultaneously two target vehicles. (a) Processed images from two independent cameras while the system is tracking two different target vehicles: Sparus (left) and Seacat (right); (b) left: lateral view of the reconstructed scene in the Rviz simulator; right: top view of the reconstructed scene in the Rviz simulator.

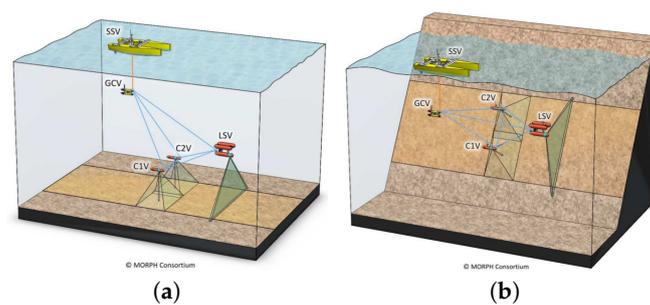


Figure 12. During the MORPH (Marine robotic system of self-organizing, logically linked physical nodes) project missions, the vehicles have the capability to change the formation geometry on-the-fly, adapting it to the changes in the terrain. When the formation navigates towards a wall, the two target vehicles, named in the figures C1V and C2V, change their altitude, which can result in one, or both, target vehicle(s) being out of the field of view of the omnidirectional camera and, therefore, without tracking capabilities. Source: MORPH Consortium. (a) Survey of a flat region. The two target vehicles navigate close to the bottom, below the leading vehicle; (b) Survey of a vertical wall. The two target vehicles have different altitudes, one or both of them above the leading vehicle.

Figure 13 presents a top or planar view of the trajectory of the G500 during the mission. The figures on the left are color-coded with the distances between the G500 and the target vehicles, while the figures on the right show the intervals of the sustained operation of the optical tracking. The trajectories and inter-vehicle distances were computed from offline optimization using all navigation data available [32]. During some parts of the mission, especially during the wall survey, the target vehicles were flying above the leading vehicle and, thus, outside the field of view of the camera. On such occasions, it was not possible to perform the tracking, even when the relative distance between the vehicles was small. Figure 14 illustrates this issue: the top plot shows the distance between the G500 and the target vehicles similarly to Figure 13; the middle plot presents the relative depth between the vehicles; the bottom plot highlights the intervals of sustained operation of the optical tracking. An initial insight into the performance of the optical tracking can be inferred from Figure 14. The plots show that under the conditions of the experiment, the tracking system performs adequately when the target vehicles navigate below the camera vehicle at distances less than 10 m, with intermittent operation for distances between 10 and 15 m. For distances of more than 15 m, the operation is unreliable. The mission was performed over a 40-min period at noon, under cycling illumination conditions of overcast cloud and direct strong sunlight. The G500 AUV was navigating at a depth between 7 and 10 m, where the Sun still causes reflections on the body of the vehicles and decreases the visibility of the light beacons. Although not tested, it is safe to assume that performance would improve even further at deeper depths or night-time conditions.

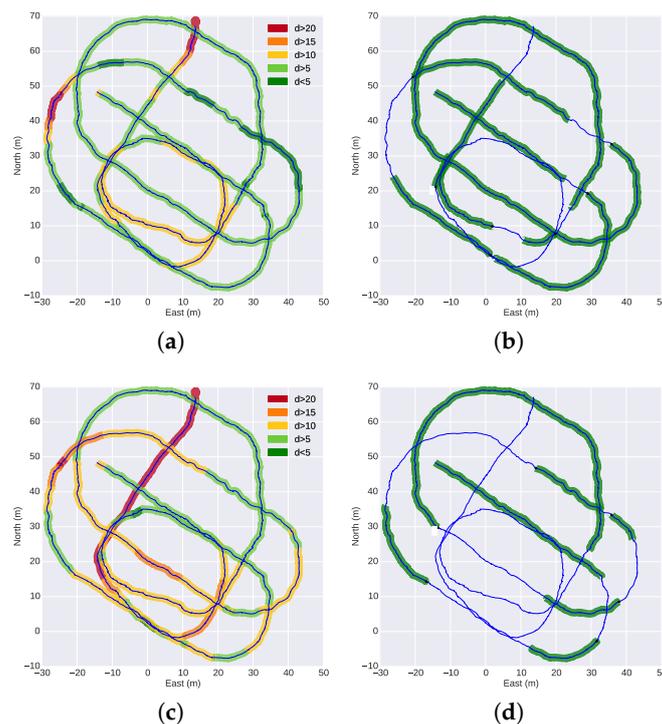


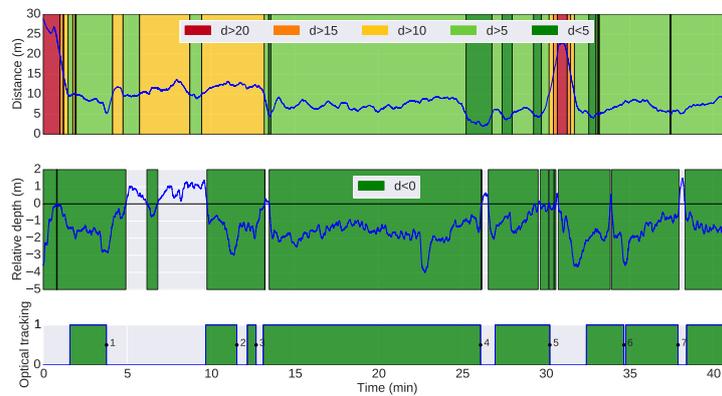
Figure 13. Top view of the trajectory of the G500 during the mission. The distances between G500 and target vehicles are color-coded (a,c); and the intervals of sustained tracking operation are marked in green (b,d). The points where the tracking was lost are numbered (b,d) for further analysis in Tables 1 and 2. The survey of the vertical wall corresponds to the lower-left corner of the images. (a) Distance between G500 and Seacat; (b) intervals of optical tracking of the Seacat; (c) distance between G500 and Sparus; (d) intervals of the optical tracking of the Sparus.

Table 1. Instants with a loss of visual tracking for Seacat and its failure mode. Points are numbered according to Figures 13b and 14a.

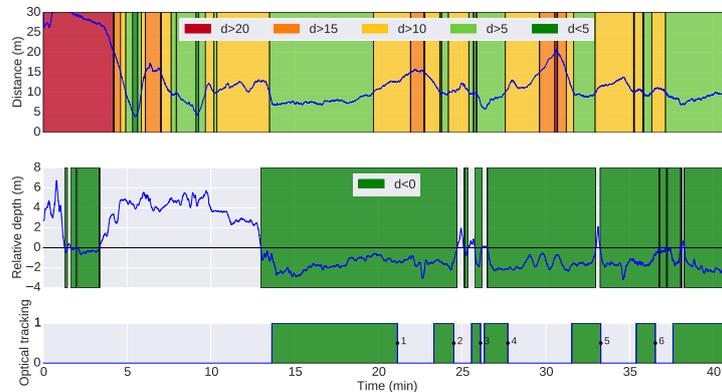
Point	1	2	3	4	5	6	7
Failure mode	Geometry	Distance	Distance	FOV	Distance	Abrupt change in depth	FOV

Table 2. Instants with loss of visual tracking for Sparus and its failure mode. Points are numbered according to Figures 13d and 14b.

Point	1	2	3	4	5	6
Failure mode	Distance	Distance + Geometry	Occlusion with Seacat	Distance	Distance	Distance



(a)



(b)

Figure 14. Analysis of intervals with optical tracking *versus* distance and relative depth between G500 and the target vehicles. Looking at the three plots in conjunction, it is easy to infer the necessary conditions for the operation of the optical tracking. (a) Top: analysis of the distance between G500 and Seacat over time; middle: analysis of the relative depth between G500 and Seacat over time. Negative values mean that Seacat was below G500; thus, the optical tracking was plausible. Bottom: intervals of optical tracking. The points where the tracking was lost are numbered for further analysis in Table 1; (b) Top: analysis of the distance between G500 and Sparus over time; middle: analysis of the relative depth between G500 and Sparus over time. Negative values mean that Sparus was below G500; thus, the optical tracking was plausible. Bottom: intervals of optical tracking. The points where the tracking was lost are numbered for further analysis in Table 2.

Regarding the failure modes, Figure 15 shows two examples of the limit conditions of operation. Both images correspond to loss of tracking after a long interval of operation. In both cases, the distance between the camera and the vehicles was a decisive factor. The limit distance is determined not only by the visibility conditions, but also by the resolution of the camera, which affects the apparent size of the light disk. In these images, the disks are less than two pixels in diameter. This was further compounded with the effects of ambient light and, in the case of Seacat, with the poor geometry of the image projected light beacons, where two beacon disks are overlapping.

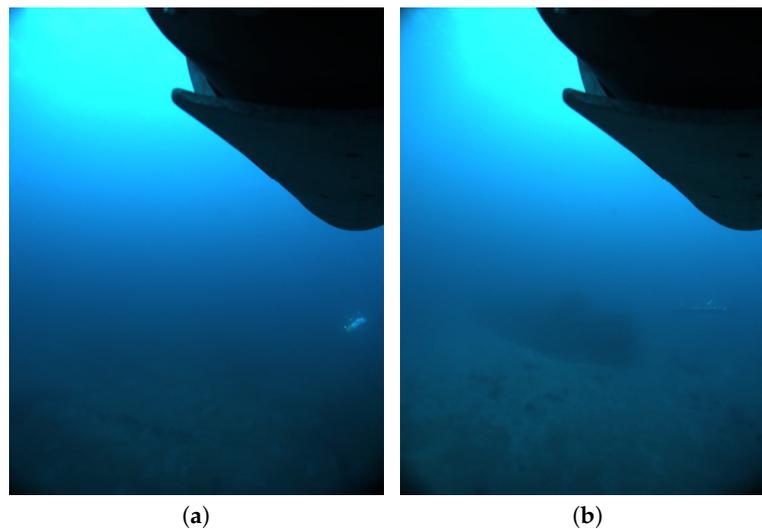


Figure 15. Two instances where the tracking system lost the target vehicles due to the visibility conditions. The lights at the distance of the vehicles were too weak for the system to discern them and keep the tracking running. (a) Sparus at approximately a 13.2-m distance; (b) Seacat at approximately a 12.4-m distance.

Tables 1 and 2 show the failure modes for each instant where tracking was lost during the mission. Three main reasons appear: (1) *distance*, the distance between the vehicles was too big and tracking became unreliable; (2) *FOV*, the target vehicle navigated higher than the camera-vehicle, and thus, it moved outside the camera field of view, or occlusions appeared between the markers and the body of the target vehicle; (3) *geometry*, the combination of the pose of the target vehicle with the geometry of the markers made it very difficult to estimate the pose correctly.

Let us now consider one of the longest sections with tracking capabilities for both vehicles. In Table 3, we compare the information of the relative navigation between the leading vehicle and the target vehicles in three different cases: (1) using only acoustic ranges; (2) using USBL updates; or (3) using the light beacon tracking system. One of the most important limitations of underwater communications using acoustics is that for a reliable and stable communication, only one emitter is allowed to send messages at any given time. For this reason, when several vehicles co-exist, their access to the media has to be scheduled using a time division media access (TDMA) protocol, assigning to each vehicle a safe slot of time for sending messages. In the case of the mission analyzed, due to the number of vehicles involved and all of the different uses of the acoustic channel, each one of the vehicles was able to send messages every 5 s. We can observe from Table 3 that the information provided by the light beacon system is the most complete, as it reports data on both the position and orientation of the target vehicle, unlike USBL or range-only systems, which report only relative position or distance. The number of updates received is another significant difference: acoustic-based systems only provided 122 updates during the period studied, while the light-beacon system reported up to

90-times more updates. This is due to the constraints on the acoustic or range-only communications when multiple vehicles are in the water, while the optical system is independent of the number of vehicles in the water.

Figure 16 shows a comparison of the distance between the G500 and the Seacat AUV computed through acoustic ranges and using the light beacon system: the two online localization systems that were available for the leading vehicle. Looking at the figure, it is clear that there is a high level of agreement between the two systems, and there is no apparent offset between the different approaches. We can also see the difference in the number of updates.

Table 3. Updates of relative navigation position received by Girona500 about Seacat during an interval of a duration of 11 min 21 s. USBL, ultra-short baseline system.

Localization system	Acoustic Ranges	USBLs	Light Beacons
Updates	122	122	10894
Information	Distance	Position	Position and orientation

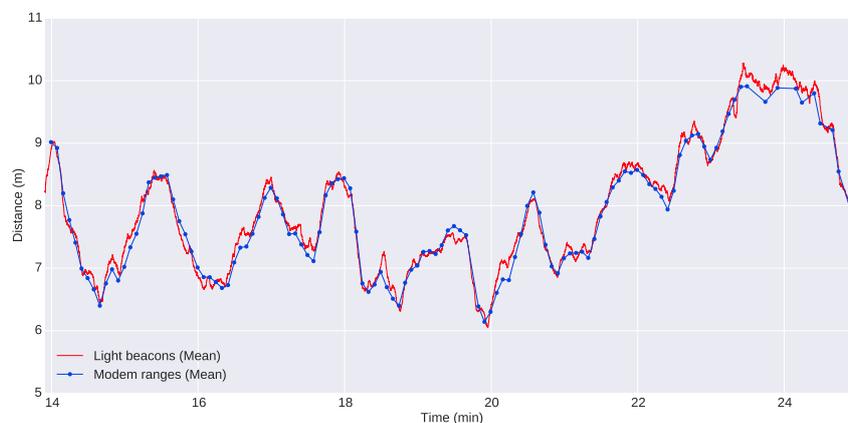


Figure 16. Comparison of the distance between the G500 and the Seacat estimated using light beacons and acoustic ranges.

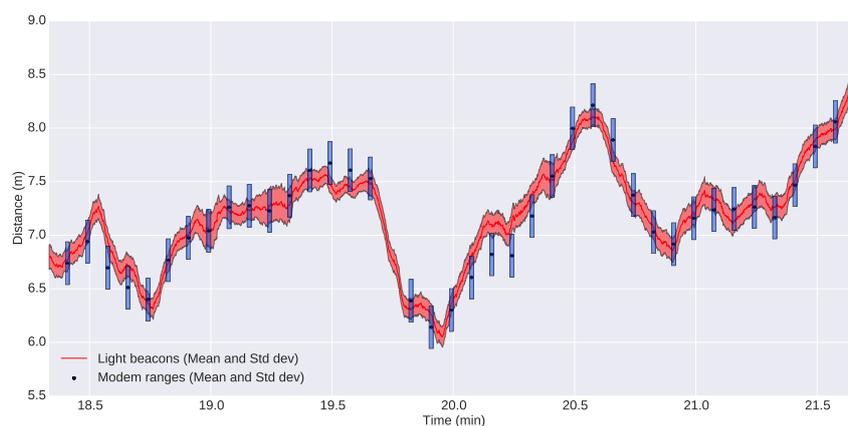


Figure 17. Detailed comparison between the distance estimated using light beacons and acoustic ranges along with their uncertainties.

Figure 17 shows a zoomed section of the former plot, but including the uncertainties of the measurements. The uncertainties of the poses computed through the optical system have been

calculated according to Section 3, while the uncertainty of the acoustic ranges has been estimated as a fixed value obtained from computing the standard deviation between data from the acoustic ranges and the estimated distances using the optical system. For both cases, we have chosen conservative values: a standard deviation of 0.2 m for the acoustic ranges and a standard deviation of two pixels for localization of the lights in the images, as explained in Section 3.2.2. We can observe from the figure that the optical system is very competitive against acoustic ranges and even for the worst cases does not exceed the 0.2-m standard deviation of the acoustic ranges.

The last two plots provide evidence of the accurate behavior of the tracking system. However, they are only a comparison in one dimension, whereas the system provides full position information. With the aim of achieving a more complete comparison, Figure 18 shows the navigation of the three vehicles according to different navigation systems. Different conclusions can be drawn from the figure:

- Comparing the trajectories computed by the internal navigation systems of each vehicle, labeled as *dead reckoning*, with the USBL updates and the trajectories computed by the other systems, we can infer that the internal navigation systems have a significant drift over time, especially Sparus, due to the fact that they are based on inertial sensors and DVL.
- The distances between the USBL updates and the trajectory computed by the light beacon system are small. This fact makes clear that the computed trajectory is consistent with the USBL updates. It also becomes clear that the number of updates is significantly higher for the light beacon system, as seen previously.
- The navigation trajectories computed by the optical method and the offline optimization method have a high degree of agreement. From this fact, we can conclude that the light beacon system is consistent with the offline optimized trajectory of the vehicles. This trajectory was computed once the mission was finished, gathering the navigation data from the vehicles participating in the mission and the acoustic ranges and USBLs received, and is assumed to be the best navigation estimate we can compute without the light beacon system. The trajectory was computed with the algorithm described in [32].

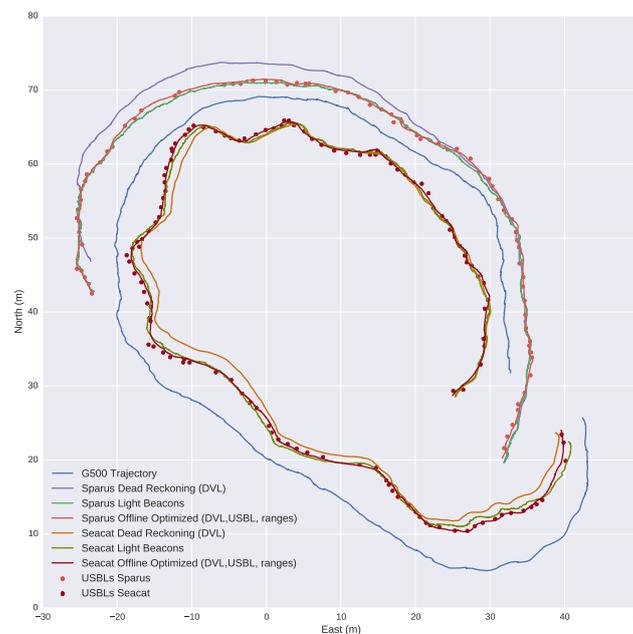


Figure 18. Comparison between the estimated position of target vehicles using light beacons and the position according to their navigation systems before and after its optimization.

Finally, Figure 19 show the evolution of the uncertainties in the position of the Seacat during a section of the mission. The plot shows the difference in the uncertainties of the estimates computed from three or four markers and how the uncertainty grows when the pose of the target vehicle only allows the sight of three of them. It is also evident how the use of the EKF allows a drastic reduction in the uncertainty of the estimates and, thus, improves significantly the performance of the optical system.

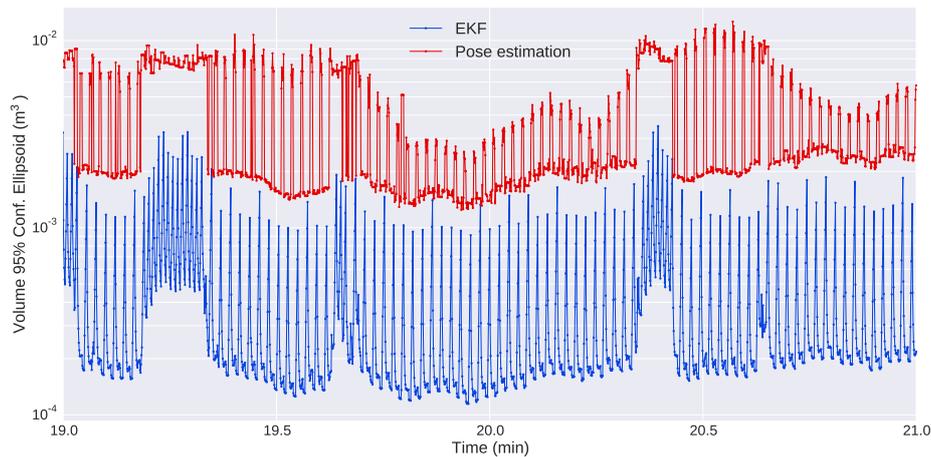


Figure 19. Volume of the ellipsoid that contains the position of the target vehicle with a probability of 95%. In red, the volume of the ellipsoid computed directly from the uncertainty given by the pose estimation module. In blue, the volume of the ellipsoid according to the uncertainty estimated by the EKF. The peaks in the red plot correspond to the pose estimates computed from only three markers, which have a greater uncertainty than the estimates computed from four markers. The peaks in the blue plot show how the uncertainty of the position grows until a new observation of the target pose is received.

5. Conclusions

This paper has presented a complete method for the tracking of AUVs when navigating in close-range based on vision and the use of active light markers. While traditional acoustic localization can be very effective from medium to long distances, it is not so advantageous at short distances, when for safety reasons, it is necessary to achieve a higher precision and faster update rate. The use of the proposed system in such conditions allows the pose estimation of a target vehicle with high accuracy and speed. To provide the system with the best possible performance, the camera used in the experiments was an omnidirectional model, which provides a coverage of 360° in the horizontal plane and allows the system to track vehicles in different positions simultaneously.

The system was tested in mapping missions in real sea conditions. The results presented focused on a mission in which three vehicles were involved: a leader vehicle at the front of the formation and two AUVs at the back for capturing images of the seafloor. These last two were fitted with one set of light beacons each, while the leader was equipped with the omnidirectional camera. The control of the formation was performed using acoustic ranging, but the light beacons enabled the possibility of reconstructing their navigation and comparing them to their own navigation and acoustic systems. As expected, the results showed that the system performs adequately for vehicle separations smaller than 10 m, while the tracking becomes intermittent for longer distances due to the challenging visibility conditions underwater.

The navigation reconstructed from the output of the light tracking system shows a high level of agreement with the navigation computed from the vehicles' navigation systems and acoustic ranging. In addition, when compared to a multi-vehicle setup with strong constraints on the use of the acoustic channel, the light beacon system is able to provide an enormous increase in the frequency of the

updates (90-fold in the case of our test setup). It can also provide information about the orientation of the target vehicles, which most common acoustic systems cannot achieve without using the explicit data communication of sensor readings.

As a central conclusion, the use of light beacon technology for the pose estimation of underwater vehicles can be considered to be at a technology readiness level of at least six, since this paper presents a working prototype operating in a relevant/operational environment. This work also contributes by raising the feasibility of active light markers for complex cooperative underwater operations in close range, such as mapping missions for 3D environments or cooperative intervention missions.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-8220/16/4/429/s1>.

Acknowledgments: The authors wish to thank all partners participating in the MORPH project: the framework of the experiments presented in this paper. We especially appreciate the contributions of ATLAS ELECTRONIK, who provided the Seacat AUV used for the mission presented here, and Instituto Superior Técnico Lisboa, who provided multiple robots for the experiments and were in charge of the multiple vehicle formation control and navigation systems. This research was sponsored by the MORPH EU FP7-Project under Grant Agreement FP7-ICT-2011-7-288704, the Spanish National Project OMNIUS (Lightweight robot for omnidirectional underwater surveying and telepresence) under Agreement CTM2013-46718-R, the ROBOCADEMY (European Academy for Marine and Underwater Robotics) EU FP7-Project under Grant Agreement 608096, the Generalitat de Catalunya through the ACCIO (Agència per a la competitivitat de l'empresa de la Generalitat de Catalunya)/TecnioSpring program (TECSPR14-1-0050), with the support of "la Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya".

Author Contributions: The work presented in this paper has been carried out in collaboration with all authors. Josep Bosch designed, implemented and tested the system with the collaboration of Klemen Istenic and Nuno Gracias. David Ribas was in charge of all system hardware. Pere Ridao and Nuno Gracias were the project leaders and in charge of the direction and supervision. All authors discussed the results obtained together and reviewed the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kalwa, J.; Pascoal, A.; Ridao, P.; Birk, A.; Eichhorn, M.; Brignone, L. The European R & D-Project MORPH: Marine robotic systems of self-organizing, logically linked physical nodes. In Proceedings of the IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles (NGCUV), Porto, Portugal, 10–12 April 2012.
2. Garcia, R.; Gracias, N. Detection of interest points in turbid underwater images. In Proceedings of the IEEE/MTS OCEANS 2011 Conference, Santander, Colombia, 6–9 June 2011.
3. Kinsey, J.; Eustice, R.; Whitcomb, L. A survey of underwater vehicle navigation: Recent advances and new challenges. In Proceedings of the IFAC Conference of Manoeuvring and Control of Marine Craft, Lisbon, Portugal, 20–22 September 2006.
4. Stutters, L.; Liu, H.; Tiltman, C.; Brown, D. Navigation Technologies for Autonomous Underwater Vehicles. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **2008**, *38*, 581–589.
5. Yun, X.; Bachmann, E.; Mcghee, R.; Whalen, R.; Roberts, R.; Knapp, R.; Healey, A.; Zyda, M. Testing and evaluation of an integrated GPS/INS system for small AUV navigation. *IEEE J. Ocean. Eng.* **1999**, *24*, 396–404.
6. Whitcomb, L. Underwater robotics: Out of the research laboratory and into the field. In Proceedings of the IEEE International Conference on Robotics and Automation, San Francisco, CA, USA, 24–28 April 2000.
7. Rigby, P.; Pizarro, O.; Williams, S. Towards Geo-Referenced AUV Navigation through Fusion of USBL and DVL Measurements. In Proceedings of the 2006 MTS/IEEE OCEANS, Boston, MA, USA, 18–21 September 2006; pp. 1–6.
8. Whitcomb, L.; Yoerger, D.; Singh, H. Advances in Doppler-based navigation of underwater robotic vehicles. In Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, USA, 10–15 May 1999.
9. Turetta, A.; Casalino, G.; Simetti, E.; Sperinde, A.; Torelli, S. Analysis of the accuracy of a LBL-based underwater localization procedure. In Proceedings of the 2014 MTS/IEEE OCEANS—St. John's, NL, USA, 14–19 September 2014; pp. 1–7.

10. Morgado, M.; Oliveira, P.; Silvestre, C. Design and experimental evaluation of an integrated USBL/INS system for AUVs. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, USA, 3–7 May 2010; pp. 4264–4269.
11. Brignone, L.; Alves, J.; Opderbecke, J. GREX sea trials: First experiences in multiple underwater vehicle coordination based on acoustic communication. In Proceedings of the 2009 MTS/IEEE OCEANS—Europe, Bremen, Germany, 11–14 May 2009; pp. 1–6.
12. Engel, R.; Kalwa, J. Coordinated Navigation of Multiple Underwater Vehicles. In Proceedings of the 17th International Offshore and Polar Engineering Conference (ISOPE-2007), Lisbon, Portugal, 1–6 July 2007; pp. 1066–1072.
13. Sanz, P.; Ridao, P.; Oliver, G.; Casalino, G.; Petillot, Y.; Silvestre, C.; Melchiorri, C.; Turetta, A. TRIDENT An European project targeted to increase the autonomy levels for underwater intervention missions. In Proceedings of the 2013 MTS/IEEE Oceans, San Diego, San Diego, CA, USA, 23–27 September 2013; pp. 1–10.
14. Prats, M.; Ribas, D.; Palomeras, N.; García, J.C.; Nannen, V.; Wirth, S.; Fernández, J.J.; Beltrán, J.P.; Campos, R.; Ridao, P.; *et al.* Reconfigurable AUV for intervention missions: A case study on underwater object recovery. *Intell. Serv. Robot.* **2012**, *5*, 19–31.
15. Censi, A.; Strubel, J.; Brandli, C.; Delbruck, T.; Scaramuzza, D. Low-Latency Localization by Active LED Markers Tracking Using a Dynamic Vision Sensor. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 891–898.
16. Faessler, M.; Mueggler, E.; Schwabe, K.; Scaramuzza, D. A Monocular Pose Estimation System Based on Infrared LEDs. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014.
17. Krupinski, S.; Maurelli, F.; Mallios, A.; Sotiropoulos, P.; Palmer, T. Towards AUV docking on sub-sea structures. In Proceedings of the 2009 MTS/IEEE OCEANS, Bremen, Germany, 11–14 May 2009.
18. Li, Y.; Jiang, Y.; Cao, J.; Wang, B.; Li, Y. AUV docking experiments based on vision positioning using two cameras. *Ocean Eng.* **2015**, *110*, 163–173.
19. Gracias, N.; Bosch, J.; Karim, M.E. Pose Estimation for Underwater Vehicles Using Light Beacons. *IFAC-PapersOnLine* **2015**, *48*, 70–75.
20. Carreras, M.; Candela, C.; Ribas, D. Sparus II, design of a lightweight hovering AUV. In Proceedings of the 5th International Workshop on Marine Technology (MARTECH), Girona, Spain, 9–11 October 2013; pp. 163–164.
21. Point Grey Research, I. 360 Spherical—Ladybug3—Firewire Camera. Available online: <http://www.ptgrey.com/products/ladybug3/> (accessed on 21 March 2016).
22. Ribas, D.; Palomeras, N.; Ridao, P.; Carreras, M.; Mallios, A. Girona 500 AUV: From Survey to Intervention. *IEEE/ASME Trans. Mech.* **2012**, *17*, 46–53.
23. Hartley, R.I.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2003.
24. Bosch, J.; Gracias, N.; Ridao, P.; Ribas, D. Omnidirectional Underwater Camera Design and Calibration. *Sensors* **2015**, *15*, 6033–6065.
25. Bradski, G. The OpenCV Library. *Dr. Dobby J. Softw. Tools* **2000**, *25*, 120–124.
26. Agarwal, S.; Mierle, K. Ceres Solver. Available online: <http://ceres-solver.org> (accessed on 21 March 2016).
27. Moyset, M.M. Automatic Source Detection in Astronomical Images. Ph.D. Thesis, Universitat de Girona, Girona, Spain, 2014.
28. Smith, R.; Self, M.; Cheeseman, P. *Autonomous Robot Vehicles*; Springer-Verlag New York, Inc.: New York, NY, USA, 1990; pp. 167–193.
29. Mahalanobis, P.C. On the generalised distance in statistics. *Proc. Natl. Inst. Sci. India* **1936**, *2*, 49–55.
30. Moré, J.J. The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical Analysis*; Springer: Berlin, Heidelberg, 1978; pp. 105–116.
31. Kalwa, J. The RACUN-project: Robust acoustic communications in underwater networks—An overview. In Proceedings of the 2011 MTS/IEEE OCEANS—Spain, Santander, Spain, 6–9 June 2011, pp. 1–6.
32. Campos, R.; Gracias, N.; Ridao, P. Underwater Multi-Vehicle Trajectory Alignment and Mapping Using Acoustic and Optical Constraints. *Sensors* **2016**, *16*, 387.

33. Abreu, P.C.; Pascoal, A.M. Formation control in the scope of the MORPH project. Part I: Theoretical Foundations. *IFAC-PapersOnLine* **2015**, *48*, 244–249.
34. Abreu, P.C.; Bayat, M.; Pascoal, A.M.; Botelho, J.; Gois, P.; Ribeiro, J.; Ribeiro, M.; Rufino, M.; Sebastiao, L.; Silva, H. Formation control in the scope of the MORPH project. Part II: Implementation and Results. *IFAC-PapersOnLine* **2015**, *48*, 250–255.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

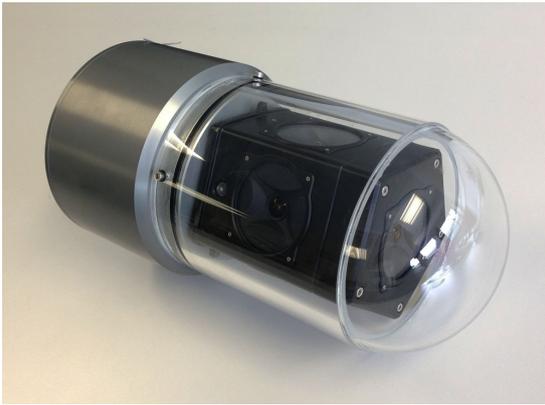
5

MAIN RESULTS AND DISCUSSION

IN this chapter we present the main results obtained during the development of the thesis. According to the objectives stated in Section 1.2, the results are presented in 5 different sections: Section 5.1: Design of an Omnidirectional Multi-Camera System, Section 5.2: Calibration of an Omnidirectional Multi-Camera System, Section 5.3: Generation of panoramic imagery and video, Section 5.4: Immersive Virtual Reality applications and Section 5.5: Close Range Tracking of multiple vehicles.

5.1 Design of an Omnidirectional Multi-Camera System

Two different OMS have been developed during this thesis. In Chapter 2 we presented a system for real time purposes based on a PointGrey Ladybug 3 camera (Fig. 5.1a). In Chapter 3 we described in detail a modular OMS based on Gopro Hero 4 cameras (Fig. 5.1b) which is intended for high-quality offline mapping. While in the first case the central work consisted in adapting the off-the-shelf system for underwater use, in the later the design of the multi-camera system was developed from scratch.



(a) Multi Camera System based on PointGrey Ladybug 3.



(b) Custom designed OMS based on Gopro Hero 4 cameras.

Figure 5.1: Omnidirectional cameras designed during this thesis.

When designing an underwater OMS there are many aspects to take into account. Among others, these aspects include the resolution of the cameras, the intended depth rate, the need of real-time access to the images and the power supply. It is especially relevant to ensure that the FOV of the designed system is able to cover 360° at the desired minimum range. The FOV of the system depends basically on the optics, relative location and orientation of the individual cameras and the waterproof housing used. The latter is especially relevant since the FOV can be drastically reduced depending on the chosen shape of the viewport. As explained in detail in chapter Chapter 2 the rays of light coming from the underwater scene must pass through two medium transitions (water-case and case-air) before reaching the camera lenses. These media transitions imply changes in the direction of the incoming rays. The change of the direction is described by Snell's law and it depends on two factors: the angle of incidence and the material of the viewport. Due to the relevant role of the geometry of the waterproofing used, a FOV simulator for OMS was implemented in order to analyze all the different combinations of cameras and waterproof housings. This tool proved to be especially useful in cases where it is not possible to design individual view ports for each of the individual cameras, such as when the relative position of the cameras is predefined and there are severe space limitations. In such cases, the camera housing needs to have a single view port shared by all cameras, which make the analysis of distortions and the calibration significantly more complicated.

The two most typical geometries used when designing an underwater camera housing are flat and hemispherical interfaces. Flat interfaces are less expensive to manufacture and easier to mount, but introduce important bending in the rays, which drastically reduces the FOV of the cameras. By contrast hemispherical interfaces preserve the FOV when positioned exactly in the optical center of the camera. However, perfect hemispherical interfaces are difficult to produce and to mount at the exact desired position. Furthermore, they reduce significantly the camera focusing range, especially for small radius domes.

For the first custom-designed system, a two-piece housing was designed. The first piece is a cylinder and covers the whole FOV of the five lateral cameras, while the second piece is an hemispherical port that covers the downward-looking camera.

The reason for choosing a cylindrical port was strongly linked to the fact that the cameras were rigidly fixed. It was not possible to use individual flat ports due to the large loss on the FOVs. Hemispherical ports were not suitable either, because the minimum dome size to ensure proper focusing could not be used due to the geometrical arrangement of the cameras. A cylindrical dome-port reduces the vertical FOV similarly to a flat port, but has the advantage of having a much wider horizontal FOV.

By contrast, for the second system designed, the spatial arrangement of the cameras was decided after a detailed analysis of their FOV [5]. The design priorities were to use the minimum number of cameras and to introduce the minimum distortion in the images. For these reasons hemispherical ports were chosen. The size of the dome was decided taking into account the focusing range of the cameras. Once all these decisions were made, the structure holding the cameras was designed to achieve the most compact system possible.

For the first system, an open-source driver for Linux was developed and published along with a ROS package to permit its integration with any robotics platform.

5.2 Calibration of an Omnidirectional Multi-Camera System

A complete pipeline for calibrating a OMS has been presented in chapter Chapter 2. The proposed method is valid for both air and underwater cameras as the strategy to do the complete calibration is done in three different stages.

The first stage consists of the estimation of the intrinsic parameters, which is done individually for each one of the cameras in air, and without the waterproof housing. This stage can be broken down into two parts. First, an initialization of the intrinsic parameters is obtained using a standard toolbox for calibration. Multiple shots of a planar grid are acquired at different positions and orientations to estimate the parameters. In a second step, a refinement procedure is performed to improve accuracy. This step is needed due to the usual high distortion of the cameras used in OMS. Standard toolboxes for calibration need to see the entire extent of a planar calibration board in the images to complete a calibration. However due to the high distortion of the images, it is not possible to place the checker board close to the corners of the image while seeing it entirely. This can lead to inaccurate results for regions of the image close to the corners. To avoid this issue, we proposed a different approach that uses a more versatile pattern which allows us to use images when only a portion of the pattern is visible. The pattern proposed in Chapter 2 consists of a large printout of a known image An aerial orthophoto was used since it provides a large number of visual features at different scales.

For each one of the shots, a camera pose with respect to the ortophoto is estimated by minimizing the re-projection error of all correct matches using the initial intrinsic parameters from the previous step and solving the perspective-n-point (PnP) problem. The PnP problem is the problem of the determination of the position and orientation of a calibrated camera given a set of n correspondences between 3D points and their 2D projections. Finally, we define an optimization problem that finds the refined intrinsic parameters. The number of variables to estimate is the totality of the intrinsic parameters (focal length, principal point coordinates and distortion parameters), and the pose of the ortophoto in all of the selected images. The initial values for the intrinsic parameters are the ones found in the initialization step, while the initial values for the pose of the ortophoto in every image are the results of the PnP problem solved previously. In order to quantify the uncertainty of the estimated parameters, a Monte Carlo analysis is carried out once the optimization has finished.

The second stage consists of the estimation of the extrinsic parameters, which is the major issue when working with a OMS. Again, this step is carried out in dry conditions and without the waterproof housing. The procedure to estimate the extrinsic parameters is very similar to the one used in the refinement of the intrinsic ones. However, images from two or more cameras acquired at the same exact time frame are needed. All cameras must acquire a recognizable section of the ortophoto while one of the other cameras acquires a different section of the ortophoto simultaneously. The observation of different parts of the ortophoto by two cameras at the same instant implicitly imposes constraints on the relative placement and relative orientation of the two cameras. The parameters to estimate are the relative orientation and position of each one of the cameras with respect to the origin of the OMS.

The final stage takes place underwater and estimates the camera pose with respect to the waterproof housing. Due to the ray bending, any small variation in the assumed relative position of the housing can significantly affect the final direction of the rays and end up generating large projection errors. For this purpose, the relative position of the housing with respect to the camera is estimated in a procedure almost identical to the one for extrinsic parameters. The main difference is that it now uses images captured underwater and with a different printout specially prepared to be placed underwater. The parameters to estimate in this stage are the position and the orientation of the global or individual housings with respect to the origin of the system.

The reason for dividing the calibration procedure into three stages, rather than a single combined step, is two-fold. Firstly, it allows for a smaller number of parameters to be estimated in each individual step, thus avoiding unwanted correlations among the parameter estimates. Secondly, it allows the use of image sets captured in air, for the estimation of the parameters that are not related with the underwater housing. This way, the intrinsic and extrinsic parameters are not affected by disturbances, such as the non-modeled geometric inaccuracies of the waterproof housing. Furthermore, it is significantly easier to find feature points in images captured in air than in water, due to light absorption and varying illumination conditions. And the use of a larger number of well-spread feature points contributes to a higher calibration accuracy.

While for the first of the cameras used in this thesis (Fig. 5.1a) the whole calibration procedure was carried out, for the second system (Fig. 5.1b), the third stage was omitted. This is due to the fact that the second system uses independent spherical housings. When spherical domes are centered exactly in the center of the camera, the direction of the rays

of light are not altered when passing through the waterproof housing. This condition implies the calibration of the camera does not change when using the waterproof housing, and makes the third stage of the calibration unnecessary.

5.3 Generation of panoramic imagery and video

In chapter 2 the generation of panoramic images from multiple images was introduced to show the results of the design and calibration of the OMS. The method used to generate the panorama consisted in the state-of-the-art strategy that assumes the world is a sphere centered on the origin of the OMS. While this strategy performs very well when the objects in the scene are all far from the camera, it performs very poorly when the scene contains elements located at very different distances. For these cases, severe mis-alignments are visible in the panoramas in the transition zones between cameras, due to parallax effects. The parallax is the difference in location of an object in an image seen by two different cameras, and is the source of mis-alignments on the final panorama if not handled correctly. In order to address this issue, new strategies were presented in chapter 3 to obtain panoramas. These strategies are an important improvement over the previous state-of-the-art methods, as they take into account the shape of the scene where the images are captured to generate the panoramas.

Among the solutions presented, two are the most relevant. The first was designed to be fast and computationally inexpensive, since it was envisioned for real-time applications. The main idea is that during an ongoing SLAM process a sparse 3d point cloud of the scene can be obtained. This point cloud can be converted to a sparse panoramic depth map, which can be later interpolated and smoothed to achieve a full dense panoramic depth map that can be texturised. Using this strategy allows a final panorama to be obtained without visible gaps and with less visible mis-alignments in the transitions between cameras. However, it could introduce some slight deformations in the final panorama.

The second strategy uses more sophisticated and time-expensive methods as it is intended to be used during a post-process stage. The method is based on the idea that a 3D reconstruction of the scene can be obtained making use of Structure from Motion (SfM) approaches. Then individual depth maps can be extracted from the model and re-projected to an equirectangular panorama with its origin on the center of the system. The equirectangular depth map is then smoothed in order to fill the gaps that were not observed by any of the cameras due to parallax. Finally, the contribution of each of the cameras is defined by the use of a graph-cut technique. The goal is to trace a seam between cameras that has the minimum accumulated error when comparing the obtained depth map with the ideal depth map generated through the 3D reconstruction. By using this strategy the visible mis-alignments between camera transitions are reduced drastically.

These methods were tested with a real underwater dataset of a sunken ship, collected using the OMS based on Gopro cameras. The baseline between the cameras is larger than the one on the Ladybug-based camera, and thus, the effects of parallax are more prominent. The methods were applied to 20-second video sequence, which was selected to contain a scene with objects at very different distances, thus allowing a better evaluation of the performance of the methods presented. Both methods significantly improved the quality of the generated panorama when compared to the spherical-world assumption. The graphcut-based solution gave the more visually pleasant results with no visible seams in the final panorama, at the cost of being a technique that needs to be applied offline.

5.4 Immersive Virtual Reality applications

During the development of this thesis, especial attention was given to the study of the different capabilities and applications where omnidirectional underwater cameras could have an impact. One of these applications is the generation of omnidirectional videos and Virtual Tours (VTs) for public dissemination.

As seen in Chapter 3, panoramic videos can be generated using different methodologies. Once the panoramic video is obtained, it can be visualized in any standard spherical video player. The number of such video players is increasing rapidly along with the popularity of 360° videos (e.g. VLC, Oculus Player, Total Cinema 360, among many others). Many social networks, such as Facebook, YouTube and Google Plus also allow the upload of 360° videos.

To achieve an even more immersive experience, panoramic videos can be watched through Virtual Reality headsets, such as the Oculus Rift. These headsets evoke a unique feeling of immersion that could be used in schools, museums and interpretation centers.

When used together with an AUV, the panoramic images captured by omnidirectional cameras can be augmented with their geo-location, obtained from the navigation data of the vehicle. This allows VTs to be created, where users can virtually visit a region of special interest by looking at a set of selected panoramas in a very similar way to how Google Street Map allows moving along images captured on the street. Many platforms exist that allow virtual tours to be created, although the most used is Google Maps, where panoramas can achieve their maximum impact by reaching a very wide audience. However, other Virtual Tour platforms have additional features that might be very useful for science education, such as the possibility of placing augmented reality markers with relevant information in the image, or displaying a customized map or informative drawings with the geolocation of the panoramas.

During the development of the thesis, multiple datasets have been collected using both AUVs and divers, in the framework of the field experiments in the projects listed in Chapter 1.

With the purpose of showing the potential of omnidirectional imagery for public education we describe briefly the results obtained with two of the datasets. We refer the reader to [5, 9] for further details on the techniques used.

The first experiment was a short survey over a shipwreck in the bay of Porto Pim in Horta, Azores Islands, that was carried out in September 2014 during trials conducted in the framework of the MORPH EU-FP7 project [80].

The Girona500 AUV [81] was equipped with the Ladybug-based omnidirectional camera and was tele-operated from a land-base by a pilot who was operating the robot through the preview images acquired from the camera. After the recovery of the vehicle, the data from the camera was processed. The panoramic images from the survey were created from the individual images of each camera composing the OMS and the calibration information, and later tagged with a GPS location according to the AUV's navigation.

A subset of the panoramic images were selected for the creation of a Virtual Tour (Fig. 5.2). In order to automate the process, we developed tools that automatically select the images depending on a chosen criterion, which can be distance-based, time-based, according to the amount of visual information, or other.

In the second experiment, the objective was to demonstrate the advantages of omnidirectional cameras on underwater robots for the rapid high-resolution mapping of ship-

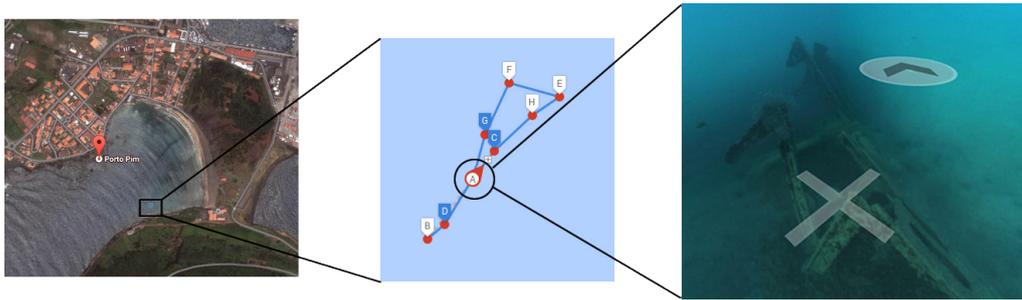


Figure 5.2: Location and sample image of a shipwreck virtual tour in Google Maps. **Left:** Location of the shipwreck in Porto Pim bay in Horta, Azores Islands. **Middle:** Different panoramas and connections available in the virtual tour. **Right:** Sample view of the shipwreck from A.

wrecks in marine archaeology applications. In collaboration with the University of Zadar, the methodology was demonstrated on the Gnalić shipwreck during the “Breaking the Surface 2016” workshop held in Biograd na Moru (Croatia). The Girona 500 AUV, equipped with the Gopro-based omnidirectional camera was programmed to survey the shipwreck and the data collected was used to build 360° panoramic videos, topological panoramic maps and 3D optical reconstructions. Results can be seen in Fig. 5.3.



(a) Map of the virtual tour of the shipwreck available at : www.tiny.cc/gnalictour .

(b) Snapshot from the Gnalić shipwreck panoramic video on YouTube. The complete video can be found at : www.tiny.cc/gnalic .

Figure 5.3: Results of the survey of the Gnalić.

5.5 Close Range Tracking of multiple vehicles

One of the objectives of this thesis was to demonstrate that omnidirectional cameras can be used as a sensor to improve the navigation of underwater vehicles. In chapter 4, this goal was studied in detail, and it was put into practice in the framework of the MORPH-FP7 project, where multiple vehicles were performing cooperative navigation missions in a close-range formation.

The objective of the method was the real-time localization of underwater vehicles at distances of less than 10m, and to obtain both position and orientation information with high update rates. The resulting system also had to be robust to short temporal occlusions of the direct line of vision to the target markers.

The proposed solution consisted in the placement of a set of light beacons, or active markers, on the target vehicles, which are optically tracked by a wide-field-of-view camera placed in a camera vehicle. The tracking of these markers allows the 3D pose of the target vehicles to be estimated. Tracking of multiple target vehicles is possible by using blinking patterns of different frequencies. The underlying assumptions are that the camera field of view covers the areas where the vehicles operate, and that the visibility conditions are not too severe for the intended inter-vehicle distances.

The performance of the light-beacon tracking and pose -estimation method was tested in several sea trials. The most relevant field testing took place in Sant Feliu de Guíxols, Spain, in March 2015 and in Horta, Azores Islands, in September 2015.

The system was tested in mapping missions in real sea conditions. The results presented focused on a mission in which three vehicles were involved: a leader vehicle at the front of the formation and two AUVs at the back for capturing images of the seafloor. These last two were fitted with one set of light beacons each, while the leader was equipped with the omnidirectional camera. The control of the formation was performed using acoustic ranging, but the light beacons enabled the possibility of reconstructing their navigation and comparing them to their own navigation and acoustic systems. As expected, the results showed that the system performs adequately for vehicle separations smaller than 10m, while the tracking becomes intermittent for greater distances due to the challenging visibility conditions underwater.

The navigation reconstructed from the output of the light tracking system shows a high level of agreement with the navigation computed from the vehicles' navigation systems and acoustic ranging. In addition, when compared to a multi-vehicle setup with strong constraints on the use of the acoustic channel, the light beacon system is able to provide an enormous increase in the frequency of the updates (90-fold in the case of our test setup). It can also provide information about the orientation of the target vehicles, which most common acoustic systems cannot achieve without using explicit data communication of sensor readings.

An adaptation of the close-tracking method was also used successfully in the context of the LOON-DOCK/SUNRISE project [6, 82]. The project aimed to demonstrate autonomous homing and docking for underwater robots. The proposed approach for locating the Docking Station (DS) consisted in combining acoustic and optical sensing. While range-only localization was used at greater distances, light beacon localization proved to be a reliable strategy at short ranges with low-cost equipment.

The proposed solution consisted in placing a set of active light beacons in distinct and known positions of the DS (Figure 5.4). Using a standard camera, the lights were detected in the images and used to estimate the pose between the DS and the camera. Unlike range-only localization, this method is able to provide information on the relative orientation of the DS, involving the full 6 DoFs (three relative translations and three rotations) which is key information for a correct approach to the DS.

Figure 5.5 presents the results obtained during the final approach to the DS for one of the experiments. The distances between the AUV and the DS estimated according to the acoustic ranges and the light beacons are compared. The methods show a high degree of agreement with the trajectory estimated offline, once the DS position was known precisely. A complete description of the approach used and the results obtained can be found in [4].



Figure 5.4: Debug image of the optical tracking system. In orange the position of the light markers according to the image analysis. In purple, the position of the markers in the DS according to the estimated relative position. AR markers were used as a complementary resource during the final approach when the light beacons were not in the FOV of the camera.

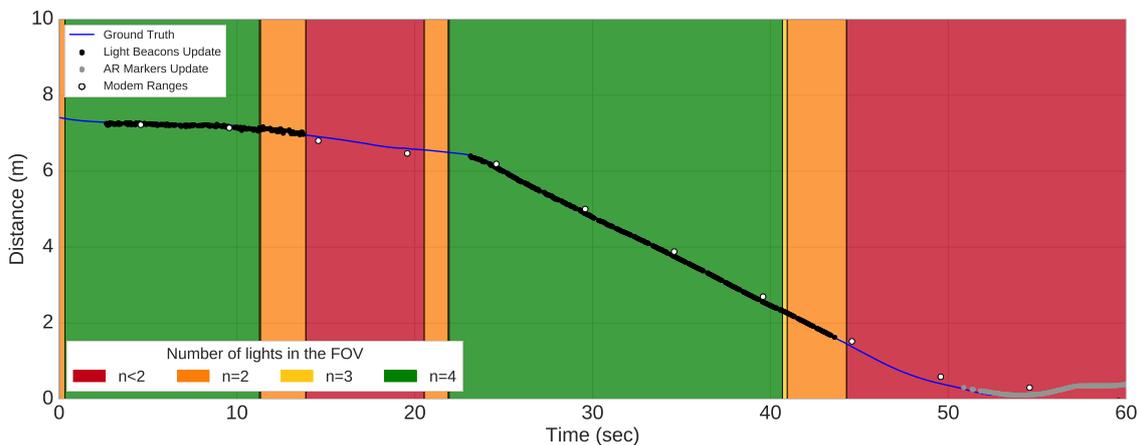


Figure 5.5: DS tracking performance during the docking command execution in one of the missions. The different colored dots represent three different updates: White dots correspond to acoustic updates, black dots are visual updates computed using light beacons, while grey dots are visual updates computed using AR markers. The blue line represents the relative distance between the DS and the AUV according to the navigation of the vehicle, and was computed offline once the position of the DS was known precisely. The colored background reflects how many light beacons were inside the FOV of the camera according to the orientation of the vehicle.

6

CONCLUSIONS AND FUTURE WORK

This chapter concludes the thesis with the main conclusions and suggestions for future research work. Section 6.1 discusses the main conclusions extracted from the work presented and summarizes the contributions to the state of the art of this field. Finally future research lines in the field are discussed in Section 6.2.

6.1 Conclusions

This thesis presented a comprehensive investigation of underwater OMS, with beginning-to-end coverage of the topic, from the design of the cameras to their practical applications in underwater robotics. In spite of the slow establishment of omnidirectional cameras in the underwater environment, they have been proven to have great potential in many different fields, and they will have a big impact once their use becomes widespread.

All objectives detailed in Chapter 1 have been successfully completed and led to the following contributions:

Design of an underwater OMS:

- We developed a FOV simulator to study how the optics of the internal cameras and the geometry of waterproof housings impact on the final FOV of the panorama.
- We designed and developed two underwater OMS, one for online purposes and a second one for high-quality mapping.

Calibration of an OMS:

- We developed a calibration method for OMS applicable to both dry and underwater systems. This method has the advantage that it does not require the calibration images to see an entire calibration pattern, for both intrinsic and extrinsic calibration. This makes it suitable for non-overlapping systems. The method can also handle the distortions introduced by the waterproof housings, thanks to a ray tracing approach.

Generating panoramic imagery and video from an OMS:

- Two new methods for generating panoramic video have been developed in order to deal with parallax effects. One is intended for online use, while the other is intended for a post-process stage.
- A new metric has been proposed for evaluating the quality of the panoramas and the alignment on camera transitions.

Exploring immersive Virtual Reality applications:

- The potential uses of OMS have been analyzed with especial emphasis on the use of omnidirectional video and Virtual Tours for public dissemination.
- Multiple datasets of interesting underwater regions and shipwrecks have been collected with OMS in order to demonstrate the appeal of omnidirectional video and Virtual Tours.

Use of an OMS for aiding underwater navigation on an AUV:

- We have proposed a new method for pose estimation and relative navigation for multiple AUVs at short ranges, based on the use of active light markers. When compared with traditional acoustic-based systems, this method has higher accuracy and update rate, has a lower cost and can provide relative orientation information.

- We tested the proposed method with a real dataset captured in a cooperative navigation mission where 3 AUV were involved. The leading AUV was equipped with a OMS while the two others were fitted with a set of blinking lights each with different blinking frequencies.
- The same method has been applied to single cameras in the context of homing and docking operations for AUV. In this case the set of blinking lights were placed on the DS and were used to localize the AUV against the DS with high precision.

6.2 Future work

Although this thesis covered a wide field related to underwater OMS, there are many different research lines for future work and many aspects of interest which were beyond the scope of this work.

Regarding the generation of panoramic images, we consider that a promising research line is to generate panoramas from larger baselines between the cameras, i.e., placing cameras on different parts of a robot (front, back, laterals), and evaluating the quality of the panoramas obtained. Due to the larger baselines between the cameras, the parallax effects on the final panoramas would be more prominent. However, this would open the door to easily adapting existing robots, which are usually equipped with multiple cameras, for the generation of omnidirectional video. Related to this aspect, we would like to fully implement ROV tele-operation through omnidirectional cameras and Virtual Reality (VR) headsets. Although in the course of developing this thesis we made a first approach to this goal, a full implementation has not been developed yet. A fully working application will need to deal with data-compression methods for the transmission of the images. It will also require an intuitive visual interface for the pilots, in order to receive comprehensive sensor information and allow piloting the ROV while using a VR headset.

Another promising research line for underwater omnidirectional cameras is their use for SLAM. The wide FOV of these cameras will make the task of self-localizing easier, especially in confined environments. Furthermore, as detailed in Chapter 3 it would improve the generation of panoramas for online streaming.

With respect to the use of panoramic imaging for public dissemination, it will be very interesting to implement a fully automatic pipeline for the generation of Virtual Tours. This application would need to deal with the automatic selection of the panoramas to cover the whole extent of a scene with the minimum number of images. While a first automatic approach has been developed in this thesis, there is still a long way to go before achieving an optimal solution that should include view-planning related strategies.

BIBLIOGRAPHY

- [1] **J. Bosch**, N. Gracias, P. Ridao, and D. Ribas. “Omnidirectional Underwater Camera Design and Calibration”. In: *Sensors* 15.3 (2015), pp. 6033–6065. ISSN: 1424-8220. DOI: 10.3390/s150306033 (cit. on pp. vii, 13).
- [2] **J. Bosch**, N. Gracias, P. Ridao, K. Istenič, and D. Ribas. “Close-Range Tracking of Underwater Vehicles Using Light Beacons”. In: *Sensors* 16.4 (2016). ISSN: 1424-8220. DOI: 10.3390/s16040429 (cit. on pp. vii, 71).
- [3] **J. Bosch**, K. Istenič, N. Gracias, R. Garcia, and P. Ridao. “Omnidirectional Multi-Camera Video Stitching using Depth Maps” (cit. on pp. vii, 47).
- [4] N. Palomeras, G. Vallicrosa, A. Mallios, **J. Bosch**, E. Vidal, N. Hurtos, M. Carreras, and P. Ridao. “AUV homing and docking for remote operations”. In: *Ocean Engineering* 154 (2018), pp. 106–120. ISSN: 0029-8018. DOI: <https://doi.org/10.1016/j.oceaneng.2018.01.114> (cit. on pp. vii, 106).
- [5] **J. Bosch**, N. Gracias, P. Ridao, D. Ribas, K. Istenič, R. Garcia, and I. R. Rossi. “Immersive Touring for Marine Archaeology. Application of a New Compact Omnidirectional Camera to Mapping the Gnalíć shipwreck with an AUV”. In: *ROBOT 2017: Third Iberian Robotics Conference*. Ed. by A. Ollero, A. Sanfeliu, L. Montano, N. Lau, and C. Cardeira. Cham: Springer International Publishing, 2018, pp. 183–195 (cit. on pp. vii, 10, 101, 104).
- [6] N. Hurtós, A. Mallios, N. Palomeras, **J. Bosch**, G. Vallicrosa, E. Vidal, D. Ribas, N. Gracias, M. Carreras, and P. Ridao. “LOON-DOCK: AUV homing and docking for high-bandwidth data transmission”. In: *OCEANS 2017 - Aberdeen*. June 2017, pp. 1–7. DOI: 10.1109/OCEANSE.2017.8084806 (cit. on pp. vii, 106).
- [7] G. Vallicrosa, **J. Bosch**, N. Palomeras, P. Ridao, M. Carreras, and N. Gracias. “Autonomous homing and docking for AUVs using Range-Only Localization and Light Beacons”. In: 49.23 (2016). 10th IFAC Conference on Control Applications in Marine Systems CAMS 2016, pp. 54–60. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2016.10.321> (cit. on p. viii).
- [8] G. Nuno, G. Rafel, R. Campos, R. Prados, **J. Bosch**, A. Elibol, T. Nicosevici, L. Neumann, and J. Quintana. “Omnidirectional Underwater Surveying and Telepresence”. In: *Proceedings of the 7th International Workshop on Marine Technology (MARTECH)*. 2016 (cit. on p. viii).

- [9] **J. Bosch**, P. Ridao, D. Ribas, and N. Gracias. “Creating 360 degrees underwater virtual tours using an omnidirectional camera integrated in an AUV”. In: *OCEANS 2015 - Genova*. May 2015, pp. 1–7. DOI: 10.1109/OCEANS-Genova.2015.7271525 (cit. on pp. viii, 104).
- [10] N. Gracias, **J. Bosch**, and M. E. Karim. “Pose Estimation for Underwater Vehicles using Light Beacons”. In: 48.2 (2015). 4th IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles NGCUV 2015, pp. 70–75. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2015.06.012> (cit. on p. viii).
- [11] S. K. Nayar. “Catadioptric omnidirectional camera”. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 1997, pp. 482–488. DOI: 10.1109/CVPR.1997.609369 (cit. on p. 8).
- [12] J. Gaspar, N. Winters, and J. Santos-Victor. “Vision-based navigation and environmental representations with an omnidirectional camera”. In: *IEEE Transactions on Robotics and Automation* 16.6 (Dec. 2000), pp. 890–898. ISSN: 1042-296X. DOI: 10.1109/70.897802 (cit. on p. 8).
- [13] T. Hayashi, H. Uchiyama, J. Pilet, and H. Saito. “An Augmented Reality Setup with an Omnidirectional Camera Based on Multiple Object Detection”. In: *2010 20th International Conference on Pattern Recognition*. Aug. 2010, pp. 3171–3174. DOI: 10.1109/ICPR.2010.776 (cit. on p. 8).
- [14] T. Yoshikawa, M. Koeda, and H. Fujimoto. “Shape recognition and grasping by robotic hands with soft fingers and omnidirectional camera”. In: *2008 IEEE International Conference on Robotics and Automation*. May 2008, pp. 299–304. DOI: 10.1109/ROBOT.2008.4543224 (cit. on p. 8).
- [15] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia. “People tracking and following with mobile robot using an omnidirectional camera and a laser”. In: *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*. May 2006, pp. 557–562. DOI: 10.1109/ROBOT.2006.1641769 (cit. on p. 8).
- [16] I. Marković, F. Chaumette, and I. Petrović. “Moving object detection, tracking and following using an omnidirectional camera on a mobile robot”. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. May 2014, pp. 5630–5635. DOI: 10.1109/ICRA.2014.6907687 (cit. on p. 8).
- [17] Y. Onoe, N. Yokoya, K. Yamazawa, and H. Takemura. “Visual surveillance and monitoring system using an omnidirectional video camera”. In: *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*. Vol. 1. Aug. 1998, 588–592 vol.1. DOI: 10.1109/ICPR.1998.711211 (cit. on p. 8).
- [18] Á. Zarándy, M. Nemeth, Z. Nagy, A. Kiss, L. Santha, and T. Zsedrovits. “A real-time multi-camera vision system for UAV collision warning and navigation”. In: *Journal of Real-Time Image Processing* 12.4 (Dec. 2016), pp. 709–724. ISSN: 1861-8219. DOI: 10.1007/s11554-014-0449-3 (cit. on p. 8).
- [19] T. Ehlgén, M. Thom, and M. Glaser. “Omnidirectional Cameras as Backing-Up Aid”. In: *2007 IEEE 11th International Conference on Computer Vision*. Oct. 2007, pp. 1–5. DOI: 10.1109/ICCV.2007.4409194 (cit. on p. 8).

- [20] T. Gandhi and M. Trivedi. “Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera”. In: *Machine Vision and Applications* 16.2 (Feb. 2005), pp. 85–95. ISSN: 1432-1769. DOI: 10.1007/s00138-004-0168-z (cit. on p. 8).
- [21] A. Rituerto, L. Puig, and J. J. Guerrero. “Visual SLAM with an Omnidirectional Camera”. In: *2010 20th International Conference on Pattern Recognition*. Aug. 2010, pp. 348–351. DOI: 10.1109/ICPR.2010.94 (cit. on p. 8).
- [22] L. Heng, G. H. Lee, and M. Pollefeys. “Self-calibration and visual SLAM with a multi-camera system on a micro aerial vehicle”. In: *Autonomous Robots* 39.3 (Oct. 2015), pp. 259–277. ISSN: 1573-7527. DOI: 10.1007/s10514-015-9466-8 (cit. on p. 8).
- [23] S. E. Chen. “Quicktime VR: An image-based approach to virtual environment navigation”. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM. 1995, pp. 29–38 (cit. on p. 8).
- [24] R. Szeliski. “Video mosaics for virtual environments”. In: *IEEE Computer Graphics and Applications* 16.2 (Mar. 1996), pp. 22–30. ISSN: 0272-1716. DOI: 10.1109/38.486677 (cit. on p. 8).
- [25] S. Peleg and M. Ben-Ezra. “Stereo panorama with a single camera”. In: *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*. Vol. 1. 1999, 401 Vol. 1. DOI: 10.1109/CVPR.1999.786969 (cit. on p. 8).
- [26] R. I. Hartley. “Self-calibration from multiple views with a rotating camera”. In: *Computer Vision — ECCV '94*. Ed. by J.-O. Eklundh. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 471–478. ISBN: 978-3-540-48398-4 (cit. on p. 8).
- [27] R. Orghidan, E. M. Mouaddib, J. Salvi, and J. J. Serrano. “Catadioptric single-shot rangefinder for textured map building in robot navigation”. In: *IET Computer Vision* 1.2 (June 2007), pp. 43–53. ISSN: 1751-9632. DOI: 10.1049/iet-cvi:20065003 (cit. on p. 8).
- [28] T. Svoboda and T. Pajdla. “Epipolar Geometry for Central Catadioptric Cameras”. In: *International Journal of Computer Vision* 49.1 (Aug. 2002), pp. 23–37. ISSN: 1573-1405. DOI: 10.1023/A:1019869530073 (cit. on p. 8).
- [29] X. Ying and Z. Hu. “Catadioptric camera calibration using geometric invariants”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.10 (Oct. 2004), pp. 1260–1271. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2004.79 (cit. on p. 8).
- [30] D. Scaramuzza, A. Martinelli, and R. Siegwart. “A Toolbox for Easily Calibrating Omnidirectional Cameras”. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2006, pp. 5695–5701. DOI: 10.1109/IROS.2006.282372 (cit. on p. 8).
- [31] C. Mei and P. Rives. “Single View Point Omnidirectional Camera Calibration from Planar Grids”. In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. Apr. 2007, pp. 3945–3950. DOI: 10.1109/ROBOT.2007.364084 (cit. on p. 8).

- [32] D. Strelow, J. Mishler, D. Koes, and S. Singh. “Precise omnidirectional camera calibration”. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. 2001, pp. 689–694. DOI: 10.1109/CVPR.2001.990542 (cit. on p. 8).
- [33] R. Swaminathan, M. D. Grossberg, and S. K. Nayar. “Non-Single Viewpoint Cata-dioptric Cameras: Geometry and Analysis”. In: *International Journal of Computer Vision* 66.3 (Mar. 2006), pp. 211–229. ISSN: 1573-1405. DOI: 10.1007/s11263-005-3220-1 (cit. on p. 8).
- [34] E. Schwalbe. “Geometric modelling and calibration of fisheye lens camera systems”. In: *Proc. 2nd Panoramic Photogrammetry Workshop, Int. Archives of Photogrammetry and Remote Sensing*. Vol. 36. Part 5. 2005, W8 (cit. on p. 8).
- [35] J. Courbon, Y. Mezouar, L. Eckt, and P. Martinet. “A generic fisheye camera model for robotic applications”. In: *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2007, pp. 1683–1688. DOI: 10.1109/IR0S.2007.4399233 (cit. on p. 8).
- [36] B. Micusik and T. Pajdla. “Estimation of omnidirectional camera model from epipolar geometry”. In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*. Vol. 1. June 2003, pp. 485–490. DOI: 10.1109/CVPR.2003.1211393 (cit. on p. 8).
- [37] D. Scaramuzza, A. Martinelli, and R. Siegwart. “A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion”. In: *Fourth IEEE International Conference on Computer Vision Systems (ICVS’06)*. Jan. 2006, pp. 45–45. DOI: 10.1109/ICVS.2006.3 (cit. on p. 8).
- [38] S. Thirthala and M. Pollefeys. “Multi-view geometry of 1D radial cameras and its application to omnidirectional camera calibration”. In: *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*. Vol. 2. Oct. 2005, 1539–1546 Vol. 2. DOI: 10.1109/ICCV.2005.158 (cit. on p. 8).
- [39] J.-Y. Bouguet. “Camera calibration tool-box for matlab”. In: http://www.vision.caltech.edu/bouguetj/calib_doc/ (2002) (cit. on p. 9).
- [40] T. Sato, S. Ikeda, and N. Yokoya. “Extrinsic Camera Parameter Recovery from Multiple Image Sequences Captured by an Omni-Directional Multi-camera System”. In: *Computer Vision - ECCV 2004*. Ed. by T. Pajdla and J. Matas. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 326–340. ISBN: 978-3-540-24671-8 (cit. on p. 9).
- [41] F. Pedersini, A. Sarti, and S. Tubaro. “Multi-camera systems”. In: *IEEE Signal Processing Magazine* 16.3 (May 1999), pp. 55–65. ISSN: 1053-5888. DOI: 10.1109/79.768573 (cit. on p. 9).
- [42] J.-M. Frahm, K. Köser, and R. Koch. “Pose Estimation for Multi-camera Systems”. In: *Pattern Recognition*. Ed. by C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 286–293. ISBN: 978-3-540-28649-3 (cit. on p. 9).
- [43] R. Szeliski. “Image alignment and stitching: A tutorial”. In: *Foundations and Trends in Computer Graphics and Vision* 2.1 (2006), pp. 1–104 (cit. on p. 9).

- [44] M. Brown and D. G. Lowe. “Automatic panoramic image stitching using invariant features”. In: *International Journal of Computer Vision* 74.1 (2007), pp. 59–73 (cit. on p. 9).
- [45] M. Brown, R. I. Hartley, and D. Nister. “Minimal Solutions for Panoramic Stitching”. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. June 2007, pp. 1–8. DOI: 10.1109/CVPR.2007.383082 (cit. on p. 9).
- [46] J. Zaragoza, T. J. Chin, Q. H. Tran, M. S. Brown, and D. Suter. “As-Projective-As-Possible Image Stitching with Moving DLT”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.7 (July 2014), pp. 1285–1298. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2013.247 (cit. on p. 9).
- [47] C. C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin. “Adaptive as-natural-as-possible image stitching”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 1155–1163. DOI: 10.1109/CVPR.2015.7298719 (cit. on p. 9).
- [48] C. H. Chang, Y. Sato, and Y. Y. Chuang. “Shape-Preserving Half-Projective Warps for Image Stitching”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 3254–3261. DOI: 10.1109/CVPR.2014.422 (cit. on p. 9).
- [49] F. Bonin, A. Burguera, and G. Oliver. “Imaging systems for advanced underwater vehicles”. In: *Journal of maritime research: JMR* 8.1 (2011), pp. 65–86 (cit. on p. 10).
- [50] M. Wu, K. Luo, J. Dang, and D. Li. “Underwater image restoration using color correction and non-local prior”. In: *OCEANS 2017 - Aberdeen*. June 2017, pp. 1–5. DOI: 10.1109/OCEANSE.2017.8084916 (cit. on p. 10).
- [51] R. Schettini and S. Corchs. “Underwater Image Processing: State of the Art of Restoration and Image Enhancement Methods”. In: *EURASIP Journal on Advances in Signal Processing* 2010.1 (Apr. 2010), p. 746052. ISSN: 1687-6180. DOI: 10.1155/2010/746052 (cit. on p. 10).
- [52] A. Sedlazeck and R. Koch. “Perspective and Non-perspective Camera Models in Underwater Imaging – Overview and Error Analysis”. In: *Outdoor and Large-Scale Real-World Scene Analysis*. Ed. by F. Dellaert, J.-M. Frahm, M. Pollefeys, L. Leal-Taixé, and B. Rosenhahn. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 212–242. ISBN: 978-3-642-34091-8 (cit. on p. 10).
- [53] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh. “Flat Refractive Geometry”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.1 (Jan. 2012), pp. 51–65. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.105 (cit. on p. 10).
- [54] T. Łuczyński, M. Pflugstorn, and A. Birk. “The Pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings”. In: *Ocean Engineering* 133 (2017), pp. 9–22. ISSN: 0029-8018. DOI: <https://doi.org/10.1016/j.oceaneng.2017.01.029> (cit. on p. 10).
- [55] C. Kunz and H. Singh. “Hemispherical refraction and camera calibration in underwater vision”. In: *OCEANS 2008*. Sept. 2008, pp. 1–7. DOI: 10.1109/OCEANS.2008.5151967 (cit. on p. 10).
- [56] F. Menna, E. Nocerino, F. Fassi, and F. Remondino. “Geometric and Optic Characterization of a Hemispherical Dome Port for Underwater Photogrammetry”. In: *Sensors* 16.48 (2016). ISSN: 1424-8220. DOI: 10.3390/s16010048 (cit. on p. 10).

- [57] A. Yamashita, R. Kawanishi, T. Koketsu, T. Kaneko, and H. Asama. “Underwater sensing with omni-directional stereo camera”. In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. Nov. 2011, pp. 304–311. DOI: 10.1109/ICCVW.2011.6130257 (cit. on p. 10).
- [58] M. González-Rivero, P. Bongaerts, O. Beijbom, O. Pizarro, A. Friedman, A. Rodríguez-Ramírez, B. Upcroft, D. Laffoley, D. Kline, C. Bailhache, R. Vevers, and O. Hoegh-Guldberg. “The Catlin Seaview Survey - kilometrescale seascape assessment, and monitoring of coral reef ecosystems”. In: *Aquatic Conservation: Marine and Freshwater Ecosystems* 24.S2 (2014), pp. 184–198. DOI: 10.1002/aqc.2505 (cit. on p. 10).
- [59] S. Negahdaripour, H. Zhang, P. Firoozfam, and J. Oles. “Utilizing panoramic views for visually guided tasks in underwater robotics applications”. In: *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No.01CH37295)*. Vol. 4. 2001, 2593–2600 vol.4. DOI: 10.1109/OCEANS.2001.968408 (cit. on p. 10).
- [60] M. Carrozzino and M. Bergamasco. “Beyond virtual museums: Experiencing immersive virtual reality in real museums”. In: *Journal of Cultural Heritage* 11.4 (2010), pp. 452–458. ISSN: 1296-2074. DOI: <https://doi.org/10.1016/j.culher.2010.04.001> (cit. on p. 10).
- [61] D. Ribas, P. Ridao, J. D. Tardós, and J. Neira. “Underwater SLAM in man-made structured environments”. In: *Journal of Field Robotics* 25.11-12 (2008), pp. 898–921. DOI: 10.1002/rob.20249 (cit. on p. 11).
- [62] A. Mallios et al. “Sonar scan matching for simultaneous localization and mapping in confined underwater environments”. PhD thesis. Universitat de Girona, 2014 (cit. on p. 11).
- [63] A. Palomer, P. Ridao, and D. Ribas. “Multibeam 3D Underwater SLAM with Probabilistic Registration”. In: *Sensors* 16.4 (2016). ISSN: 1424-8220. DOI: 10.3390/s16040560 (cit. on p. 11).
- [64] G. Vallicrosa and P. Ridao. “H-SLAM: Rao-Blackwellized Particle Filter SLAM Using Hilbert Maps”. In: *Sensors* 18.5 (2018). ISSN: 1424-8220. DOI: 10.3390/s18051386 (cit. on p. 11).
- [65] A. Carrera, N. Palomeras, N. Hurtós, P. Kormushev, and M. Carreras. “Cognitive system for autonomous underwater intervention”. In: *Pattern Recognition Letters* 67 (2015). Cognitive Systems for Knowledge Discovery, pp. 91–99. ISSN: 0167-8655. DOI: <https://doi.org/10.1016/j.patrec.2015.06.010> (cit. on p. 11).
- [66] N. Palomeras, A. Peñalver, M. Massot-Campos, G. Vallicrosa, P. L. Negre, J. J. Fernández, P. Ridao, P. J. Sanz, G. Oliver-Codina, and A. Palomer. “I-AUV docking and intervention in a subsea panel”. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Sept. 2014, pp. 2279–2285. DOI: 10.1109/IRoS.2014.6942870 (cit. on p. 11).
- [67] D. Youakim, P. Ridao, N. Palomeras, F. Spadafora, D. Ribas, and M. Muzzupappa. “MoveIt!: Autonomous Underwater Free-Floating Manipulation”. In: *IEEE Robotics Automation Magazine* 24.3 (Sept. 2017), pp. 41–51. ISSN: 1070-9932. DOI: 10.1109/MRA.2016.2636369 (cit. on p. 11).

- [68] P. Cieslak, P. Ridao, and M. Giergiel. “Autonomous underwater panel operation by GIRONA500 UVMS: A practical approach to autonomous underwater manipulation”. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. May 2015, pp. 529–536. DOI: 10.1109/ICRA.2015.7139230 (cit. on p. 11).
- [69] A. Palomer, P. Ridao, D. Youakim, D. Ribas, J. Forest, and Y. Petillot. “3D Laser Scanner for Underwater Manipulation.” In: *Sensors* 18.4 (2018), pp. 1–14. DOI: 10.3390/s18041086 (cit. on p. 11).
- [70] E. Hernández, M. Carreras, J. Antich, P. Ridao, and A. Ortiz. “A topologically guided path planner for an auv using homotopy classes”. In: *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2011, pp. 2337–2343 (cit. on p. 11).
- [71] E. Galceran Yebenes. “Coverage path planning for autonomous underwater vehicles”. PhD thesis. Universitat de Girona, 2014 (cit. on p. 11).
- [72] J. D. Hernández, K. Istenič, N. Gracias, N. Palomeras, R. Campos, E. Vidal, R. García, and M. Carreras. “Autonomous Underwater Navigation and Optical Mapping in Unknown Natural Environments”. In: *Sensors* 16.8 (2016), p. 1174 (cit. on p. 11).
- [73] M. Carreras, J. D. Hernández, E. Vidal, N. Palomeras, D. Ribas, and P. Ridao. “Sparus II AUV-A Hovering Vehicle for Seabed Inspection”. In: *IEEE Journal of Oceanic Engineering* (2018) (cit. on p. 11).
- [74] D. Ribas, N. Palomeras, P. Ridao, M. Carreras, and A. Mallios. “Girona 500 AUV: From Survey to Intervention”. In: *IEEE/ASME Transactions on Mechatronics* 17.1 (2012), pp. 46–53 (cit. on p. 11).
- [75] K. Istenič, V. Ila, L. Polok, N. Gracias, and R. García. “Mission-time 3D Reconstruction with Quality Estimation”. In: *Proc. of the MTS/IEEE OCEANS 2017 Conference*. MTS/IEEE. Aberdeen, UK, June 2017 (cit. on p. 11).
- [76] C. O. Ancuti, C. Ancuti, C. D. Vleeschouwer, L. Neumann, and R. Garcia. “Color transfer for underwater dehazing and depth estimation”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. Sept. 2017, pp. 695–699. DOI: 10.1109/ICIP.2017.8296370 (cit. on p. 11).
- [77] R. Prados, R. Garcia, N. Gracias, J. Escartin, and L. Neumann. “A Novel Blending Technique for Underwater Giga-Mosaicing”. In: *IEEE Journal of Oceanic Engineering* 37.4 (2012), pp. 626–644 (cit. on p. 11).
- [78] R. Campos, N. Gracias, and P. Ridao. “Underwater Multi-Vehicle Trajectory Alignment and Mapping Using Acoustic and Optical Constraints”. In: *Sensors* 16.3 (2016), p. 387. ISSN: 1424-8220. DOI: 10.3390/s16030387 (cit. on p. 11).
- [79] R. Prados, R. García, N. Gracias, L. Neumann, and H. Vagstol. “Real-time fish detection in trawl nets”. In: *OCEANS 2017 - Aberdeen*. June 2017, pp. 1–5. DOI: 10.1109/OCEANSE.2017.8084760 (cit. on p. 11).
- [80] J. Kalwa, M. Carreiro-Silva, F. Tempera, J. Fontes, R. Santos, M.-C. Fabri, L. Brignone, P. Ridao, A. Birk, T. Glotzbach, M. Caccia, J. Alves, and A. Pascoal. “The MORPH concept and its application in marine research”. In: *OCEANS - Bergen, 2013 MTS/IEEE*. June 2013, pp. 1–8. DOI: 10.1109/OCEANS-Bergen.2013.6607988 (cit. on p. 104).

-
- [81] D. Ribas, N. Palomeras, P. Ridao, M. Carreras, and A. Mallios. “Girona 500 AUV: From Survey to Intervention”. In: *Mechatronics, IEEE/ASME Transactions on* 17.1 (Feb. 2012), pp. 46–53. ISSN: 1083-4435. DOI: 10.1109/TMECH.2011.2174065 (cit. on p. 104).
- [82] C. Petrioli, R. Petroccia, D. Spaccini, A. Vitaletti, T. Arzilli, D. Lamanna, A. Galizial, and E. Renzi. “The SUNRISE GATE: Accessing the SUNRISE federation of facilities to test solutions for the Internet of Underwater Things”. In: *2014 Underwater Communications and Networking (UComms)*. Sept. 2014, pp. 1–4. DOI: 10.1109/UComms.2014.7017144 (cit. on p. 106).