Universitat de Girona

UNDERWATER IMAGE-BASED 3D RECONSTRUCTION WITH QUALITY ESTIMATION

Klemen Istenič

Universitat de Girona

Doctoral Thesis

Underwater Image-Based 3D Reconstruction With Quality Estimation

---

Klemen Istenič

2021

Universitat de Girona

Doctoral Thesis

Underwater Image-Based 3D Reconstruction With Quality Estimation

Klemen Istenič

2021

Doctoral Program in Technology

Supervised by:

Nuno Gracias
Rafael Garcia

Thesis submitted to the University of Girona for the degree of

Doctor

CERTIFICATE OF THESIS DIRECTION

Dr. Nuno Gracias, member of the *Departament d'Arquitectura i Tecnologia de Computadors* of *Universitat de Girona*, and Dr. Rafael Garcia, member of the *Departament d'Arquitectura i Tecnologia de Computadors* of *Universitat de Girona*,

DECLARE:

That the work entitled *Underwater Image-Based 3D Reconstruction With Quality Estimation* presented by Klemen Istenič to obtain the degree in Doctor of Philosophy has been developed under our supervision and fulfills the requirements to obtain the International Mention.

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

Girona, February 2021

<div style="text-align: right;">

_____

Dr. Nuno Gracias

</div>

<div style="text-align: right;">

_____

Dr. Rafael Garcia

</div>

To my mother Irena
who never got the chance to get bored reading this thesis

Acknowledgments

In the beginning, I naively had one goal in mind. To travel the world and experience things I would not have had the chance to otherwise. After a few more years than I expected, I look back, the world is nearly traveled and I have a long list of people to thank...

First and foremost, I would like to express my gratitude to my supervisors, Dr. Rafael Garcia and Dr. Nuno Gracias. Rafa and Nuno, thank you for your support during these past several years. I consider it one of my greatest fortunes that you both decided to mentor me. Aside from providing me with life-changing opportunities, teaching me an incomprehensible amount of lessons about research, and showing me that there is more to research than just research, most of all you have shown me what it truly means to support someone unconditionally. You opened doors for me when they needed to be opened, you always gave me the freedom to do things my way, and most importantly, you understood and supported me when it mattered most. I will never forget that.

I thank all the members of the CIRS and ViCOROB research group. Juan David , you introduced me to the ins and outs of the lab and helped me at the beginning of my journey. Your *"todo va a salir bien"* has stuck with me to this day. Ricard C., you always let me pick your brain and learn from you without ever making me feel like I was bothering you. Jep and Eduard, involving me in your research has given me the opportunity to learn things I would have never known otherwise. Also, thank you Dina, Khadija, and Joseta for your support and random conversations that always left me feeling better, Quintana and Ricard P., for always helping me with all the technical things and Laszlo for a number of insightful discussions and suggestions. A big thank you also goes to Marc Carreras. Not only for all the effort in collecting all those datasets, but also for never letting me forget it and, more importantly, for making me feel welcome and supported at CIRS. It is more than necessary for me to also thank Lluis and Carles for your help with the robots and Joseta, Mireia, Anna, Olga and Bego for your constant help with all the bureaucratic pickles I seem to constantly find myself in. My experience in the lab would not be the same without the environment created by Pere Ridao and the rest of you: Guillem, Albert, Tali, Patryk, Narcis, David, Eduardo, Miguel, Angelos, Eric, Roger and Bruno.

I would also like to thank Viorela Ila for hosting and mentoring me during my research stay at Australian National University. Your motivation and countless explanations helped me to improve my understanding in vision and robotics. Lukas, thank you for your insights into coding and all the little tricks you taught me along the way. A big thanks also goes to Mina, Monty, Cristian and Perla for making my stay in Canberra so enjoyable.

This thesis would not have been the same without the support of Javier Escartín.

Your leadership on the SUBSAINTES cruise and ever positive attitude in our subsequent collaboration helped me to carry out a significant part of the research covered in this thesis. I must also thank Aurélien for his help and willingness to share his knowledge, as well as all the other people who participated in the scientific missions in which I took part. Each mission taught me something new about research and gave me at least one extremely memorable story to tell.

I also had the pleasure of participating in Robocademy, a Marie Curie Research Training Network (FP7-PEOPLE-2013-ITN-608096). I am grateful to everyone involved, the organizers and fellow students and their supervisors for making the three years full of fun and memorable events and giving us the opportunity to gain knowledge on a wide range of topics.

I thank my friends for the support and friendship I needed. Friends from back home (too many to list here, but you know who you are!), thank you for always welcoming me when I showed up in Slovenia, as if we had seen each other the day before. When it came down to it, I could always rely on your honest advice. My so-called second family in Girona - Ana Maria, Jordi and Anna. You were there for most of this journey and were always more than understanding no matter what I presented you with. Without your support it would have been really hard to persevere and get to this point. Ana Maria, I have lost count of how many conversations we have had about the moment we will both finish our dissertations - here we finally are! The ViBOT group (Maruf, Vibhav, Oksana, Dani, and Igor), even though we are scattered around, I always knew I could reach out whenever I needed advice that only you guys could give me. And then there are many others in Girona, Canberra and San Diego, who have supported me, listened to my complaining and always managed to put a smile on my face when I needed it most. I hope that each and every one of you will find yourselves here.

These past several years have not been an easy ride, both academically and personally. I thank Karen for sticking by my side and always managing to lift me up and make me realize what was important in life, even when things were hard and life presented challenges.

I especially thank my mom, dad, brother and aunt as well as my grandparents (nona and nono). You have given me unconditional love and care and taught me perseverance. Without your moral and emotional support, I would not have made it this far and experienced the wonderful things I have. I am also grateful to my other family members who have supported me along the way.

Finally, I would like to thank the members of my PhD committee for their valuable comments that helped improve this thesis, as well as the many anonymous reviewers of the various publications derived from this work.

Mami in Nono - Medtem ko se drugim lahko osebno zahvalim, se vama na žalost nikoli ne bom mogel. Brez naših neštetih pogovorov nikoli ne bi razvil take želje po znanju in razumevanju zakaj in kako vse deluje. Za to in za vse še enkrat hvala!

List of Publications

Publications in the compendium

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

- Juan David Hernández, **Klemen Istenič**, Nuno Gracias, Narcis Palomeras, Ricard Campos, Eduard Vidal, Rafael Garcia, and Marc Carreras. "Autonomous underwater navigation and optical mapping in unknown natural environments". In: *Sensors* 16.8 (2016), page 1174. DOI: [10.3390/s16081174](10.3390/s16081174)
  Quality index: JCR2016 Instruments & Instrumentation, Impact Factor: 2.677, Q1 (10/58).
  Contributions: This publication presented an approach for autonomous exploration and subsequent 3D reconstruction of previously unexplored underwater areas. The procedures for collision-free online path-planning and an image-based pipeline capable of producing photo-realistic textured 3D models are presented and tested using Sparus II AUV in a challenging real-world natural environment.
  Author contributions: The author of this thesis was responsible for the design and implementation of the 3D reconstruction pipeline and for the analysis of its results on different data sets. In addition, the author contributed to the design of the adjustable multi-camera system used in the experiments conducted jointly by the authors. The 3D reconstruction part of the manuscript was written by the author of this thesis while Juan David Hernández was responsible for the path-planning part

- **Klemen Istenič**, Nuno Gracias, Aurélien Arnaubec, Javier Escartín, and Rafael Garcia. "Scale Accuracy Evaluation of Image-Based 3D Reconstruction Strategies Using Laser Photogrammetry". In: *Remote Sensing* 11.18 (2019), page 2093. DOI: [10.3390/rs11182093](10.3390/rs11182093)
  Quality index: JCR2019 Remote Sensing, Impact Factor: 4.509, Q2 (9/30).
  Contributions: The paper presented two novel methods for scaling SfM-based 3D models using commonly available laser scalers. The methods have been extensively evaluated using data generated from a realistic 3D model as well as using data collected during an oceanographic cruise in 2017.
  Author contributions: The author of this thesis is the main developer of the two proposed approaches. He was also responsible for the experimental evaluation and wrote the substantial part of the article. Nuno Gracias, Aurélien Arnaubec and

Rafael Garcia contributed to the design of the methods and approach for the evaluation, while Javier Escartín supervised the collection of the data sets and provided a geological perspective on the applicability of the approaches.

- **Klemen Istenič**, Nuno Gracias, Aurélien Arnaubec, Javier Escartín, and Rafael Garcia. "Automatic scale estimation of structure from motion based 3D models using laser scalers in underwater scenarios". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (2020), pages 13–25. DOI: 10.1016/j.isprsjprs.2019.10.007
  Quality index: JCR2019 Remote Sensing, Impact Factor: 7.319, Q1 (3/30).
  Contributions: This publication introduced a novel framework for scale estimation that extends our earlier work on scaling methods. In addition to the proposed framework, which is capable of estimating the uncertainty of scale estimates, a new automatic laser spot detection method has been introduced. The main novelty of the approach is the process of estimating uncertainties based on the recovery of characteristic shapes of laser spots. By exploiting these novelties, we were able to perform a comprehensive analysis of scale errors for various image-based 3D reconstruction and navigation fusion strategies in underwater applications. We provided an insight into the best possible reconstruction strategies as well as precise quantification of the uncertainty of the estimated scale for models created using the *SUBSAINTES* data sets.
  Author contributions: The author of this thesis is the main developer of the methods and their extensions. He also supervised the acquisition of the optical data used in the analysis during the *SUBSAINTES* cruise, and is the main responsible for the evaluation and writing of the article. Nuno Gracias, Aurélien Arnaubec and Rafael Garcia contributed to the design and improvements of the scaling methods and evaluation. Javier Escartín provided guidance on the geological perspective of the usefulness of our approach and supervised the collection of the data sets as the Chef de Mission of the *SUBSAINTES* cruise.

- **Klemen Istenič**, Viorela Ila, Lukáš Polok, Nuno Gracias, and Rafael García. "Mission-time 3D reconstruction with quality estimation". In: *OCEANS 2017-Aberdeen*. IEEE. 2017, pages 1–9. DOI: 10.1109/OCEANSE.2017.8084708
  Quality index: Not a journal.
  Contributions: This paper presented a SfM-based system capable of providing a globally consistent 3D reconstruction together with an estimate of the uncertainty in mission-time or shortly thereafter. The publication is a result of the work during the author's research stay at the ARC Centre of Excellence for Robotic Vision at the ANU. The published work received the Norman Miller Prize for the Best Student Poster Award at the MTS/IEEE OCEANS 2017 Conference.
  Author contributions: The author, with the help of Viorela Ila, are the main developers of the algorithm. Lukáš Polok helped extensively with the integration of SLAM ++ library and gave guidance on the software development of the approach. Rafael Garcia and Nuno Gracias provided feedback on the approach and reviewed the paper before publication.

Publications derived from this thesis

The work developed in this thesis also led to the following publications (to be updated and checked):

- Juan David Hernández, **Klemen Istenič**, Nuno Gracias, Rafael García, Pere Ridao, and Marc Carreras. "Autonomous seabed inspection for environmental monitoring". In: *Robot 2015: Second Iberian Robotics Conference.* Springer. 2016, pages 27–39. DOI: 10.1007/978-3-319-27149-1_3

- Viorela Ila, Lukáš Polok, Marek Solony, and **Klemen Istenič**. "Fast incremental bundle adjustment with covariance recovery". In: *2017 International Conference on 3D Vision (3DV).* IEEE. 2017, pages 175–184. DOI: 10.1109/3DV.2017.00029

- Josep Bosch, Nuno Gracias, Pere Ridao, David Ribas, **Klemen Istenič**, Rafael Garcia, and Irena Radic Rossi. "Immersive Touring for Marine Archaeology. Application of a New Compact Omnidirectional Camera to Mapping the Gnalić shipwreck with an AUV". in: *Iberian Robotics conference.* Springer. 2017, pages 183–195. DOI: 10.1007/978-3-319-70833-1_16

- Eduard Vidal, Juan David Hernández, **Klemen Istenič**, and Marc Carreras. "Online view planning for inspecting unexplored underwater structures". In: *IEEE Robotics and Automation Letters* 2.3 (2017), pages 1436–1443. DOI: 10.1109/LRA.2017.2671415

- Javier Escartin, Frédérique Leclerc, Feuillet Nathalie, Anne Le Friant, Jérémy Billant, Jean-Arthur L Olive, Mathilde Henri, Muriel Andreani, Aurelien Arnaubec, Alexandre Dano, et al. "Mapping the Mw6. 3 2004 Les Saintes earthquake seafloor rupture with deep-sea vehicles: Length, displacement, nature, and links between coseismic deformation and erosion/sedimentation". In: *AGU Fall Meeting Abstracts.* 2018

- Jérémy Billant, Frédérique Leclerc, Javier Escartin, Nuno Gracias, **Klemen Istenič**, Aurelien Arnaubec, Feuillet Nathalie, and Christine Deplus. "Submarine expression of the earthquake cycle: morphotectonics analysis of the normal Roseau fault (Lesser Antilles, France) from high resolution bathymetry and 3D photogrammetric models to reconstruct its slip history". In: *AGU Fall Meeting Abstracts.* 2018

- Eduard Vidal, Juan David Hernández, **Klemen Istenič**, and Marc Carreras. "Optimized environment exploration for autonomous underwater vehicles". In: *2018 IEEE International Conference on Robotics and Automation (ICRA).* IEEE. 2018, pages 6409–6416. DOI: 10.1109/ICRA.2018.8460919

- Josep Bosch, **Klemen Istenič**, Nuno Gracias, Rafael Garcia, and Pere Ridao. "Omnidirectional Multicamera Video Stitching Using Depth Maps". In: *IEEE Journal of Oceanic Engineering* (2019). DOI: 10.1109/JOE.2019.2924276

- Eduard Vidal, Narcís Palomeras, **Klemen Istenič**, Juan David Hernández, and Marc Carreras. "Two-Dimensional Frontier-Based Viewpoint Generation for Exploring and Mapping Underwater Environments". In: *Sensors* 19.6 (2019), page 1460. DOI: 10.3390/s19061460

- Maria Montseny, Cristina Linares, Núria Viladrich, Alejandro Olariaga, Marc Carreras, Narcís Palomeras, Nuno Gracias, **Klemen Istenič**, Rafael Garcia, Stefano Ambroso, et al. "First attempts towards the restoration of gorgonian populations on the Mediterranean continental shelf". In: *Aquatic Conservation: Marine and Freshwater Ecosystems* 29.8 (2019), pages 1278–1284. DOI: `10.1002/aqc.3118`

- Javier Escartin, Jeremy Billant, Frederique Leclerc, Jean-Arthur Olive, **Klemen Istenič**, Nuno Gracias, Aurelien Arnaubec, Rafael Garcia, Christine Deplus, Nathalie Feuillet, et al. "Seafloor rupture of a normal fault: geometry, displacement, and links between erosion/sedimentation and seismicity." In: *Geophysical Research Abstracts*. Volume 21. 2019

- Jérémy Billant, Frédérique Leclerc, Javier Escartin, Nuno Gracias, **Klemen Istenič**, Rafael Garcia, Aurélien Arnaubec, Alexandre Dano, and Claire Marchand. "Development of a Unity package allowing GIS-like mapping in a Virtual Reality environment." In: *Geophysical Research Abstracts*. Volume 21. 2019

- Jérémy Billant, Júlia Bozzinio, Frédérique Leclerc, Javier Escartin, Nuno Gracias, **Klemen Istenič**, Aurelien Arnaubec, and Rafael Garcia. "Performing submarine field survey without scuba gear using GIS-like mapping in a Virtual Reality environment". In: *OCEANS 2019-Marseille*. IEEE. 2019, pages 1–6. DOI: `10.1109/OCEANSE.2019.8867408`

- Eduard Vidal, Narcís Palomeras, **Klemen Istenič**, Nuno Gracias, and Marc Carreras. "Multisensor online 3D view planning for autonomous underwater exploration". In: *Journal of Field Robotics* (2020). DOI: `10.1002/rob.21951`

Other publications

Parallel work at the time of this thesis led to the following publications:

- Josep Bosch, Nuno Gracias, Pere Ridao, **Klemen Istenič**, and David Ribas. "Close-range tracking of underwater vehicles using light beacons". In: *Sensors* 16.4 (2016), page 429. DOI: `10.3390/s16040429`

- Eduardo Ochoa, Nuno Gracias, **Klemen Istenič**, Rafael García, Josep Bosch, and Patryk Cieslak. "Allowing untrained scientists to safely pilot ROV's: Early collision detection and avoidance using omnidirectional vision". In: *OCEANS Global 2020*. IEEE. 2020

Finally, I also had the pleasure to co-supervise two MSc theses:

- *Early collision detection using panoramic cameras for underwater robotics applications.* Thesis written by Eduardo Ochoa for the degree of Master of Science, Erasmus Mundus Masters in VIsion and roBOTics (VIBOT), University of Burgundy, University of Girona and Heriot-Watt University, 2019.

- *Combined use of a frame camera and a linear pushbroom camera for hyperspectral underwater mapping.* Thesis written by Peeyush kumar for the degree of Master of Science, Institute of Geodesy and Geoinformation, University of Bonn, 2020.

ACRONYMS

ANU   Australian National University.
AUV   autonomous underwater vehicle.
BA   bundle adjustment.
CIRS   Underwater Robotics Research Center.
DL   deep learning.
FUM   fully unconstrained method.
GIS   geographic information system.
GPS   global positioning system.
IFREMER   French Research Institute for Exploitation of the Sea.
IPGP   Paris Institute of Earth Physics.
LBL   long baseline.
MC   Monte Carlo.
NLS   non-linear least squares.
PCM   partially constrained method.
RMSE   root mean square error.
ROV   remote operated vehicle.
SfM   structure from motion.
SLAM   simultaneous localization and mapping.
UdG   University of Girona.
USBL   ultra-short baseline.
UUV   unmanned underwater vehicle.
UVL   Underwater Vision Lab.
V-SLAM   visual SLAM.
ViCOROB   Computer Vision and Robotics Research Group.
VR   virtual reality.

CONTENTS

LIST OF FIGURES

xviii

Abstract

Despite the undeniable importance of the marine ecosystem, vast areas of the seabed remain largely unexplored. Accurate and detailed 3D models of the environment yield high added value to any marine survey, as such results convey immense information easily interpretable by humans. The wealth of information enables experts (biologists, archaeologists and geologists, among others) to carry out further in-depth investigations of the areas of interest after the missions and can also serve as a base map for long-term environmental monitoring.

This thesis presents a comprehensive investigation and use of image-based 3D reconstruction techniques for practical applications in underwater environment. It focuses on the research and development of resources for accurate scaling and uncertainty estimation of 3D models for scientific purposes based on data acquired with monocular or un-synchronized camera systems in difficult-to-access GPS-denied (underwater) environments.

Two novel methods for automatic scaling of structure from motion (SfM)-based 3D models using commonly available laser scalers were proposed, namely the fully unconstrained method (FUM) and the partially constrained method (PCM). The methods are able to compensate the geometry of the scene together with the geometry of the laser scaler setup. Furthermore, they also eliminate the need for the manual, time-consuming and error-prone step of image-model point association. The approaches are therefore no longer limited to certain camera poses and/or the roughness of the observed scene when estimating the scale. In addition, the confidence level for each of the scale estimates is also evaluated independently by propagating the uncertainties associated with image features and laser spot detection using a Monte Carlo simulation. Both methods are considered universal as they are not tied to data collected with specific sensors or hardware and can even be used with diver-based surveys. Moreover, the minimal requirement for knowledge of laser configurations allows the methods to be used with a wide variety of legacy data. The methods have been fully validated by a series of generated data sets based on a real 3D model of a deep-sea hydrothermal vent, and their applicability in real scenarios has been demonstrated by a data set from the *SUBSAINTES* scientific cruise.

Utilizing the newly proposed scaling methods, a scale error analysis on large-scale models of deep-sea underwater environments was performed. It evaluated and determined the effects of the most commonly used image-navigation fusion strategies. Two distinct underwater survey scenarios encapsulating realistic deep-sea field working condi-

tions were considered. The study demonstrated that the surveys with multiple overlaps of non-sequential images result in a nearly identical solution regardless of the strategy, while surveys with weakly connected sequentially acquired images are prone to produce broad-scale deformation (doming effect) when navigation is not included in the optimization. Additionally, we evaluated the effects of multi-survey data sets. We showed that the introduction of auxiliary displacement parameters in the optimization successfully accounted for offset changes present in the underwater USBL-based navigation data, and thus minimize the effect of contradicting navigation priors.

In addition, the analysis also provided a quantitative estimate of the model scaling which is a requirement for applications that demand precise measurements (such as distances, areas, volumes and others). Since the analysis was performed using the *SUBSAINTES* data set, it enabled the geologists from Paris Institute of Earth Physics to use the corresponding models in their respective studies. Previously performed analyses using a single model are now being performed with information from more than 30 different sites with 41 different models.

To facilitate the detection and characterization of the uncertainty of laser spots, we have also developed a method that robustly detects the centers of laser beams by taking into account the texture of the scene. At the same time, the uncertainty of each of these detections is determined based on the recovered characteristic shapes of the laser spots. The versatility and robustness of the proposed method has been demonstrated on several data sets.

The dissertation also introduces the developed image-based 3D reconstruction framework. The end-to-end pipeline was built by integrating several open-source solutions and by developing additional functionalities (e.g. navigation fusion, soft enforcement of known relative constraints, underwater color correction). It enables the creation of photo-realistic textured 3D models based on optical and navigation data and is independent of a specific robot or diver-based platform, camera or mission. We report on an extensive use of the framework in real-world applications with six different underwater robotic platforms and numerous different cameras and camera types. The created models were used in numerous collaborations with different scientists and researchers. They enabled accurate measurements in scientific fields, the generation of superior omnidirectional video, augmented reality applications or as proof that algorithms like mission planning have successfully ensured satisfactory data acquisition. The processing of data from the *SUBSAINTES* scientific cruise led to the creation of one of the largest collections of image-derived underwater 3D models ever made by deep-sea vehicles for geological purposes.

Finally, we also made a contribution by demonstrating the feasibility of mission-time 3D reconstruction and uncertainty estimation. We proposed a novel SfM-based system capable of producing a globally consistent 3D reconstruction together with an estimate of its uncertainty while the robot is still in the water or shortly after. The system integrates an incremental way of solving the bundle adjustment problem as well as fast covariance recovery and an outlier rejection scheme based on a dual-map approach. This enables the possibility of quality-conscious data acquisitions, which consequently will not only increase the quality of the final data collected and the efficiency of the survey, but also reduces the possibility of conducting unsatisfactory optical surveys.

Resum

Malgrat la indubtable importància dels ecosistemes marins, grans àrees del fons del mar permaneixen encara inexplorades. Models de l'entorn 3D precisos i detallats proporcionen un alt valor afegit en qualsevol investigació marina, donat que els seus resultats transmeten una extensa informació fàcilment interpretable pels humans.La riquesa d'aquesta informació permet als experts (biòlegs, arqueòlegs y geòlegs, entre d'altres) portar a terme investigacions més profundes de les àrees d'interès després de las missions i també pot servir com a mapa base per a la vigilància ambiental a llarg plaç.

Aquesta tesis presenta una investigació exhaustiva i l'ús de tècniques de reconstrucció tridimensional basades en imatges per a aplicacions pràctiques en el medi ambient submarí. Se centra en la investigació i el desenvolupament de recursos per a l'escalat precís i l'estimació de la incertesa dels models tridimensionals amb finalitats científiques basades en dades adquirides amb sistemes de càmeres monoculars o no sincronitzades en entorns (subaquàtics) de difícil accès.

Es van proposar dos mètodes nous per a l'escalat automàtic de models 3D basats en SfM utilitzant mesuradors làser comunament disponibles, principalment, el FUM i el PCM. Aquests mètodes són capaços de compensar la geometria de l'escena juntament amb la geometria del mesurador làser. A més, també eliminen la necessitat del pas manual, que porta temps i és propens a errors, de l'associació de punts de models d'imatges. Per tant, els enfocaments ja no es limiten a determinades posis de la càmera i/o a la rugositat de l'escena observada a l'estimar l'escala. A més, el nivell de confiança per a cadascuna de les estimacions de l'escala també s'avalua de forma independent mitjançant la propagació de les incerteses associades a les característiques de la imatge i la detecció de punts de làser utilitzant una simulació de Monte Carlo. Tots dos mètodes es consideren universals ja que no estan relacionats amb dades recollides amb sensors o hardware específics i fins i tot es poden utilitzar amb dades adquirides per submarinistes. A més, aquests mètodes requereixen tan poc coneixement sobre la configuració dels làsers, que és possible utilitzar-los amb dades ja existents, adquirides en altres campanyes. Els mètodes han estat plenament validats per una sèrie de conjunts de dades generats sobre la base d'un model real en 3D d'un respirador hidrotermal en aigües profundes, i la seva aplicabilitat en escenaris reals ha quedat demostrada per un conjunt de dades del creuer científic *SUBSAINTES*.

Utilitzant els mètodes d'escalat proposats, es va realitzar una anàlisi d'error d'escala en models a gran escala dels entorns submarins de les profunditats. Es van avaluar i determinar els efectes de les estratègies de fusió de navegació d'imatges més utilitzades. Es van

considerar dos escenaris d'estudi submarí diferents que encapsulaven condicions de treball realistes en el camp de les profunditats marines. L'estudi va demostrar que els estudis amb múltiples superposicions d'imatges no seqüencials donen com a resultat una solució gairebé idèntica independentment de l'estratègia, mentre que els estudis amb imatges adquirides seqüencialment i dèbilment connectades són propensos a produir una deformació a gran escala (efecte de cúpula) quan la navegació no està inclosa en l'optimització. A més a més, vam avaluar els efectes dels conjunts de dades d'estudis múltiples. Vam demostrar que la introducció de paràmetres de desplaçament auxiliars en l'optimització va tenir en compte amb èxit els canvis de desplaçament presents en les dades de navegació submarina basats en USBL, i per tant, va minimitzar l'efecte de inconsistències en les dades de navegació.

A més a més, l'anàlisi també va proporcionar una estimació quantitativa de l'escalament del model, que és un requisit per a les aplicacions que requereixen mesures precises (com ara distàncies, àrees, volums i altres). Atès que l'anàlisi es va realitzar utilitzant el conjunt de dades de *SUBSAINTES*, va permetre als geòlegs de l'Institut de Física del Globus de París utilitzar els models corresponents en els seus respectius estudis. Les anàlisis realitzades anteriorment amb un sol model s'estan realitzant ara amb informació de més de 30 de llocs diferents amb 41 de models diferents.

Per a facilitar la detecció i caracterització de la incertesa dels punts de làser, també hem desenvolupat un mètode que detecta de forma robusta els centres dels raigs làser tenint en compte la textura de l'escena. Al mateix temps, la incertesa de cadascuna d'aquestes deteccions es determina en base a les formes característiques recuperades dels punts làser. La versatilitat i la robustesa del 'mètode proposat han estat demostrades en diversos conjunts de dades. La tesi també introdueix el marc desenvolupat de reconstrucció tridimensional basat en imatges. La línia de processament de d'extrem a extrem es va construir integrant diverses solucions de codi obert i desenvolupant funcionalitats addicionals (per exemple, fusió de navegació, aplicació suau de les limitacions relatives conegudes, correcció del color submarí). Permet la creació de models 3D texturitzats foto-realistes basats en dades òptics i de navegació i és independent d'una plataforma robòtica específica, càmera o missió. S'ha realitzat un ampli ús del marc en aplicacions del món real amb sis plataformes robòtiques submarines diferents i nombroses càmeres i tipus de càmeres diferents. Els models creats es van utilitzar en nombroses col·laboracions amb diferents científics i investigadors. Van permetre realitzar mesures precises en els camps científics, generar vídeo omnidireccional de qualitat superior, aplicacions de realitat augmentada o com a prova que els algoritmes com la planificació de missions han assegurat satisfactòriament l'adquisició de dades. El processament de les dades del creuer científic *SUBSAINTES* va portar a la creació d'una de les majors col·leccions amb finalitat geològica mai reunides de models 3D submarins derivats d'imatges realitzades per vehicles d'alta mar.

Finalment, també contribuïm demostrant la viabilitat de la reconstrucció tridimensional en temps de missió i l'estimació de la incertesa. Vam proposar un nou sistema basat en el SfM capaç de produir una reconstrucció 3D globalment consistent juntament amb una estimació de la seva incertesa mentre el robot està encara a l'aigua o poc després. El sistema integra una forma incremental de resoldre el problema de l'bundle adjustment, així com una ràpida recuperació de la covariància i un esquema de rebuig de valors atípics basat en un enfocament de doble mapa. Això permet la possibilitat d'adquirir dades de qualitat, la qual cosa no només augmenta la qualitat de les dades finals recollides i l'eficiència del mesurament, sinó que també redueix la possibilitat de realitzar mesures òptiques insatisfactòries.

Resumen

A pesar de la indudable importancia de los ecosistemas marinos, grandes áreas del fondo del mar permanecen aún inexploradas. Los modelos 3D precisos del entorno proporcionan un alto valor añadido en cualquier investigación marina, puesto que sus resultados transmiten una vasta información fácilmente interpretable por los humanos. La riqueza de dicha información permite a los expertos (biólogos, arqueólogos y geólogos, entre otros) llevar a cabo investigaciones más detalladas de las áreas de interés después de las misiones y también puede servir como mapa base para la vigilancia ambiental a largo plazo.

Esta tesis presenta una investigación exhaustiva de técnicas de reconstrucción tridimensional basadas en imágenes y su uso para aplicaciones prácticas en el entorno submarino. El estudio se centra en la investigación y el desarrollo de recursos para el escalado preciso y la estimación de la incertidumbre de los modelos tridimensionales con fines científicos basados en datos adquiridos con sistemas de cámaras monoculares o no sincronizadas en entornos (subacuáticos) de difícil acceso.

Se propusieron dos métodos novedosos para el escalado automático de modelos 3D basados en SfM utilizando medidores láser comúnmente disponibles, principalmente, el FUM y el PCM. Estos métodos son capaces de compensar la geometría de la escena junto con la geometría del medidor láser. También eliminan la necesidad del paso, que lleva tiempo y es propenso a errores, de la asociación manual de puntos de modelos de imágenes. Por lo tanto, los enfoques ya no se limitan a determinadas poses de la cámara y/o a la rugosidad de la escena observada al estimar la escala. El nivel de confianza para cada una de las estimaciones también se evalúa de forma independiente mediante la propagación de las incertidumbres asociadas a las características de la imagen y la detección de puntos de láser utilizando una simulación de Monte Carlo. Ambos métodos se consideran universales ya que no están relacionados con datos recogidos por sensores o hardware específicos e incluso se pueden utilizar con datos adquiridos mediante buceadores. Estos métodos requieren tan poco conocimiento sobre la configuración de los láseres que es posible utilizarlos con datos ya existentes. Los métodos han sido plenamente validados por una serie de conjuntos de datos generados sobre la base de un modelo real en 3D de un respiradero hidrotermal en aguas profundas, y su aplicabilidad en escenarios reales ha quedado demostrada por un conjunto de datos del crucero científico *SUBSAINTES*.

Utilizando los métodos de escalado recientemente propuestos, se realizó un análisis de error de escala en modelos a gran escala de los entornos submarinos de las profundidades. Se evaluaron y determinaron los efectos de las estrategias de fusión de navegación

de imágenes más utilizadas. Se consideraron dos escenarios de estudio submarino distintos que encapsulaban condiciones de trabajo realistas en el campo de las profundidades marinas. El estudio demostró que los estudios con múltiples superposiciones de imágenes no secuenciales dan como resultado una solución casi idéntica independientemente de la estrategia, mientras que los estudios con imágenes adquiridas secuencialmente y débilmente conectadas son propensos a producir una deformación a gran escala (efecto domo) cuando la navegación no está incluida en la optimización. Además, evaluamos los efectos de los conjuntos de datos de estudios múltiples. Demostramos que la introducción de parámetros de desplazamiento auxiliares en la optimización tuvo en cuenta con éxito los cambios de desplazamiento presentes en los datos de navegación submarina basados en USBL, y por lo tanto, minimizó el efecto de la contradicción de los datos de navegación.

Además, el análisis también proporcionó una estimación cuantitativa de la escala del modelo, que es un requisito para las aplicaciones que requieren mediciones precisas (como distancias, áreas, volúmenes y otros). Dado que el análisis se realizó utilizando el conjunto de datos de *SUBSAINTES*, permitió a los geólogos de IPGP utilizar los modelos correspondientes en sus respectivos estudios. Los análisis hechos anteriormente con un solo modelo se están realizando ahora con 41 modelos diferentes en 30 sitios distintos.

Para facilitar la detección y caracterización de la incertidumbre de los puntos de láser, también desarrollamos un método que detecta de forma robusta los centros de los rayos láser teniendo en cuenta la textura de la escena. Al mismo tiempo, la incertidumbre de cada una de estas detecciones se determina en base a las formas características recuperadas de los puntos láser. La versatilidad y la robustez del método propuesto han sido demostradas en varios conjuntos de datos.

La tesis también introduce el marco desarrollado de reconstrucción tridimensional basado en imágenes. La solución de procesamiento se construyó integrando varias soluciones de código abierto y desarrollando funcionalidades adicionales (p. ej., fusión de navegación, aplicación permissiva de las limitaciones relativas conocidas, corrección del color submarino). Permite la creación de modelos 3D texturizados foto-realistas basados en datos ópticos y de navegación y es independiente de una plataforma robótica específica, cámara o misión. Realizamos un amplio uso del marco en aplicaciones del mundo real con seis plataformas robóticas submarinas y numerosas tipos de cámaras diferentes. Los modelos creados se utilizaron en colaboraciones con diferentes científicos e investigadores. Permitieron realizar mediciones precisas en los campos científicos, generar vídeo omnidireccional de calidad superior, aplicaciones de realidad aumentada o servir como prueba de que los algoritmos como la planificación de misiones pueden asegurar satisfactoriamente la adquisición de datos. El procesamiento de los datos del crucero científico *SUBSAINTES* llevó a la creación de una de las mayores colecciones de modelos 3D submarinos derivados de imágenes jamás reunidas por vehículos de alta mar con fines geológicos.

Por último, contribuimos demostrando la viabilidad de la reconstrucción tridimensional en tiempo de misión y la estimación de la incertidumbre. Propusimos un sistema basado en el SfM capaz de producir una reconstrucción 3D globalmente consistente junto con una estimación de su incertidumbre mientras el robot está todavía en el agua o poco después. El sistema integra una forma incremental de resolver el problema de la BA, así como una rápida recuperación de la covarianza y un esquema de rechazo de valores atípicos basado en un enfoque de doble mapa. Esto permite adquirir datos de calidad, lo que en consecuencia no sólo aumentará la calidad de los datos finales recogidos y la eficiencia de la medición, sino que también reducirá la posibilidad de realizar mediciones ópticas insatisfactorias.

1

INTRODUCTION

THIS chapter presents the motivation behind the development of this Ph.D. thesis. In Section 1.1 a brief overview of main underwater 3D reconstruction methods is presented together with the principal challenges of optical-based mapping in underwater environment. Next, Section 1.2 states the objectives of the thesis and Section 1.3 describes the context in which this work has been developed. Finally, Section 1.4 concludes with a summary of the organization of this document.

1.1 Motivation

The underwater world is more complex than most of us could ever imagine. As it covers 71% of the Earth's surface [21], the role of the world's largest biosphere in climate regulation, energy resources, and as a habitat for breathtaking biodiversity has made it a vital source of food, transport and trade throughout history. Despite its undeniable importance, the ocean ecosystem and its innumerable processes are still poorly understood, as vast areas of the seabed remain largely unexplored. According to some estimates, less than twenty percent of the world's oceans and seabed have been explored, and less than ten percent have been fully mapped and explored with modern technologies [22].

Despite numerous challenges, seafarers and marine researchers have been committed to mapping the ocean since the early days of seafaring. Be it with simple measurements of the ocean depths with heavy ropes dropped over the side of the ship as early as in the 16$^{\text{th}}$ century, with sound waves at the beginning of the 20$^{\text{th}}$ century and today with state-of-the-art submersibles, which are equipped with arrays of highly accurate optical and acoustic sensors. With the introduction of unmanned underwater vehicles (UUVs) from the military to the industrial and research sector in the late 1970s and 1980s, the availability of various remote operated vehicles (ROVs) and autonomous underwater vehicles (AUVs), such as work-class deep-sea vehicles (e.g. ROV Jason [23], ROV VICTOR 6000 [24], Sentry AUV [25], REMUS 6000 AUV [26]), smaller vehicles (e.g. Girona 500 AUV [27], Seabed AUV [28], Seaeye Cougar XT ROV [29], Sparus II AUV [30]) and their lightweight and low cost alternatives (e.g. Sofar Trident [31], SeaDrone Inspector [32], BlueRov2 [33]) has increased dramatically. Together with towed systems (e.g. TowCam [34], Deep Towed Imaging system - DTIS [35], Wide Angle Seafloor Photography [36], Ocean Floor Observation and Bathymetry System - OFOBS [37]) and various diver-based handheld devices, scientists, industry and even mere enthusiasts now have unprecedented access to large marine areas in both shallow and deep-sea regions [38].

The immense amount of diverse data being collected in previously unreachable areas has already enabled researchers to revolutionise not only our understanding of the fundamental processes responsible for the formation and alteration of the oceanic crust, but has also led to the discovery of new life forms and sources of energy and improved our overall understanding of the oceanic ecosystem. One of the most important discoveries of the 20$^{\text{th}}$ century is considered to be the discovery of hydrothermal vent ecosystems at the Galapagos Rift in 1977 by the world's first deep-sea submersible Alvin [39]. This finding enabled the confirmation of chemosynthesis [40] as the method of survival without sunlight that was previously considered impossible. Since this discovery, modern technologies and maps of the underwater landscape have been widely used in fields such as marine biology, geology and archaeology to survey and detect marine benthic habitats, study hydrothermal vents, as well as observe, record and categorise ancient shipwrecks and settlements. In addition, several commercial companies now offer routine inspections of underwater structures, including pipelines, oil platforms, bridges, dams and other waterfront structures with automatic detection of problematic areas based on the recorded maps.

1.1.1 Underwater environment

Regardless of the enormous improvements in recent years, access to and operation in an underwater environment remains a major challenge and often an extremely costly undertaking. Apart from the countless technical difficulties involved in the design and manufac-

ture of subsea vehicles/devices, the acquisition of potentially useful mapping data remains highly unpredictable due to the particular electromagnetic and acoustic propagation characteristics of the water medium.

While acoustic waves achieve good propagation by causing a longitudinal movement of particles in the direction of wave transmission [41], the same numerous particles of different sizes (from molecules to zooplankton [42]) cause a strong wavelength-dependent attenuation of light. This effectively leads to an exponential loss of energy on the propagation path of light (absorption) and/or to a change of direction of individual photons (scattering). These together with other degradation processes that affect image formation are shown in Figure 1.1.

The effects of attenuation can be described by the inherent optical properties of the water medium (as absorption and volume scattering coefficients) [43] with their values depending on several complex factors, including water temperature and salinity, and the type and amount of particles in the water [44]. Famously, Jerlov [45, 46] classified natural waters into a set of typical oceanic and coastal water types based on their spectral optical attenuation. Examples of attenuation of a full sunlight spectrum in a clear oceanic and coastal water are shown in Figure 1.1(right). The penetration of different wavelengths is shown together with the human perception of water color. As expected, in a clean ocean the light reaches much deeper and causes the water to appear blue, while in coastal waters laden with chlorophyll and dissolved organic matter, such as the English Channel, the stronger attenuation of blue compared to green makes the water appear green [42].
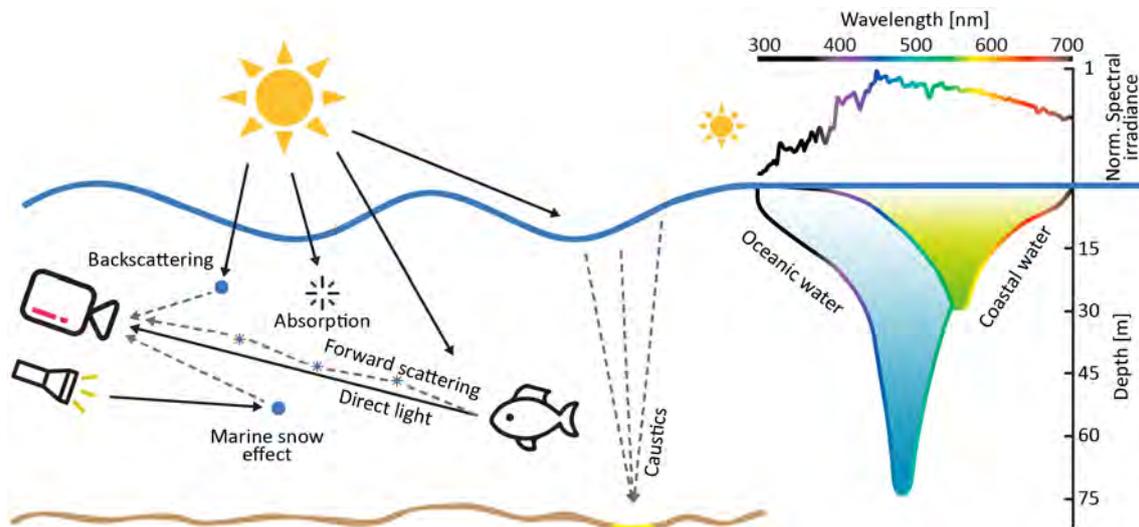
Figure 1.1: Underwater image formation model with different degradation processes (attenuation, scattering, marine snow and caustics). An example of the attenuation of the full sunlight spectrum (top right) occurring in different water types (oceanic and coastal) is shown along with the human perception of water color (blue in clear ocean water and green in coastal waters) on the right. Ultraviolet wavelengths, which are not perceptible to human vision, are colored black (adapted from [42]).

The selective attenuation of light, which is normally greater at red wavelengths (i.e. 600 to 700 nm) than that of green and blue [47], not only causes the color distortion of the light that illuminates the scene itself, but also disproportionately reduces the red intensity of the light reflected (emitted) by the scene. Forward scattering of photons

further reduces the amount of light from the incident beam, while backward scattering introduces additional residual light that is scattered from various other beams. These phenomena can be observed in an image as image blur and light pollution, i.e. a veil covering the object. Figures 1.2 and 1.3 show the effects of light attenuation. The color charts in Figure 1.2 were photographed at various depths z with a constant distance between the chart and the camera. Since the emitted light is equally attenuated in all cases, the change in perceived color is caused by the increasing wavelength dependent attenuation of the sunlight illuminating the chart. On the other hand, the images in Figure 1.3 are deteriorated due to the attenuation caused by the greater distance between the object and the camera (at constant depth). The greater volume of water between the camera and the object results in exponentially greater absorption of the light reflected by the object, as well as increased scattering, making the images appear increasingly bluish and blurred.

$z = 29.1$m $z = 22.6$m $z = 16.2$m $z = 11.9$m $z = 6.0$m $z = 3.8$ m
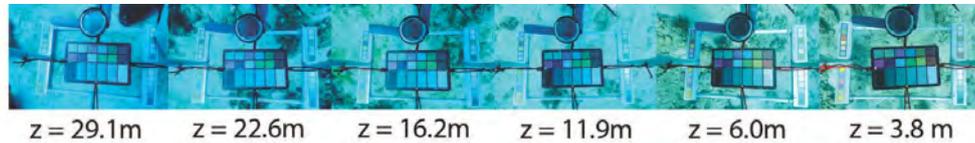
Figure 1.2: The effect of wavelength-dependent attenuated sunlight illuminating a color chart at different depths z (adapted from [48]).
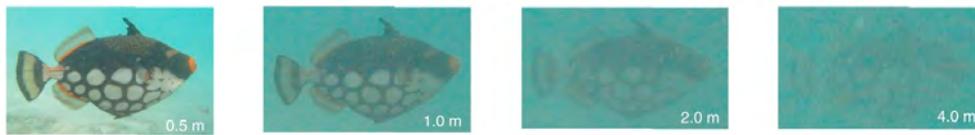
Figure 1.3: The effects of wavelength-dependent attenuation of light emitted by an object at different observation distances (adapted from [42]).

The rapid decay of sunlight intensity with depth therefore requires the use of artificial light sources if optical based acquisitions are to be performed in the deep-sea or correct color information is to be restored (Figure 1.4a and Figure 1.4b). While additional light sources are of great benefit, they can also cause a number of problems. Because such setups usually illuminate the scene unevenly, collected images contain a characteristic bright spot in the center with diminishingly illuminated areas around it. The new light source can also create shadows in the scene, which inevitably change with movement and can therefore cause visual inconsistencies in the collected data. The emitted light can also be excessively reflected by the small observable particles floating in the water, causing disproportionately bright spots known as marine snow (Figure 1.4c). Since vehicles often need to be close to the seafloor to obtain high-quality images, the propulsion of their thrusters can quickly stir up suspended matter, which greatly degrades the data and may even prevent subsequent processing [49].

Conversely, images taken in shallow waters (<10 m) can be significantly affected by the refraction of sunlight at waves (i.e. at the air-water interface) [50]. Strong light fluctuations, also known as caustics/flickering, can cause constant random changes in the perceived scene and therefore, similar to moving shadows, can lead to inconsistencies in the captured data (Figure 1.4d) [51].

(a)

(b)

(c)

(d)
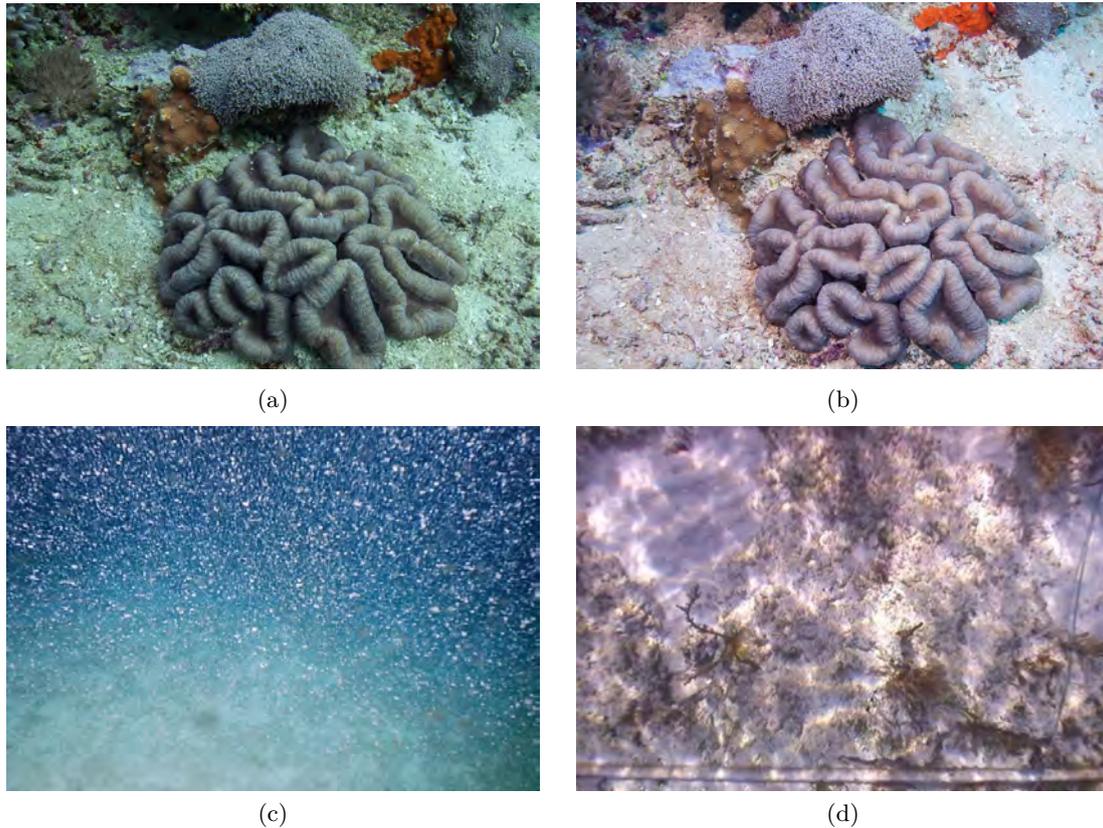
Figure 1.4: Examples of images taken under various conditions under water: (**a**) without and (**b**) with artificial light [52]. (**c**) Marine snow effect due to reflection of artificial light by particles in the water. (**d**) Caustics / flickering caused by refracted sunlight [50].

1.1.2   Underwater 3D reconstruction methods

The most common underwater 3D mapping techniques, which are extensively described by Massot *et al.* [53], can be classified into either acoustic or optical, based on the type of sensor used.

Given the ability of sound to penetrate water much further than light (kilometers compared to tens of meters), it is not surprising that the most commonly used underwater mapping devices are based on acoustics. Modern sonar systems work on the principle that the distance between the sensor and the scene (seabed) can be determined by considering the elapsed time between the pulse and its echo, taking into account the speed of sound in the water. While multibeam profiling and side-scan sonars are usually used to obtain 2.5D elevation maps of the sea floor (i.e. bathymetric maps), their use for performing fine-scale 3D reconstructions has so far been limited (e.g. [54, 55, 56, 57, 58]). Due to their small spatial extent and strict path restrictions, these approaches require very dense sampling, expensive navigation sensors and complex processing, which significantly limits their practical applicability [59]. On the other hand, (imaging) sonars with a wider angle of aperture increase coverage and can thus encapsulate information from larger volumes of water at once. While they are often used for mapping the seafloor (e.g. [60, 61, 62, 63]), the problems of recorded sonar images in the form of blurring along the aperture and

the significant dependence of the perceived appearance of objects on the viewing direction has for the time being, limited their widespread use for 3D reconstruction to a few works reported in the literature (e.g. [59, 64, 65, 66]).

Optical-based 3D reconstruction methods

On the other hand, optical scanning still offers the best option for restoring an accurate 3D representation of smaller areas of interest in higher resolution [67], despite various problems with light penetration as described above. Optical-based 3D reconstruction methods can be further divided into *active* and *passive*, depending on the needs of a technique for additional projection of a signal (e.g. laser line) [53, 68].

Several *active* underwater optical systems [68, 69, 70, 71, 72, 73, 74] have been presented. In most cases a structured light, based on laser (e.g. a laser line), is projected over the field of view of the camera. By exploiting a known relative pose between camera and laser emitter, highly precise individual 3D points can be obtained through a triangulation process. Long acquisition times, caused by the need to move the line projection over the scene, limit its usability under real underwater conditions, where a constant relative movement between the sensor and the observed scene often produces additional reconstruction errors [53]. In addition, the strong absorption and scattering of the emitted laser, even in clear water [75], limits its use to small working distances and prevents the simultaneous use of other light sources. Therefore, intricate strategies such as those presented in [72, 76] have to be employed to acquire both 3D and color information.

In contrast, *passive* methods fully rely on the information acquired by the optical images to reconstruct the 3D model of the observed environment. Additional light sources may be used to illuminate the scene, and the methods are still considered passive as long as they do not aim to cast features used for triangulation of the 3D points [77].

Nowadays, images are normally acquired using RGB cameras, which have undergone a tremendous development in recent years. The evolution of modern and increasingly sensitive image sensors allows their versatile use in various underwater conditions. In addition, their increasingly compact design allows them to be placed on virtually any submersible or to be used hand-held. While in the past the complex handling of the cameras and the high price limited their use underwater to specialized research institutes and companies with specially trained personnel, the above mentioned improvements and the recent emergence of low-cost cameras (e.g. GoPro) have spearheaded their use in a variety of underwater applications even by non-vision experts.

Depending on the number of cameras and their configuration, the camera systems are usually classified as monocular, stereo or multi-camera systems. Stereo systems typically consist of two synchronized cameras offset by a fixed distance (baseline), resulting in a large viewing overlap, while multi-camera systems can consist of any number of synchronized cameras in any pose. Typically, multi-camera systems are designed with the desire to maximize the total field of view of the system [12], and hence the cameras tend to have small overlap areas. Therefore, the main differences between the systems related to 3D mapping are in the number of occasions that any point in the observed scene is viewed in the images recorded at each instance of image acquisition and in the number of known geometric constraints between the cameras in the system. In most underwater scenarios, the limited payload capabilities often mean that the systems have only a single high-resolution camera or stereo camera system at best. While multi-camera underwater metrology systems are gradually coming into the market as commercially available prod-

ucts and services (Rovco; Comex. SA), such systems are still too large and expensive for most scientific applications.

In the context of image-based 3D mapping, monocular cameras have been extensively studied in the computer vision community. Using the concept known as structure from motion (SfM) [78, 79], the geometric information about the structure can be derived from the information encoded in the change of location of projected scene points onto the images taken from different viewpoints. Recent improvements in both speed and robustness [80, 81, 82, 83] have led to a significant increase in the use of SfM in underwater reconstruction scenarios, as the data can be easily captured with monocular cameras mounted on UUVs [84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94], or else is acquired by divers using low-cost, commercially available cameras [95, 96, 97, 98, 99, 100, 101]. To jointly estimate the structure (3D coordinates of scene points) and the movement (relative camera trajectory), modern methods solve the problem by nonlinear optimization of the re-projection error as a cost function, i.e. bundle adjustment (BA) [83]. In order to describe the geometry of the scene completely in high detail, dense and surface reconstructions are built on top of the obtained sparse point clouds by means of triangulation [102].

An example of a result obtained with a SfM-based 3D reconstruction is shown in Figure 1.5. The 3D model of the *Cap del Vol* shipwreck in Port de la Selva (Costa Brava, Spain) was created using our 3D reconstruction framework with images taken by a diver with a standard off-the-shelf camera. The sparse reconstruction of the scene together with the location from which the images were captured is shown in Figure 1.5a. Enlarged views in Figures 1.5c-d show the fine details of the reconstruction. For comparison, Figure 1.6 shows a 2D forward-looking sonar-based mosaic of the same area. This result is taken from the work of Hurtós *et al.* [103] where an ARIS Explorer 3000 sonar was mounted on a Girona 500 AUV tele-operated at a constant altitude of 3m. Since the data were not collected with the same surveying strategy and the methods pursued different objectives (3D vs. 2D), the comparison should only be interpreted as a comparison of the typical representations of the results obtained nowadays with one of the two types of sensor.

Due to the nature of projective geometry, performing a SfM-based reconstruction using only data from a monocular camera precludes obtaining a metric scale of the resulting model, i.e. the model is defined only up to an unknown scale [79, 104]. Since the electromagnetic signals from satellites are heavily damped in water [105] and therefore global positioning system (GPS) is not usable, the scale of the model can be determined by a combination of acoustic positioning (e.g. ultra-short baseline (USBL) or long baseline (LBL) [106] and inertial navigation system [90, 91, 107]). Although acoustic-based navigation provides position estimates that can be used for scale estimation, the associated uncertainties that increase with increasing range (i.e. depth) in addition to the possible loss of communication (navigation gaps) mean that the scale is estimated from data that is often noisy and/or poorly resolved. In addition, many smaller submersible vehicles do not have sufficiently accurate navigation capabilities due to the high cost and large payload requirements, resulting in pure dead reckoning navigation solutions which often result in excessive drift in the estimated trajectory.

The scale can also be estimated through the introduction of known distances or object dimensions [92]. Since in real underwater scenarios only rarely are known measurements readily available, the latter requires the placing of objects on site (e.g. [95, 96, 97, 98, 108]) and is therefore practically limited to small-scale reconstructions and shallow water environments due to the time and logistical effort involved. Alternatively, knowledge of
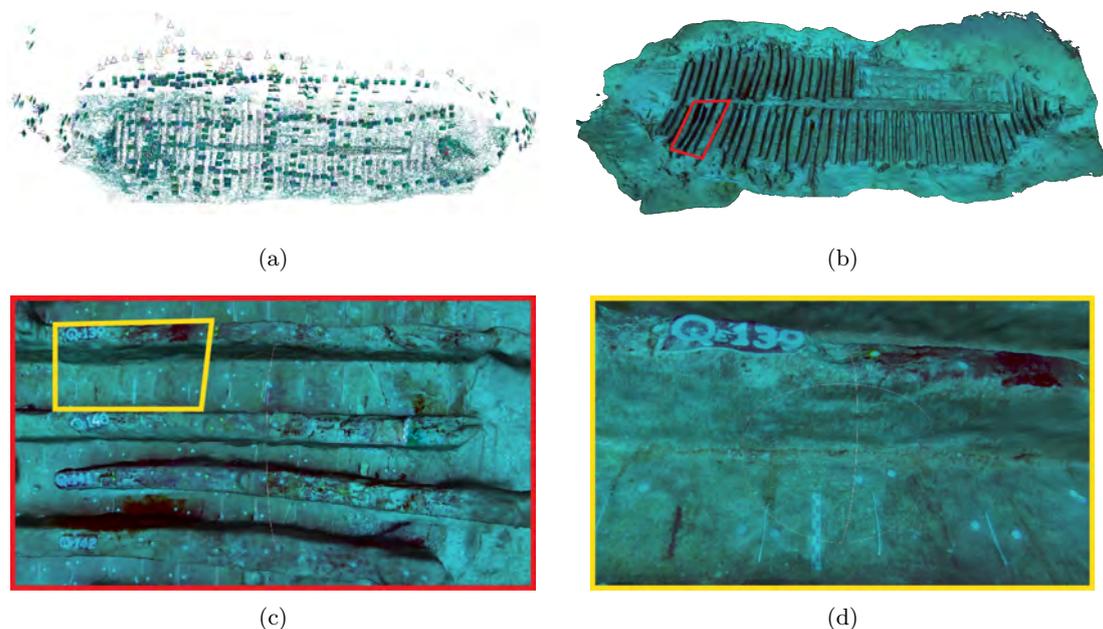
Figure 1.5: SfM-based 3D reconstruction of the *Cap del Vol* shipwreck at Port de la Selva (Costa Brava, Spain). (**a**) Sparse point cloud , and (**b**) textured 3D model of the reconstructed scene with magnified details in (**c,d**).
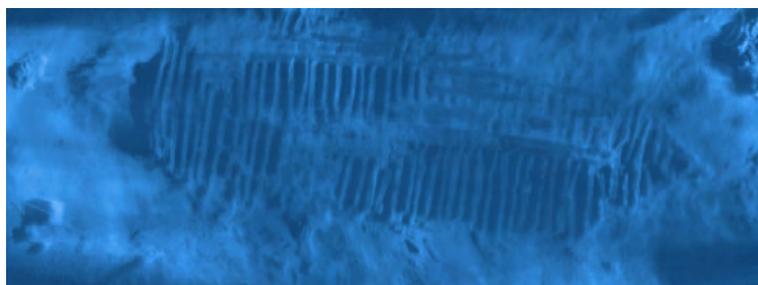
Figure 1.6: Acoustic-based 2D mosaic of the *Cap del Vol* shipwreck, created using forward-looking sonar data. .The results are taken from Hurtós *et al.* [103].

the distance between different points can be obtained by projecting laser beams in a known configuration onto the scene (e.g. [94, 109, 110, 111, 112]). These image scaling methods require precise alignment of the laser scaler to the camera and projection of the lasers onto flat areas of the scene, ideally perpendicular to the camera, to overcome the lack of knowledge of the scene geometry and camera-scene distance. Both conditions prove to be extremely limiting in real scenarios with rough surfaces. In order to transfer the knowledge of scale from the image to the 3D model, each laser point must additionally be identified on the model itself, which leads to an extremely error-prone and time-consuming manual process.

As a special case of SfM with a sequential nature of observation in a robotic setup, visual SLAM (V-SLAM) enables precise localization of an underwater robot in real time while it is moving through the environment and simultaneously creates a map of the explored areas [113]. Since pre-calibrated and synchronized cameras allow for any scene

point viewed by multiple cameras to be efficiently identified, converted to a 3D representation and subsequently merged into a single 3D reconstruction of a large area, numerous commercial (e.g. [113, 114, 115]) and custom-build stereo camera systems (e.g. [102, 116, 117, 118, 119, 120]) have also been employed on underwater robots or used by scuba divers [121]. Besides the fact that visual sensors are a cost-effective alternative to more specialized sensors (e.g. Doppler velocity log, USBL, inertial measurement unit and sonar), V-SLAM methods can provide information about the location and the environment even in cluttered areas where solutions such as dead-reckoning may not be sufficiently accurate [122]. A variety of methods have been developed that can be roughly divided into feature-based, direct methods and a few recent attempts using deep learning [123, 124]. When used underwater, dramatic lighting variations and the common use of radial corrections to compensate for refraction effects at air-glass-water interfaces [125] lead to inconsistencies in the results of direct methods [114, 126], while purely feature-based V-SLAM approaches often suffer from poor imaging conditions, fast movements and dynamic scenes [127]. This problem is often solved by merging visual information with inertial measurements, Doppler velocities, an/or depth pressure measurements [128] (e.g. [118, 129, 130, 131, 132, 133]), which in the case of monocular-based systems (e.g. [129, 131, 133, 134]), also helps with initialization, scale ambiguity and scale drift problems [135]. To suppress accumulated estimation errors within the global map or robot trajectory, the global solution should constantly optimized. While in the past filter-based approaches (Extended Kalman Filter [136, 137] and its variants [130]) were used, most modern simultaneous localization and mapping (SLAM) approaches perform a graph-based optimization [113, 138, 139] through BA. To limit the growing computational complexity due to the increasing number of points and cameras, common solutions decouple the problem into a local BA step that minimizes the re-projection error over a few recently added cameras and 3D points, and a global optimization of the pose-graph optimization [135, 140, 141]. While this allows for real-time results, such solutions are only locally consistent when executing large loops and cannot provide uncertainty estimates as indicators of the quality of the acquired optical data.

Given the number and unpredictability of degradation effects that severely affect the quality of optical images taken underwater, the lack of real-time feedback on data quality requires strong human intervention and an overly conservative approach during surveys to try to ensure the collection of adequate data. Not only does this significantly increase the time required for a mapping mission, but it can also lead to the need for additional missions in the same area. It is therefore, together with scale estimation, one of the most limiting problems of optical-based 3D reconstruction with monocular cameras in underwater scenarios.

1.2 Objectives

With the motivations of this thesis described, we can now state that the main goal is:

> *To research and develop resources for accurate scaling and uncertainty estimation of 3D models for scientific purposes, built based on data acquired with monocular or un-synchronized camera systems in difficult-to-access GPS-denied (underwater) environments.*

This general aim can be divided into the following objectives:

- Survey the state of the art in SfM-based 3D reconstruction and navigation fusion to understand the work and research contributions of other authors in the field.

- Develop a versatile 3D reconstruction framework for the generation of photo-realistic textured 3D models based on optical and navigation data, independent of a specific robot platform, camera or mission.

- Systematically analyze and determine the most appropriate 3D reconstruction and navigation fusion strategy for large-scale reconstructions in GPS-denied environments using a monocular camera.

- Design and develop an automatic scaling method for SfM-based 3D models with uncertainty estimation that can be used in routine underwater investigations using commonly available underwater laser scalers, with emphasis on reducing the need for manual labor in the process.

- Extend existing laser spot detection algorithms to allow automatic estimation of the uncertainty in the position of detection.

- Extensively validate the framework, both in simulation and in several field tests performed in representative environments with real robot platforms.

Surveying the state of the art and the validation of the reconstruction framework have been a constant effort throughout the development of the thesis and as such reflected in all publications of this compendium as well as in numerous other manuscripts published in collaboration with fellow researchers. The two proposed scale estimation methods have been covered in separate publications with thorough evaluations in simulation, with a comprehensive application to real data was carried out in conjunction with the systematic analysis of fusion strategies, which has been covered in a third publication of the compendium.

1.3   Context

The work presented in this thesis was supported by the Marie Skłodowska-Curie Early Stage Researcher Fellowship of an European Academy for Marine and Underwater Robotics - Robocademy (FP7- PEOPLE -2013-ITN-608096) and was developed at the Underwater Vision Lab (UVL) of the Underwater Robotics Research Center (CIRS) within the Computer Vision and Robotics Research Group (ViCOROB) of the University of Girona (UdG). The group started its activities in the field of underwater vision and robotics in 1992 and is now considered one of the leading groups in the research and development of AUVs for accurate seafloor mapping and light intervention. The team, consisting of pre-doctoral researchers, engineers, technicians, post-doctoral fellows and permanent staff, has participated in many national and European projects and is actively involved in various technology transfer projects. Over the years, the group has developed several AUVs prototypes and currently has two fully-functional robots: the Sparus II AUV [30] and the Girona 500 AUV [27], both of which have been used for data collection in the context of this dissertation. Other robotic platforms used in this dissertation are VICTOR 6000 [24], Seaeye Cougar XT ROV [29] and PlaDyPos ASV [142].

Throughout this work, various collaborations have taken place with other researchers from Computer Vision and Robotics Research Group (ViCOROB) and from abroad. The

(a) Sparus II AUV [30]

(b) Girona 500 AUV [27]
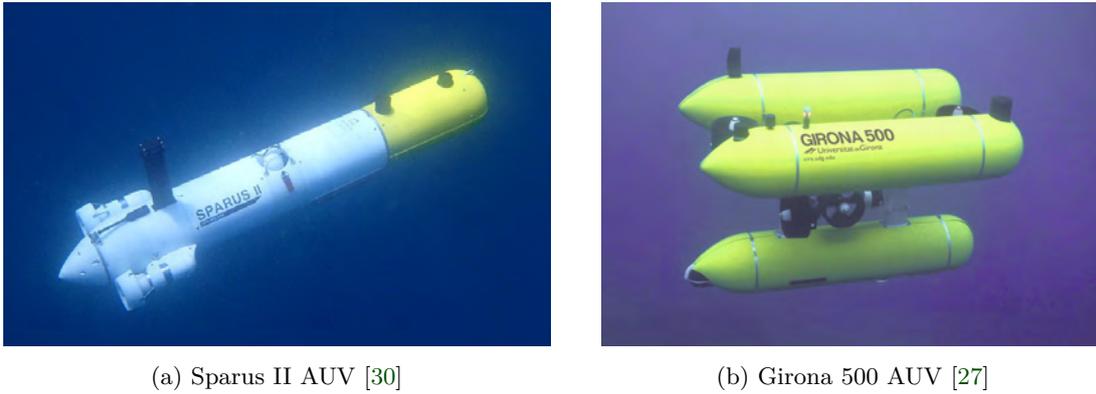
Figure 1.7: Robots currently in use at the CIRS research center.

use of 3D models in omnidirectional video stitching enabled the collaboration in [12]. The developed reconstruction framework was extended to use the rigidity information of the multi-camera system and to provide spatial information about the scene to enable seamless panoramic stitching. The same system has been used in collaboration with the University of Zadar to demonstrate the methodology for the use of omnidirectional cameras on underwater robots for rapid high-resolution mapping of shipwrecks in marine archeology [7]. As part of a successful collaboration with fellow researchers of ViCOROB focusing on mission planning, the 3D reconstruction framework has also been significantly improved and extended to work with different camera configurations and robot platforms [1, 8, 11, 13, 18]. The 3D reconstruction framework has also received a number of improvements and extensions to work with different camera configurations and robot platforms. Other collaborations with researchers developing GIS-like mapping in a virtual reality (VR) environment [16, 17] and studying gorgonian populations in the Mediterranean [14] have also been possible through our 3D modeling capabilities.

Due to the close cooperation with marine scientists, the author also had the opportunity to participate in three international research missions. In the framework of the EU project MIDAS (Grant No. 603418, FP7) we took part in an 8-day demonstration cruise at the Mid-Atlantic Ridge (Figure 1.8a), during which the 3D reconstruction framework was tested and evaluated for its applicability to industry-standard environmental baseline surveying and impact assessment. The 3D reconstruction abilities with imagery acquired using omnidirectional and stereo camera systems were also tested during the 7-day archaeological campaign in Caesarea, Israel, in cooperation with the Israel Antiquity Authorities, the University of Zagreb and the University of Rhode Island. Most recently, the author participated in the second stage (10 days) of the 21-day mapping cruise SUBSAINTES (Figure 1.8b) near Les Saintes (Guadeloupe, France Antilles). In collaboration with Paris Institute of Earth Physics (IPGP) and the French Research Institute for Exploitation of the Sea (IFREMER), the 3D reconstruction framework was extensively used to create 3D models of the Roseau fault rupture [9, 10, 15, 16].

Finally, this thesis also benefited from a 5-month research stay at the ARC Centre of Excellence for Robotic Vision within the Australian National University (Canberra, Australia). During this time an optical-based system for mission-time 3D reconstruction with quality estimation for difficult environments was developed. The published papers [4, 6] received the Best Paper Honorable Mention Award at the International Conference on

(a) *MIDAS* cruise at Mid-Atlantic Ridge   (b) *SUBSAINTES* cruise at French Antilles

Figure 1.8: The scientific cruises attended to validate the developed algorithms.

3D Vision - 3DV 2017 (Qingdao, China) and the Norman Miller Best Student Poster Award at IEEE OCEANS'17 (Aberdeen, UK).

The experiments, equipment and infrastructure resources used in this thesis have been partially funded by the following projects:

- EU H2020 Project ROBUST (ref. 690416-H2020-CS5-2015-onestage) funded by the European Commission.

- EU H2020 Project Eurofleets Plus (ref. 824077) funded by the European Commission.

- MINECO Project UDRONE (ref. CTM2017-83075-R) funded by the Spanish Ministry of Education, Culture and Sport.

- ANR Project SERSURF (ref. ANR-17-CE31-0020, France) funded by the French National Research Agency.

- MINECO Project OMNIUS (ref. CTM2013-46718-R), funded by the Spanish Ministry of Economy, Industry and Competitiveness.

- MINECO Project ARCHROV (part of MERBOTS) (ref. DPI2014-57746-C3-3-R) funded by the Spanish Ministry of Science and Innovation.

- EU H2020 Project EXCELLABUST (ref. H2020-TWINN-2015(CSA)-691980) funded by the European Commission.

- EU H2020 Project STRONGMAR (ref. H2020-TWINN-2015(CSA)-692427) funded by the European Commission.

1.4 Document structure

This document is structured into the following chapters:

- **Chapter 2:** In this chapter, the publication *Autonomous Underwater Navigation and Optical Mapping in Unknown Natural Environments* introduces our first image-based 3D reconstruction pipeline developed for the purpose of automatically generating photo-realistic textured 3D models of previously unexplored scenes. It addresses underwater specific problems that normally occur under changing conditions and in

dynamically planned missions, such as color correction and inconsistent speeds of image content changes. The work was carried out in collaboration with research colleagues at the Underwater Robotics Research Center (CIRS) research center, focusing on mission planning algorithms.

- **Chapter 3:** In this chapter we present our work on the problem of the exact scaling of 3D models created from monocular camera images in GPS-denied environments. In the publication *Automatic Scale Estimation of Structure from Motion based 3D Models using Laser Scalers*, we present two novel methods of scaling based on laser scalers. Together they cover the majority of different laser scaler configurations, one completely unrestricted method which allows the use of any laser setup, while the partially constrained method reduces the required knowledge of the geometry of the setup by assuming the laser parallelism and equidistance of the laser origin to the camera. The extended validation of the methods on a simulated real world example is reported, as well as their applicability to the real scenario.

- **Chapter 4:** In the publication of this chapter, entitled *Scale Accuracy Evaluation of Image-Based 3D Reconstruction Strategies Using Laser Photogrammetry*, a complete scale accuracy framework is presented first. As an extension of our earlier work, the scaling methods are incorporated into a Monte Carlo (MC) simulation capable of estimating the confidence levels for each of the scale estimates. The detection and uncertainty estimation of laser spots is addressed through a new automatic laser spot detection method, which aims to facilitate the detection of even strongly attenuated laser beams. The framework is then used to perform a systematic analysis of scale errors in 3D reconstruction and navigation fusion strategies applied to the reconstruction of 3D models on a larger scale in GPS-denied environments using monocular camera images. Two distinct surveys from the *SUBSAINTES* scientific cruise are used to assess the effects of different real deep-sea field working conditions.

- **Chapter 5:** The work done during my research stay at the Australian National University (ANU) is presented in this chapter by a publication entitled *Mission-time 3D Reconstruction with Quality Estimation*. It presents a novel SfM-based system capable of producing globally consistent 3D reconstruction together with an estimate of its uncertainty in mission-time or shortly thereafter. This publication describes the outlier rejection scheme, which is based on a dual-map approach, as well as the integration of the incremental fashion of solving the BA problem using an incremental non-linear least squares (NLS) solver and a method for fast covariance recovery, both implemented in the SLAM++ library by the collaborating authors.

- **Chapter 6:** This chapter contains a summary of the results obtained in the context of this thesis, as well as some additional new and previously unpublished results. The last section of this chapter also gives an overview of the most important models reconstructed due to various collaborations with other researchers and scientists during the development of the thesis.

- **Chapter 7:** Finally, the last chapter presents the conclusions and some guidelines for future work.

2

Autonomous Underwater Navigation and Optical Mapping in Unknown Natural Environments

In this chapter we describe an approach for autonomous exploration and subsequent image-based 3D reconstruction of underwater structures. First, the procedure for online planning of a collision-free path is presented. This endows an AUV with the capability to autonomously acquire optical data in close proximity to the environment without the requirement for any previous knowledge of the observed area. Subsequently, we focus on an image-based 3D reconstruction pipeline. It addresses underwater specific problems that normally occur under changing conditions and during dynamically planned missions, such as color correction and inconsistent speeds of image content changes. Through several steps the pipeline is able to create photo-realistic textured 3D models of the inspected area. The work, carried out in collaboration with a research colleague at CIRS was evaluated using the Sparus II AUV in a challenging real-world natural environment. For the experiments, a custom-made multi-camera setup was designed and built, which allowed for the reconstruction of a complex underwater canyon between two rocky formations with a single pass of the robot.

sensors   MDPI

Article

Autonomous Underwater Navigation and Optical Mapping in Unknown Natural Environments

Juan David Hernández *, Klemen Istenič *, Nuno Gracias, Narcís Palomeras, Ricard Campos, Eduard Vidal, Rafael García and Marc Carreras

Underwater Vision and Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, C\Pic de Peguera, 13 (La Creueta), 17003 Girona, Spain; ngracias@eia.udg.edu (N.G.); npalomer@eia.udg.edu (N.P.); rcampos@eia.udg.edu (R.C.); eduard.vidalgarcia@udg.edu (E.V.); rafael.garcia@udg.edu (R.G.); marc.carreras@udg.edu (M.C.)

* Correspondence: juandhv@eia.udg.edu (J.D.H.); klemen.istenic@udg.edu (K.I.);
   Tel.: +34-972-419-651 (J.D.H. & K.I.)

Abstract: We present an approach for navigating in unknown environments while, simultaneously, gathering information for inspecting underwater structures using an autonomous underwater vehicle (AUV). To accomplish this, we first use our pipeline for mapping and planning collision-free paths online, which endows an AUV with the capability to autonomously acquire optical data in close proximity. With that information, we then propose a reconstruction pipeline to create a photo-realistic textured 3D model of the inspected area. These 3D models are also of particular interest to other fields of study in marine sciences, since they can serve as base maps for environmental monitoring, thus allowing change detection of biological communities and their environment over time. Finally, we evaluate our approach using the Sparus II, a torpedo-shaped AUV, conducting inspection missions in a challenging, real-world and natural scenario.

1. Introduction

Environmental science is an interdisciplinary field that gathers together different natural sciences to study and determine the interactions of physical, chemical and biological components of the environment, as well as their effects on the organisms which inhabit it. An important objective in such studies is to establish a baseline that permits detecting changes and correlating them with possible underlying factors. In order to correctly identify such changes, it is necessary to conduct long-term and high-frequency observations of the studied ecosystem. To this end, and especially during the last decades, robotic systems have started being used to systematically collect such environmental data [1].

Marine scientists were among the first to capitalize on the use of robotic vehicles for environmental monitoring. Oceanographers, for instance, started using unmanned underwater vehicles (UUVs) to study deep marine environments and the seafloor [2]. However, although the majority of such early applications were devoted to monitoring marine habitats, nowadays there are a significant and increasing number of robots that contribute to other environmental science domains. The interested reader is encouraged to look into [1] for an extensive review of such contributions.

UUVs, are divided into two categories: remotely operated vehicles (ROVs), which need to be controlled by a human operator, and autonomous underwater vehicles (AUVs) that conduct (autonomously) a pre-established mission. While the former group has as a main drawback its dependence on a surface vessel to operate the vehicle, the second, on the other hand, involves important research challenges around localization, perception, mapping, path planning, and safety—just to

mention a few. Therefore, an important part of the work in underwater environmental robotics research, especially that involving AUV, has concentrated on developing the basic functional modules that allow autonomous operation.

In their simplest form, robotic monitoring applications in underwater environments involve an AUV that follows a sequence of pre-calculated waypoints in order to collect data, which are retrieved after concluding the mission. In other words, the AUV behaves as a mobile sensor that explores and measures aspects of interest in an underwater environment. For this reason, another important body of research has been dedicated to developing pipelines that automatically and systematically process large amounts of information.

With the increased selection of sensors now included in the AUV allowing performance of underwater explorations, underwater 3D mapping now relies on acoustic multibeam [3,4] or sidescan sonars [5] to produce elevations maps. This ability to accurately map underwater environments yields high added value to any survey, as such results convey immense information easily interpretable by humans [6].

However, while these maps are indispensable for providing a rough approximation of the terrain, they are not able to sense more complex structures (e.g., they cannot represent concavities). For this reason, optical imaging is used to recover high quality 3D representation of small areas of interest in high resolution [7]. For an extensive review of various methods for underwater 3D reconstruction, the interested reader is referred to [8].

Furthermore, as a result of numerous readily-available off-the-shelf underwater camera systems, as well as custom-made systems for deep-sea explorations, an increasing number of biologists, geologists and archaeologists rely on optical imagery to survey marine benthic habitats [9–12], study hydrothermal vents and spreading ridges [13,14] as well as ancient shipwrecks and settlements [6,15,16]. Underwater imagery has also been used to identify and classify different benthic elements in the surveyed area [17–19] as well as to detect changes in the environment [20–22].

Whatever the type of information required to study an underwater environment (e.g., thermal, chemical, acoustic, optic, etc.), most of the surveys done with AUVs are conducted in a previously explored area so that the vehicle can navigate at a constant and safe altitude from the seafloor. In a typical application, the vehicle uses its on-board sensors to gather (environmental) data that is used to build thematic maps. However, recent and potential new applications require the AUV to navigate in close proximity to underwater structures and the seafloor. An example is the imaging and inspection of different structures such as underwater boulders [23] or confined natural spaces (e.g., underwater caves) [24]. In some of these cases, preliminary information about the structure to be inspected, such as its location and shape, permits determining the region of interest in advance, so that a coverage path is pre-calculated. The information obtained during the mission is used to correct or adjust the path to the goal online in order to adapt to the real shape of underwater structures [23]. Nonetheless, there are applications in which no previous information is available on the environment, or cannot be obtained autonomously. In such cases, preliminary work has focused on gathering data to characterize such environments [24], while relying on human supervision to ensure vehicle safety.

On that basis, the purpose of this paper is to propose a framework that endows an AUV with the capability to autonomously inspect environments for which no previous information is available. The framework consists of two main functional pipelines: (1) one that computes collision-free paths while simultaneously mapping the surroundings incrementally; (2) another that allows the reconstruction of various 3D representations (i.e., sparse, dense, meshed, textured) of the surveyed area using images gathered by an arbitrary camera setup during the mission. Due to the aforementioned constraints of underwater optical mapping, the latter pipeline establishes a range of distances at which the AUV must navigate, which represents a path constraint to be considered by the former functional pipeline. The resulting 3D reconstructions will serve as base maps for environmental monitoring of interest areas, allowing the detection of any change in biological communities and their

environment on a temporal scale, and enabling a new way to visualize the evolution of wide areas in that temporal scale.

The remainder of this paper is organized as follows. Section 2 presents our proposed path planning pipeline that permits an AUV to autonomously navigate in unknown environments, and also discusses the extensions necessary in order to calculate paths that attempt to maintain a desired range of visibility, i.e., distance to inspected structure. Section 3 reviews the reconstruction pipeline that builds a 3D textured model of the inspected area using optical imagery. In Section 4, we present a real-world mission that validates our approach. Results include both an autonomous inspection conducted by the Sparus II (University of Girona, Girona, Spain) AUV (see Figure 1) in a challenging and natural environment, and its corresponding 3D reconstruction done with optical data gathered during the inspection mission. This clearly extends our preliminary work, where only simulated or real-world but structured (non-natural) environments have been used [25,26]. Finally, concluding remarks and directions for further research are given in Section 5.

Figure 1. Sparus II, a torpedo-shaped AUV.

2. Path Planning Pipeline

This section reviews our path-planning pipeline that solves start-to-goal queries online for an AUV that operates in unknown environments [25]. In order to accomplish this, the pipeline is composed of three functional modules. The first of them incrementally builds an occupancy *map* of the environment using on-board perception sensors. The second one *plans* safe (collision-free) paths online. The third and last functional module works as a high-level coordinator that *handles the mission* execution by exchanging information with the other two modules and the AUV's controllers. Figure 2 depicts how these functional modules are connected to one another. Additionally, we explain how to extend this pipeline by incorporating a criterion to maintain a desired distance to guarantee visibility constraints while conducting a mission in close proximity.
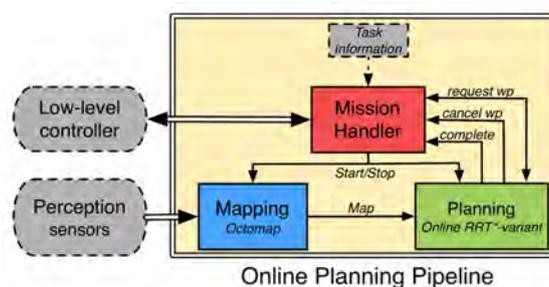
Figure 2. Pipeline for online path planning for AUV.

2.1. Module for Incremental and Online Mapping

The *mapping* module incrementally builds a representation of the environment by using data received from different kinds of perception sensors, such as multibeam or mechanically scanned profiling sonars, echosounders, etc. Such sensors provide a range of information about nearby obstacles that, combined with the vehicle's navigation (position and orientation), permits establishing the free and occupied space with respect to an inertial coordinate frame. In order to represent this data, we use an octree-based framework called Octomap [27], which has three main characteristics that permit efficient modelling such volumetric information. The first characteristic is the probabilistic state representation that not only allows us to modify the map when updated environment information is available, but also protects it from noisy measurements, i.e., a position state considers previous information and calculates its new value according to probabilistic functions. The second characteristic is the capacity of representing unexplored areas, which can be relevant for guiding exploration in unknown environments. Finally, Octomap offers a computationally efficient way to enlarge or extend the map as demanded. Figure 3 shows a breakwater structure and its representation with an Octomap, which has been built using multibeam sonar data obtained by a surface vessel.

| (a) | (b) |

Figure 3. (**a**) breakwater structure in the harbor of Sant Feliu de Guíxols in Catalonia, Spain. The structure is composed of a series of blocks, each of which is 14.5 m long and 12 m wide; (**b**) Octomap created from real-world data obtained with a multibeam sonar.

2.2. Module for (Re)Planning Paths Online

The *planning* module is in charge of calculating a collision-free path for the AUV. For doing so, this module receives a query to be solved that is specified with a start configuration (q_{start}) and a goal configuration (q_{goal}), and other parameters, such as the available computing time and minimum distance to the goal. Furthermore, given that the vehicle navigates in an unknown environment, this module is required to continuously verify and repair (if necessary) the path from the current vehicle's position to q_{goal}. In order to calculate a collision-free path under such constraints, i.e., incrementally and online, we have modified the asymptotic optimal RRT (RRT*) [28], which is one of the most relevant sampling-based path planning algorithms.

The rapidly-exploring random tree (RRT) and its variants are algorithms that incrementally build a tree of collision-free configurations. Their main characteristic is the rapid and efficient exploration of the C-Space [29]. The state-of-the-art method for calculating optimal paths is the RRT* with its concept of asymptotic optimality (an algorithm is said to be probabilistically complete when the probability that the planner finds a path, if one exists, asymptotically approaches one as the number of samples increases), which was firstly introduced in 2010 by Karaman and Frazzoli [28,30]. This property states that the total cost of the solution, measured by a user-defined function, decreases as the number of samples increases. In this approach, new configurations are connected to the closest and best

configuration, i.e., the one that guarantees a minimum cost. Furthermore, an additional step of sample reconnection allows improving the associated cost of the surrounding configurations.

However, the RRT* used in this work not only permits us to progressively improve the path (through its property of asymptotic optimality), but has also been modified to incorporate concepts of *anytime* computation and *delayed collision checking*. These extensions enable the enhancement of the performance for online (re)planning applications, such as the one proposed for this work.

2.2.1. Anytime Approach for (Re)Planning Online

Even though RRT* has been previously extended to behave as an *anytime* algorithm [31], our alternative approach grows a single tree and prunes it to discard those branches that result under collision after updating the map, similarly to what Bekris and Kavraki proposed for a standard RRT [32]. Like other RRT-based algorithms, our variant consists of two procedures, `build` and `extend`. The former procedure, which is presented in Algorithm 1, works similarly to other RRTs that sample uniformly distributed configurations (line 4) and attempt to expand the tree towards them (line 5) (see also Algorithm 2)). However, our variant has two main modifications.

The first modification arises from the necessity of correcting or adjusting the path according to the new elements discovered in the environment. In order to deal with this, `updateTree` procedure is called before sampling new configurations (line 2). With this procedure, the modified RRT* traverses the tree using a depth-first search (DFS) algorithm to check if any node or edge is under collision. If a new collision is detected, the corresponding subtree (i.e., the tree that includes the nodes or edges under collision) will be discarded. Nonetheless, if the tree root is one of the nodes under collision or if the path from the current vehicle's configuration to the root is not feasible, our modified RRT* (i.e., the *planning module*) informs the *mission handler* to cancel the current waypoint and starts again planning a new path from the current vehicle's position. This latter situation occurs because the tree root always corresponds to the configuration (or position) that the vehicle is moving towards, as explained below.

The purpose of the second modification in Algorithm 1 is to make the RRT* behave in an *anytime* fashion. To do this, if the new configuration resulted from the tree expansion meets the specified minimum distance to the goal (line 7), it is added to a list of possible solutions (line 8). After concluding the tree expansion, if the *mission handler* has requested a new waypoint and there is at least one available solution stored in the list (line 10), the planner selects the solution with the minimum associated cost, sends the *mission handler* the configuration connected to the root of that solution (line 13), and prunes the tree in such a way that the configuration sent becomes the new tree root (line 14). During this *pruning* process, subtrees connected to the initial root (excepting the corresponding to the new root) are discarded.

In our modified RRT*, `extend` procedure (Algorithm 2) remains as originally proposed in [28]. This means that it receives a random configuration (q_{rand}) towards which the tree will be expanded. To do so, it first finds the node (configuration) q_{near} that is the nearest to q_{rand} (line 2). It then calculates a path of length δ from q_{near} towards q_{rand} which, in turn, generates a new configuration q_{new} (line 3) (in a geometrical case, i.e., when no motion constraints are considered, the connection between two configurations results in a straight line segment). If both q_{new} and the path that connects q_{near} and q_{new} are proved to be safe (collision-free), this procedure will not only incorporate q_{new} into the tree, but will also check its surrounding nodes to reconnect them in case better (less expensive) connections are possible (lines 6–9). Finally, this procedure returns a value that confirms whether the expansion was successful (line 10) or not (line 12).

Algorithm 1: buildRRT

 Input:

 T: tree of collision-free configurations.

1 begin

2 updateTree()

3 **while not** *stop_condition* **do**

4 $q_{rand} \leftarrow$ sampleConf()

5   *result*, $q_{new} \leftarrow$ extendRRT(T, q_{rand})

6   **if** *result* $\neq$ *TRAPPED* **then**

7   **if** dist(q_{new}, q_{goal}) $< \epsilon_{goal}$ **then**

8   addSolution(q_{new})

9   *solution_found* $\leftarrow$ *true*

10 **if** *solution_found* **and** *wp_req* **then**

11   *result_path* $\leftarrow$ getBestSolution()

12   *new_root* $\leftarrow$ *result_path*[1]

13   sendWaypoint(*new_root*)

14   pruneTree(*new_root*)

Algorithm 2: extendRRT*

 Input:

 T: tree of collision-free configurations.

 q_{rand}: state towards which the tree will be extended.

 $\mathcal{C}$: C-Space.

 Output:

 Result after attempting to extend.

1 begin

2 $q_{near} \leftarrow T$.findNearestNeighbor(q_{rand})

3 q_{new}, *collision* $\leftarrow$ calculatePath($q_{near}, q_{rand}, \delta$)

4 **if not** *collision* **then**

5   addNewNode(T, q_{new})

6 $Q_{near} \leftarrow$ findNearestNeighbors(T, q_{new})

7 $q_{min\_cost} \leftarrow$ findMinCost(T, Q_{near}, q_{new})

8   addNewEdge(T, q_{min_cost}, q_{new})

9   reconnectNearNeighbors(T, Q_{near}, q_{new})

10   **return** *ADVANCED*

11 **else**

12   **return** *TRAPPED*

Figure 4 shows a simulation of the Sparus II AUV conducting a mission in an environment that resembles the breakwater structure presented in Figure 3. In this case, the vehicle is assumed to have a mechanically scanning profiler with a perception distance of 20 m, the tree generated by our modified RRT* is presented in dark blue, the path to the goal is drawn in red, and the path to the current waypoint appears in yellow. The mission, specified as a start-to-goal query, consists in navigating from one side of a series of blocks (obstacles) to the other (see Figure 4d). The environment is initially undiscovered and is incrementally mapped as the AUV navigates towards the goal. When the vehicle starts the mission, and no obstacle has been detected, the waypoint sent to the AUV's controllers (the tree root) coincides with the goal, since a straight path to it is feasible (see Figure 4a). This situation

can persist even when obstacles have been detected, as long as the path from the vehicle's position to the goal is collision-free (see Figure 4b). However, when such a straight path is not possible, the planner starts again calculating a new path from the vehicle's current position, as previously explained (see Figure 4c). More details about the simulation environment and test scenarios will be provided in Section 4.
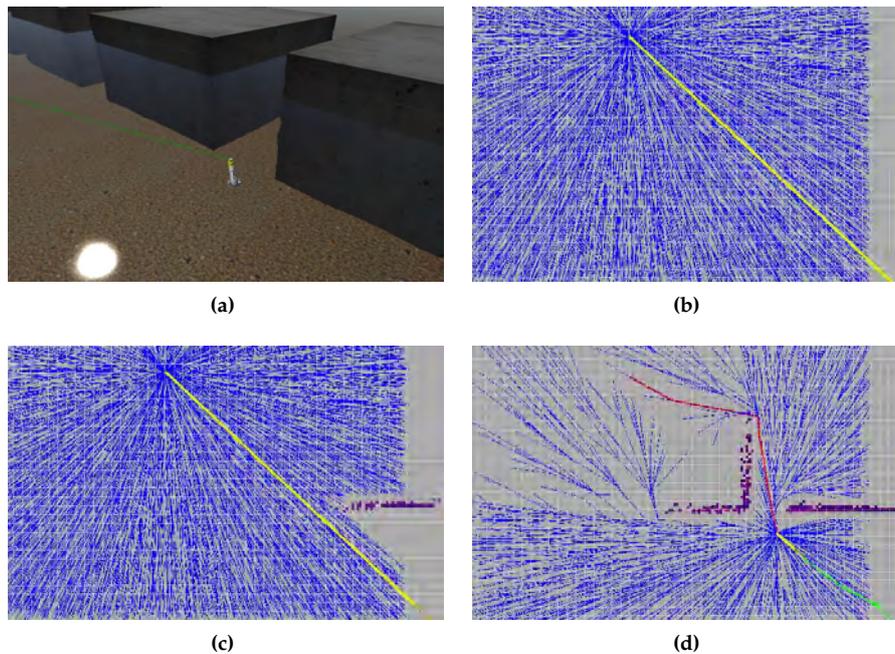
Figure 4. Sparus II AUV conducting an autonomous mission in a simulated scenario (**a**), where it incrementally maps the environment (**b**), (**c**) and (re)plans a collision-free path to the goal (**d**). The tree of configurations is presented in dark blue, the path to the goal in red, the path to the current waypoint in yellow, and the vehicle's trajectory in green.

2.2.2. Delayed Collision Checking for (Re)Planning Incrementally and Online

When navigating in unknown or partially known environments, the information about the static and dynamic surrounding elements (obstacles) is progressively acquired as the vehicle moves. Because of this, an important number of configurations (sampled or obtained after expanding the tree) are located in unexplored regions of the environment. In these situations, it is not only impossible but also unnecessary to attempt to determine if a configuration is at risk of collision. To compensate for this, our strategy is to assume as safe (collision-free) any configuration that is out of the explored area, which can be efficiently determined when using Octomaps (as described in Section 2.1). Furthermore, given that the tree expansion is periodically interleaved with updating the map, such parts initially assumed as safe will be verified and discarded if found under collision as the vehicle explores the environment. This approach is inspired by the *lazy collision checking* strategy introduced by Bohlin and Kavraki [33].

Figure 5 depicts another simulation in the same scenario used in Figure 4, but, in this case, the explored regions of the maps are presented in light blue and the occupied ones are presented in green. With this visualization, the importance of delaying the collision checking for those configurations located in undiscovered regions can be appreciated. As mentioned before, the tree is checked and reshaped as the environment is being explored (see changes from Figure 5a–b).

(a)   **(b)**
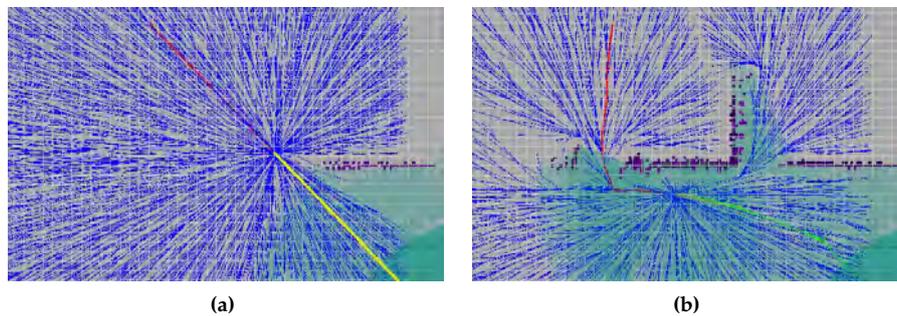
Figure 5. Sparus II AUV conducting an autonomous mission in the same simulated scenario (breakwater-structure). (**a**) The explored region, presented in light blue, expands as the vehicle moves towards the goal. It is important to notice that a significant part of the tree (dark blue) is located in undiscovered areas of the workspace; (**b**) Those branches are initially assumed as safe (collision-free) until the corresponding region has been explored, thus avoiding unnecessary collision-checking routines computation.

2.3. Mission Handler

The third functional module that constitutes the path-planning pipeline is the *mission handler*. This module is in charge of controlling and coordinating the previously explained modules (*mapping* and *planning*). It also verifies whether the AUV is prepared to start solving and conducting a task; to do so, this module communicates with other functional modules on the vehicle to verify both that navigation data is correctly being generated and that the vehicle's low-level controllers are not conducting any safety manoeuver. After completing the checking stage, the *mission handler* starts requesting waypoints from the *planning* module, which, after being received, are adapted and sent to the vehicle's low-level controllers. Finally, this module is also responsible for cancelling any ongoing waypoint if it is notified by the *planning* module, as explained in Section 2.2.

2.4. Conducting Surveys at a Desired Distance Using a C-Space Costmap

As mentioned in the introduction to this paper, environmental science requires conducting long-term and high-frequency observations in order to determine changes over the studied ecosystem. One of the alternative ways of conducting such a study—and the one presented in this work—is to gather optical data that can be used to build 3D representations of the environment. In this way, changes in the environment can be detected by inspecting the different reconstructions built over time. However, visibility is highly variable in underwater environments and can play a critical role when collecting optical information. For this reason, the planning pipeline has to be adjusted according to both the visibility conditions given when conducting the data collection and the camera parameters.

In order to tackle this issue, we propose establishing a costmap over the C-Space that specifies a distance constraint as an attempt to guarantee the visibility with respect to the surface of interest. To do so, the cost that is associated with each configuration q must be calculated according to Equation (1), where $0 \leq Cost \leq 100$, with 0 and 100 as the minimum and maximum costs, respectively. This cost function is used as the optimization objective for our planning pipeline and its RRT*. In this formulation, the *Cost* is dependent on the distance (d) and has additional parameters that permit adjusting the costmap. These parameters include the expected distance to the inspected structure (d_e) and the range of distance (Δd_a) that defines an admissible interval in which the cost is minimal (clearly observed in Figure 6):

$$Cost(d) = \begin{cases} \left(1 - \frac{d}{d_e}\right) 100, & d < d_e - \frac{\Delta d_a}{2}, \\ 0, & d_e - \frac{\Delta d_a}{2} \leq d \leq d_e + \frac{\Delta d_a}{2}, \\ \left(\frac{d}{d_e} - 1\right) 100, & d > d_e + \frac{\Delta d_a}{2}. \end{cases} \quad (1)$$

Lastly, there is an additional important aspect to note when using Equation (1) within the planning pipeline. When high-cost values are defined for certain zones, it does not imply that a restriction is applied to planning paths over those zones (thus discarding possible paths), but it does means that those paths should be avoided as far as possible. To put it in another way, the proposed approach does not define restricted areas in which the vehicle would not be allowed to move through. This situation may occur when the only feasible path coincides with the highest cost one (e.g., narrow passages), in which case the planner will also admit the path as a valid solution.
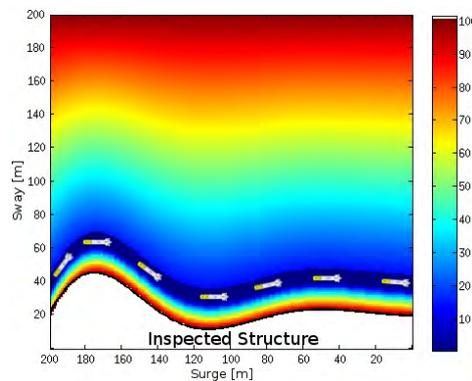
Figure 6. Costmap projected in vehicle's X(surge)-Y(sway) plane. Dark blue indicates the zone that meets visibility constraints.

3. 3D Reconstruction Pipeline

Underwater environments with their particular phenomena (i.e., light attenuation, blurring and low contrast) [7], can be regarded as hostile places for optical surveying. Acquisitions have to be performed at close range, significantly limiting the area viewed in a single image, thus enormous amounts of data have to be collected and processed to generate a wide area view enabling the extraction of valuable information on a broader space scale [34]. While 2D photomosaics have been successfully used in describing large areas [13,35,36], they can only be accurately generated if the scene is relatively planar [37] and images taken in a nearly orthographic setup (e.g., [38]). In scenarios with significant 3D structures, the aforementioned prerequisites cannot be met, resulting in obvious distortions (see [39]). However, redundant information from multiple images can be used to provide visually rich 3D reconstructions of the underwater terrain. Furthermore, as the camera poses do not have to be orthogonal to the seafloor, images can convey more meaningful information regarding the global shape of the object, especially in the case of intricate structures (e.g., underwater hydrothermal vents), which are impossible to capture accurately using downward looking cameras [7].

In this section, an optical-based 3D reconstruction pipeline is presented. Through a series of sequential modules (see Figure 7), the pipeline is able to reconstruct a 3D model based on the optical imagery acquired from an arbitrary number of cameras in various poses. Cameras do not have to be synchronized and can be freely re-positioned between missions, based on the mission's goals and the expectations of the terrain in the surveyed area. However, in order to obtain a metric-scaled reconstruction, the cameras have to be connected with the AUV sensors [40] or known measurements of the environment have to be introduced [41].
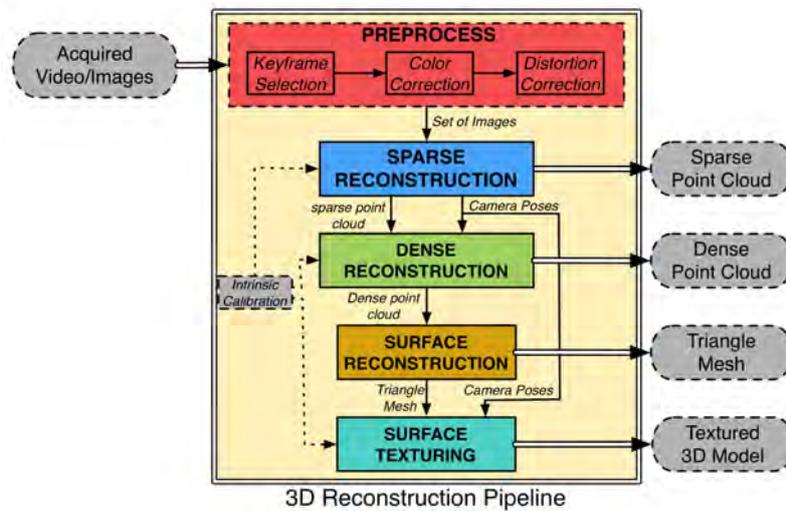
Figure 7. 3D reconstruction pipeline.

Each of the reconstruction steps results in a different representation of the reconstruction. While the preferable representation for visualization and dissemination is a textured triangle mesh, some applications may require solely dense or even a sparse point cloud of points describing the observed area. In such cases, substantial reduction in the computational cost can be achieved, as none of the subsequent steps have to be performed. It is also worth noting that an important pre-processing step of color correction is required to reduce the image degradation effects of the water medium.

3.1. Keyframe Selection

Once the AUV has been recovered, the data acquired during the mission has to be collected and preprocessed in order to obtain a consistent set of quality images used in the subsequent reconstruction process. The optical data can be acquired either by still imagery or through recording of high resolution videos. Challenging underwater conditions in which the acquisition is performed often lead to blurry and low-contrast images, which have to be detected and removed before the start of the reconstruction process. Videos provide a large number of images (i.e., frames) with large overlap and enable a greater flexibility in image selection. However, as the movement of the AUV is typically slow, the length of the videos and consequently the number of frames may become overwhelming and inefficient to process. Furthermore, as the displacement of the camera (i.e., baseline) between consecutive frames is extremely small, the reconstruction process cannot reliably infer the depth information [42]. Thus, a smaller subset of quality images with sufficient movement is extracted and used in the following steps to build an accurate 3D model.

The initial keyframe selection step employed in our reconstruction pipeline is performed as a twofold process on data acquired by each of the cameras used in our setup. To identify the frames in which the vantage point has sufficiently changed, we use a similar approach to Cavan in [43]. The Lucas–Kanade tracking algorithm [44] has been used to track features detected as points with strong gradients in both image directions (minimum eigenvalue corners) [45]. The camera motion is then estimated based on the movement of the tracked features. Despite the fact that this does not allow us to directly estimate real camera movement, it does, however, enable us to detect frames where the content has sufficiently changed. Similarly, if the number of tracked features between the frames is insufficient, we predict a significant change and extract a new keyframe. While Cavan selected each candidate that met this criteria, we perform an additional filtering step to avoid extracting blurry frames. A new frame is selected among a few closely consecutive frames based on the blurriness

score, which is estimated as the variance of its Laplacian [35]. As the Laplacian highlights the rapid change of intensities, higher variance represents a more in-focus image with sharper edges.

This approach enables us to automatically adjust the extraction rate of the frames depending on the movement of the AUV, as opposed to time-dependent frame extraction (e.g., selecting a frame every second). However, the approach requires an empirical selection of per-video threshold value for sufficient feature movement on the image plane. On the contrary, the least blurry image is detected based on the relative highest variance and is not conditioned by any thresholds.

3.2. Color Correction

The inherent properties of the water medium induce several effects causing the color and brightness of the images to vary significantly depending on the distance to the scene [46]. The red components of the visible spectrum are strongly absorbed, resulting in typical greenish-blue images [47] (see Figure 8a). The image quality can be further degraded by the scattering and absorption effects caused by the presence of suspended particles in the medium (e.g., plankton) [48,49]. For heavily degraded images, applying color correction not only benefits human perception and interpretation of its contents, but also improves the quality and quantity of successful scale-invariant feature transform (SIFT) matches between image pairs [50] used in the subsequent reconstruction process.

The color correction techniques can be based either on a physical model of the image formulation process (image restoration) or on subjective qualitative criteria (image enhancement). While the former requires knowledge of the physical parameters of the medium, the latter can be performed based on various suggested criteria (e.g., histogram stretching). As our goal is to perform the reconstruction without the requirement of prior knowledge about physical parameters or medium properties, we use an image enhancement technique proposed by Bianco et al. [47]. The method is based on the "white-world" assumption in Ruderman opponent color space $l\alpha\beta$ [51] and uniform illumination of the scene. The $l\alpha\beta$ space is used to separate the luminance (l) and two chromatic components (α, β) and subsequently shift their distributions around the white point $(0, 0)$. This can be seen as the correction of image-cast and the adjustment is analogous to the "grey-world" assumption in the RGB space. Additionally, histogram cutoff and stretching are performed on the luminance component to improve the image contrast. An example image before and after color correction and contrast enhancement is depicted in Figure 8.

(a) **(b)**

Figure 8. (**a**) image acquired with a downward oriented camera; (**b**) the same image after color correction.

3.3. Distortion Correction

Aside from the effects of the light passing through water, the incident light beam's path is additionally altered due to the difference in the density in the water–glass–air interface (between the camera sensor and the scene). The change in path destroys the collinearity between the point in water,

the camera's center of projection, and the image point [52], resulting in a distorted image, causing the scene to appear wider on the image than it actually is [53].

For planar interfaces, such as we use (see Section 4.1), the deformation increases with respect to the distance from the camera's center of projection. It can be regarded as a pin-cushion distortion, and its effects can be reduced by considering it as radial distortion [54]. Unknown radial distortion parameters introduce additional ambiguity in the subsequent structure from motion (SfM) process and can lead to ambiguous reconstructions (e.g., bending of the model) [55]. This can be avoided by either using pre-calibrated radial distortion parameters or avoiding critical acquisition configurations [55].

As we cannot ensure the absence of critical configurations, due to the unknown structure of the environment, we pre-calibrate the intrinsic parameters of the cameras in an underwater environment prior to the mission using a calibration pattern and a standard calibration procedure [56].

3.4. Sparse Reconstruction

Using a set of previously pre-processed images, the 3D geometry of the scene, simplified to sparse 3D points, is simultaneously estimated with the motion of the cameras through a process known as structure from motion (SfM). The problem solved by SfM can be seen as an inverse process of image formulation. Instead of finding the points of intersection between the image plane and the rays connecting the camera's centre of projection and the points in space, the goal is to recover the position of the points in space together with the pose of the camera. The estimation is done entirely from the texture features extracted and matched across the 2D image set. By using the equations of projective geometry, connecting the image projections and the position of real world points, the solution is estimated through a non-linear minimization of the reprojection errors—also known as bundle adjustment [57]. As the process is based on projective geometry, the obtained reconstruction can be defined only up to scale [42].

3.4.1. Feature Detection and Matching

Given that the structure and motion parameters are inferred entirely from the projections of the points on the images, these interest points should be salient features robustly detected and associated across multiple views. In our approach, we detect such features using Wu's [58] graphics processing unit (GPU) implementation of SIFT [59]. SIFT is widely accepted as one of the highest quality feature descriptors [60] as it has a high degree of invariance to scale and rotation, as well as being partially invariant to changes in illumination, noise, occlusions and small changes in the viewpoint. The points are detected as extremes of the difference of gaussians (DOG) at multiple scales, and described based on local gradients using a 128-element normalized unit vector.

The association of features across the image set is done image pairwise based on the Euclidean distance using Lowe's ratio test [59] and subsequently filtered through a geometric filtering procedure to eliminate possible outliers. The behaviour of individual matches is evaluated with respect to the global estimate of the transformation between the two images, i.e., epipolar constraints [42]. Without prior knowledge of the cameras positions, the fundamental/essential matrices are computed using a robust statistical method to prevent the influence of possible outliers on the estimation of the model. In order to avoid the empirical selection of the inlier/outlier threshold value in the widely used method of random sample consensus (RANSAC) [61], we use the parameter-free evolution called a contrario-RANSAC (AC-RANSAC) [62] implemented in the open-source library OpenMVG [63]. The method uses the *acontrario* methodology, which relies on the Helmholtz principle of meaningful deviations and regards any model that is unlikely to be explained by chance as conspicuous. As the meaningfulness of the model is determined by data-specific statistical criteria, explicit threshold values are not required.

3.4.2. Structure from Motion

Using the established feature correspondences, the 3D scene and extrinsic camera parameters are gradually estimated through a sequential SfM [64] implemented in the open-source library OpenMVG [63]. The process starts with the initialization step, where a seed pair of images is selected. Without prior knowledge of the scene, initial poses are decomposed directly from the estimated fundamental/essential matrix relating the pair. As the erroneous initial estimation can cause the algorithm to converge to wrong local minima, from which it is unlikely to recover, the selected pair has to be well conditioned, i.e., have a wide baseline and not many coplanar common points, to ensure robust estimation of the fundamental/essential matrix [42]. The reconstruction is then incrementally expanded with newly observed points and cameras one view at a time. Extrinsic parameters of each new camera are initialized through a direct linear transform technique [42] using already estimated 3D points. After each step of the expansion, bundle adjustment is performed to minimize and evenly propagate the re-projection error.

As the reconstruction is performed sequentially from a series of relative motion estimations between the images, the accumulation of small errors can lead to a drift in the final reconstruction. This can be significantly reduced with the introduction of non-sequential constraints in the optimization process. An effective strategy is to perform a loop closure by re-visiting the same area. Using the image matching between the non-sequential images the relative motion is restricted and drift minimized.

While the algorithm enables the recovery of both extrinsic (i.e., the position and orientation of the camera at the moment of the acquisition) and intrinsic cameras parameters (i.e., focal length, principle points and radial distortions of the lens), this can lead to ambiguous reconstructions (excluding the inherited scale ambiguity) [55] (see Section 3.3). We avoid the radial ambiguity by pre-calibrating the cameras in an underwater environment and consider the intrinsic parameters constant during the reconstruction process. This additionally reduces the complexity of the problem and subsequently reduces the possibility of convergence to a wrong solution.

As a final result, an estimate of the external camera parameters together with the sparse 3D structure is obtained as shown in Figure 9.

Figure 9. Example of sparse 3D scene reconstruction together with with camera positions (red).

3.5. Dense Reconstruction

As sparse 3D points describing the observed scene recovered in the previous step are usually not sufficient to describe the underlining structure in detail, we perform a densification step in which we aim to obtain a detailed globally consistent dense representation of the scene. This is achieved by trying to estimate the 3D coordinates of each pixel in the acquired images. Using the mathematical models estimated in the SfM process (e.g., camera projection matrices) [65], the correspondence

information required for such estimations can be determined through a dense image matching procedure, either by using stereo pairs (stereo matching) or by identifying correspondences in multiple images (multi-view stereo).

Modern dense matching algorithms can be categorized either as local or global [66]. While local methods are efficient, due to pixel-wise correspondence evaluation and local "winner-take-all" optimization [65], they tend to produce incorrect results at sudden depth variations and detailed areas, as well as lacking the ability to reconstruct surfaces in locally ambiguous areas (e.g., occlusions, repeated patterns and uniform texture regions) [67]. Relative robustness in such areas can be achieved using global methods, which determine the solution by minimizing a global cost function extended to all image pixels [68] but require significant computational effort. An efficient solution can be found using a Semi-Global Matching [69] method, which is a local pixel-wise method that approximates the minimization of a global 2D smoothness constraint by combining several independent 1D constraints, thus allowing the recovery of object boundaries and fine details.

We perform the dense reconstruction using a semi-global matching-like multi-image method [70,71] implemented in an open-source MicMac photogrammetry [72]. The reconstruction is formulated as an energy minimization problem and solved by finding a minimal cut in a graph. For each hypothetical 3D point, a patch in the master image is identified and projected to all the neighbouring images, which are then used in the computation of the global similarity estimate through Normalized Cross Correlation [65]. An energy minimization approach, similar to [69], is then applied to enforce surface regularities and avoid undesirable jumps [71].

To reduce the computational complexity, the method is based on a multi-resolution pyramidal approach using a coarse-to-fine extension of the maximum-flow image matching algorithm presented in [73]. At each pyramid level, matching results for the relevant resolution are computed and used to guide the matching process at a higher level. This produces a multi-stereo correlation result for each master image in the form of a depth map. A 3D point cloud is later obtained by projecting the points to space according to the camera's pose and associated depth value. Additionally, RGB attributes can be assigned to each of the 3D points from the appropriate master image [71]. The result can be seen in Figure 10, depicting the scene from the same viewpoint as in Figure 9.

Figure 10. Example of dense 3D scene reconstruction.

3.6. Surface Reconstruction

In sparse and dense point clouds obtained in the previous steps, the scene has only been described using an unorganized noisy point cloud without any assumption about their connectivity. This representation is frequently not sufficient for further processing of the data, and it also does not enable proper visualization, as visibility information is not established, preventing the user from easily distinguishing points that should be either visible or occluded from a specific viewpoint [7].

In the surface reconstruction step, our aim is to describe the geometry of the scene using a triangle mesh, i.e., finding the most probable surface based on the sampling represented by the point cloud obtained from the noisy dense reconstruction. Optical-based reconstructions created in the underwater environment are normally corrupted by both noise and outliers due to poor imaging conditions [7]. Surface reconstruction methods can be classified as interpolation- or approximation-based depending on their approach [7]. Interpolation methods consider the points in the point set only as possible vertex candidates in the resulting triangle mesh, thus making them directly dependent on the quality of the point cloud. For this reason, such methods should only be used with ideal (or nearly ideal) point clouds, i.e., noise and outlier free [7]. On the contrary, the approximation-based methods are able to mitigate the effects of noise in the data by considering the point set only as information about the surface and not necessarily as final vertices. However, the implicit smoothing hampers its ability to recover sharp features (e.g., edges, corners). Considering that our main focus is on the recovery of the scenes in the underwater scenarios, where sharp edges are rare (with the exception of man-made structures) and the point cloud is noisy, we decided to use the Poisson method [74,75] as one of the most representative approximation-based methods.

The Poisson surface reconstruction method forms a unique implicit surface representation (iso-surface) through the reconstruction of an indicator function, i.e., a function having the value 0 if inside the object, and 1 if outside. As the oriented points in the point cloud can be seen as samples of the indicator function's gradient, the point cloud can be used to represent the function's gradient field. By computing the inverse of the gradient, that is, by finding the scalar function of which gradient best approximates the gradient field defined by the input point set, the problem can be transformed to a Poisson problem and its solution found by determining the scalar function whose Laplacian equals the divergence of the gradient field. To efficiently represent the 3D function, an octree adapted to the distribution of the samples is used.

By using the indicator function, the Poisson method implicitly requires the possibility of determining the inside and outside of the surface, i.e., the object has to be watertight [7]. As the underwater scenarios are predominately focused on the reconstruction of the seafloor with various additional 3D structures, the scene can not be properly viewed from all angles. Such scenarios force the Poisson method to define false surfaces in areas lacking needed information. These areas can be subsequently eliminated by removing triangles with edges longer than a certain threshold (triangles increase in size in non-sampled parts) [76]. The resulting model is presented in Figure 11.
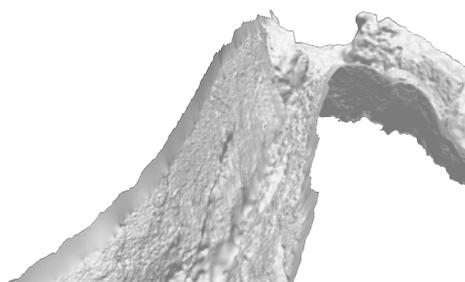
Figure 11. Example of surface reconstruction.

3.7. Surface Texturing

In the final step of our reconstruction pipeline, the goal is to obtain a photo-realistic 3D model of the observed scene. A consistent texture for the reconstructed surface mesh can be retrieved by mapping high-quality textures from the input images.

Generating a globally consistent and seamless texture is challenging due to changes in the acquisition conditions (e.g., changes in illumination, light attenuation, white-balancing, presence

of unreconstructed occluding objects (particles in the water)), varying image scales (e.g., close-ups, distant overview images), as well as unavoidable imperfections in the reconstructed geometry) [77]. As multiple images observe each surface element, their information can either be fused using blending techniques or the information from a selected image can be used. In blending, due to the inaccuracies in the camera poses, slight inaccuracies in the reconstruction and different attenuation of the light depending on the camera poses, can lead to ghosting and blurring of details in the final texture. Alternatively, by selecting the most appropriate image for each texel, seams between regions mapped from different images can be made clearly visible [78].

In our pipeline, we use the work of Waechter et al. [77], which performs the mapping of textures from multiple registered images in two steps. First, for each surface face, a single image is selected through a process of energy minimization, preferring close focused orthogonal views with high resolution and similar adjacent patches. Additional photo-consistency checking is employed to detect and reject any inconsistent views caused by unreconstructed occluding objects. Color discontinuities between the patches are then adjusted in an attempt to minimize the visibility of the seams. Per-vertex-based globally optimal luminance correction terms are computed using a weighted average of the vertex color along all adjacent seam edges, and are used together with local Poisson image editing [79] to generate the final coherent texture of the reconstructed scene.

An arbitrary user-defined view of the textured 3D model is presented in Figure 12.

Figure 12. Example of surface reconstruction.

4. Results

In order to evaluate and validate the presented approach, we selected a natural and unexplored underwater scenario in which the Sparus II AUV had to autonomously navigate. The vehicle not only successfully conducted multiple start-to-goal missions, but also gathered optical data that was used to build a 3D reconstruction of the surroundings. The results demonstrate the capabilities of our approach in natural real-world conditions, and validate our preliminary work conducted in a simulated environment [26]. This section explains the vehicle setup and presents the results obtained in one of the conducted missions.

4.1. Experimental Setup and Simulation Environment

Sparus II is the most recent AUV developed at the CIRS [80]. Rated for depths up to 200 m, the torpedo-shaped robot has three thrusters (two horizontal and one vertical, see Figure 13a) and can be actuated in surge, heave and yaw modes degrees of freedom (DOF), which endows it with hovering capabilities. For estimating its position and orientation, the vehicle is equipped with a navigation sensor suite that includes a pressure sensor, a doppler velocity log (DVL), an inertial measurement unit (IMU) and a GPS to receive fixes while at surface. Furthermore, the vehicle has different perception sensors that are located within the vehicle's payload (front) area, including a mechanically scanning pencil-beam sonar used online to create the surroundings map, and a set of three GoPro Hero 4 Black edition cameras (GoPro, San Mateo, CA, United States) that gather the images required to create the

3D environment's reconstruction. The cameras are positioned systematically (see Figure 13) to ensure the highest possible coverage, while still maintaining the ability to perform feature matching between images taken from different perspectives. Two cameras were placed in a downward configuration at an angle of 20 degrees while the third camera was positioned forward looking at 40 degrees.

(a)   (b)

Figure 13. Sparus II AUV. (**a**) CAD model, where the three thrusters can be observed, as well as the profiling sonar and cameras located in the payload area (yellow); (**b**) real-world vehicle's payload.

Sparus II AUV is controlled through the component oriented layer-based architecture for autonomy (COLA2) [81], which is completely integrated with the robot operating system (ROS). Furthermore, COLA2 not only operates the real vehicle, but can also interact with the underwater simulator (UWSim) [82], thus permitting the use of 3D environment models and simulation of the vehicle's sensors and dynamics. Before conducting real-world trials, we used UWSim with different virtual scenarios in order to extensively simulate and test most of the vehicle's functional modules, including our path-planning pipeline. For this particular work, we designed two virtual scenarios: one that resembles the breakwater structure mentioned in Section 2 (see Figures 3 and 4), and one that includes an underwater canyon located between two rocks (see Figure 14). In both cases, the vehicle succeeded in conducting start-to-goal missions without having a priori information on the environment. Another important aspect to mention is that we make use of the open motion planning library (OMPL), which is a general path-planning library that can be extended and adapted to different contexts and problems [83].

(a)   (b)
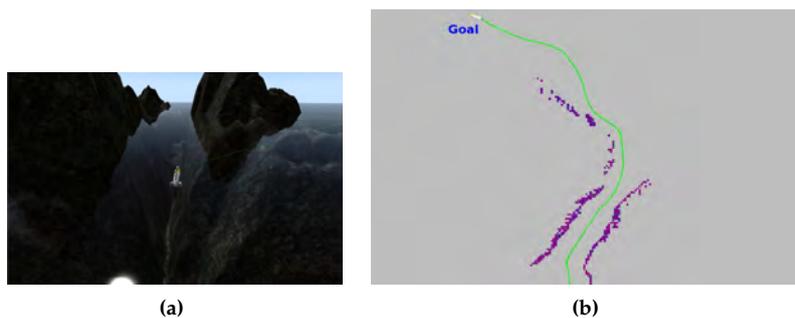
Figure 14. (**a**) Sparus II AUV conducting autonomous missions in a simulated environment (UWSim), which resembles an underwater canyron created by a rocky formation; (**b**) The vehicle after traveled successfully through the canyon. The map, generated online, can be observed in purple, while the vehicle trajectory appears in green.

The following sections present an inspection mission conducted by the Sparus II AUV in a real-world and natural environment. The results include not only the navigation map created online from the profiling sonar data, but also the 3D reconstruction of the underwater surroundings.

4.2. Online Mapping and Path Planning in Unexplored Natural Environments

In order to evaluate our approach, we tested its effectiveness in a challenging real-world natural environment in Sant Feliu de Guíxols, Spain (see Figure 15). The testing area contains rocky formations that create an underwater canyon. In order to inspect this environment, two different start-to-goal queries were established by extracting GPS coordinates from Google Maps [84]. The first query required the Sparus II AUV to traverse the canyon towards the shore. The second query goal was chosen on the outside of the rocky formation in such a way that the vehicle had to circumnavigate the outer rock. Furthermore, after completing the second query, the first query was executed again until the vehicle overlapped its initial trajectory in the canyon, in order to close the imaging acquisition loop and thus improve the reconstruction results. The profiling sonar of the AUV only covers the horizontal plane, which restricts the safe motion of the vehicle to planes of constant depth. For this reason, the navigation was set at a constant depth of 3 m for both queries.

Figure 15. The test scenario that consists of rocky formations that create an underwater canyon.

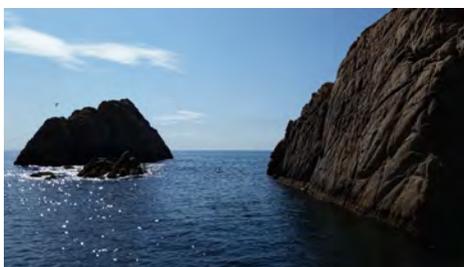Figure 16 depicts the results of the inspection mission, where the AUV not only created a map of a complex and unknown environment, but also planned a collision-free path, simultaneously and incrementally. The map and the vehicle's trajectory are shown overlapping a satellite image. In the initial part of the mission, i.e., when the vehicle traverses the canyon for the first time, the map coincides with the satellite image (see Figure 16b); however, disparities can be clearly observed after some time (see Figure 16c,d). Such differences are due to the accumulation of errors in the dead reckoning system (position and orientation estimation) that depends on the DVL, which may provide incorrect data when navigating over rocks, as occurred in this test scenario. Despite this situation, the vehicle succeeded in conducting the mission because both the map and the path are created online, which permits correcting or adjusting them even when moving in previously visited areas (see Figure 16d when accessing the canyon for a second time).

4.3. 3D Reconstruction

During the autonomous inspection mission, each of the three cameras captured 725 seconds of video, comprising a total of $21,750$ frames. The imagery was acquired in 2.7 K HD 4:3 video mode with a resolution of 2704×2028 pixels. In order to test our pipeline, we used the complete set of frames as input. Using the keyframe selection method described in Section 3.1, we automatically identified 264, 376, and 316 representative frames from the left, right and forward looking camera, respectively.

The visibility varied between different sections of the mission from $\sim$4.5 m inside the canyon to $\sim$3 m on the most exposed outer part. These conditions, combined with the different viewing angles of the cameras, caused the acquired images to have low contrast and different appearances depending on the cameras, with predominant blue and green tones (see Figure 17a–c). This was especially noticeable in the last part of the mission, where the AUV pulled away from the rocky formation. For this reason, the images were color corrected and contrast enhanced, to unify the appearance and improve the feature detection and matching (see Figure 17d–f).
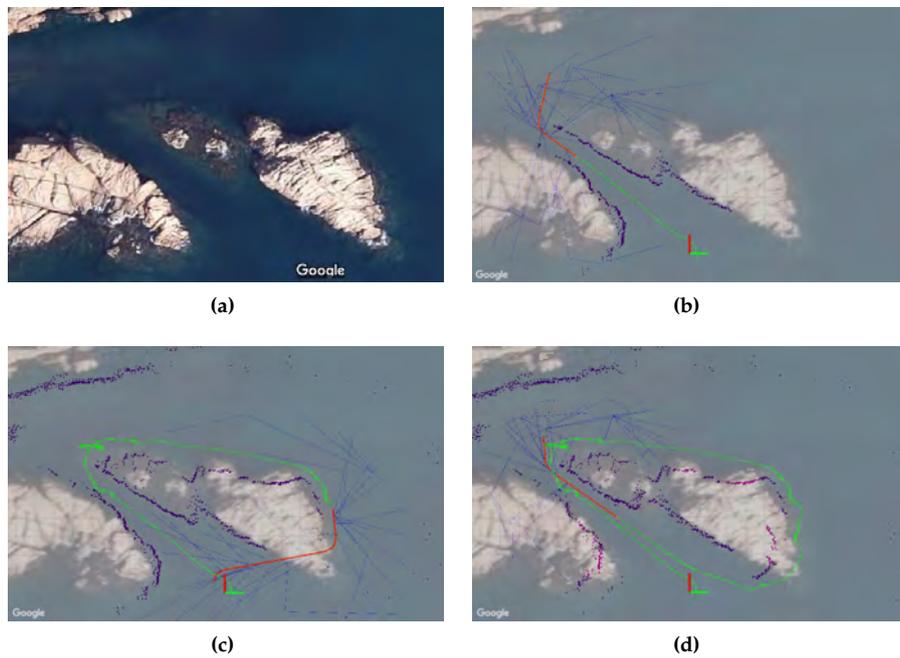
Figure 16. (**a**) the test scenario consists of rocky formations that create an underwater canyon; (**b**) Sparus II AUV conducting the first part of the inspection mission that requires traversing the canyon; (**c**) during the second part of the mission, the vehicle circumnavigates one of the rocks on its way back to the initial position; (**d**) the AUV partially repeats the first start-to-goal query in order to close the loop and obtain overlapped images.
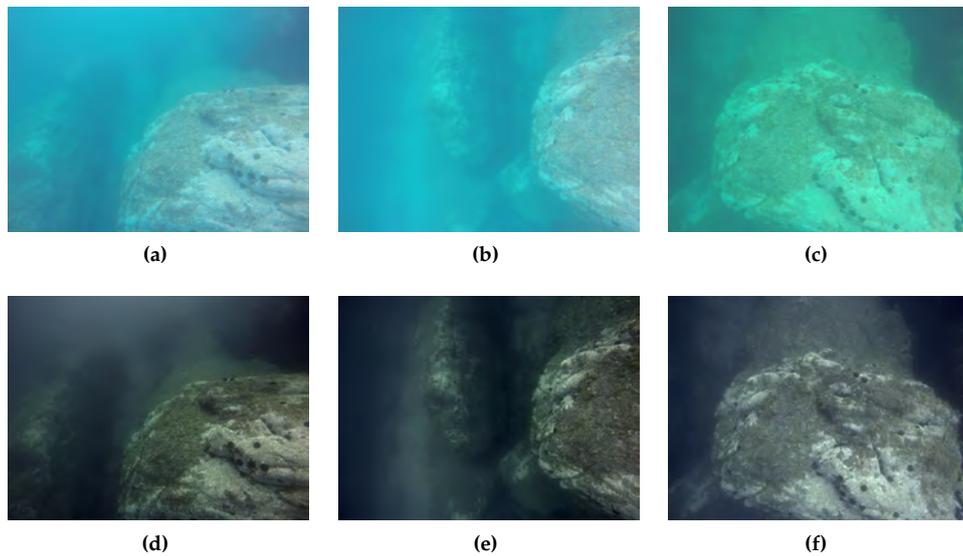
Figure 17. (**a**–**c**) images captured by forward, left and right camera respectively; (**d**–**f**) images after color correction and contrast enhancement.

The reconstruction was obtained using the series of sequential steps explained in Section 3, each producing an intermediate representation, e.g., sparse/dense point clouds and 3D triangle mesh (see Figures 9–11), as well as the final photo-realistic 3D model depicted in Figure 18a, presented in a standard top-down view. The recovered 3D model also enables us to generate arbitrary user-defined views (see Figures 18b,c and 19b, and also supplementary material Video S1 to observe the complete reconstruction). Since the reconstruction is performed exclusively from image information, the areas which were not properly imaged (such as crevices, or areas occluded by other rocks along the robot path) can not be reconstructed. Similarly, the bottom of the canyon was not properly observed on the images due to the poor visibility conditions, thus preventing its reconstruction in high detail.

While a 3D reconstruction could be achieved with a single camera, each of the cameras used in our mission contributed an important part of the final result. This can be seen in Figure 19a where the reconstruction is color coded with respect to the different combinations of cameras used for different areas. It is not surprising that the largest area is reconstructed using the images captured by the right camera, as this camera was within the visibility range of the rocky formations. This was not true for the remaining cameras, as the visibility conditions on several occasions prevented the observation of any distant objects. While the majority of the reconstruction is done using the combination of one or two cameras (either due to the visibility conditions, or small overlap between downward oriented cameras), we can see that the area in which the texture is especially rich (enlarged area in Figure 19b), many points have been reconstructed from different combinations of cameras (see Figure 19c). When features detected are sufficiently dominant, they can be successfully matched across various views from different cameras.

(a)
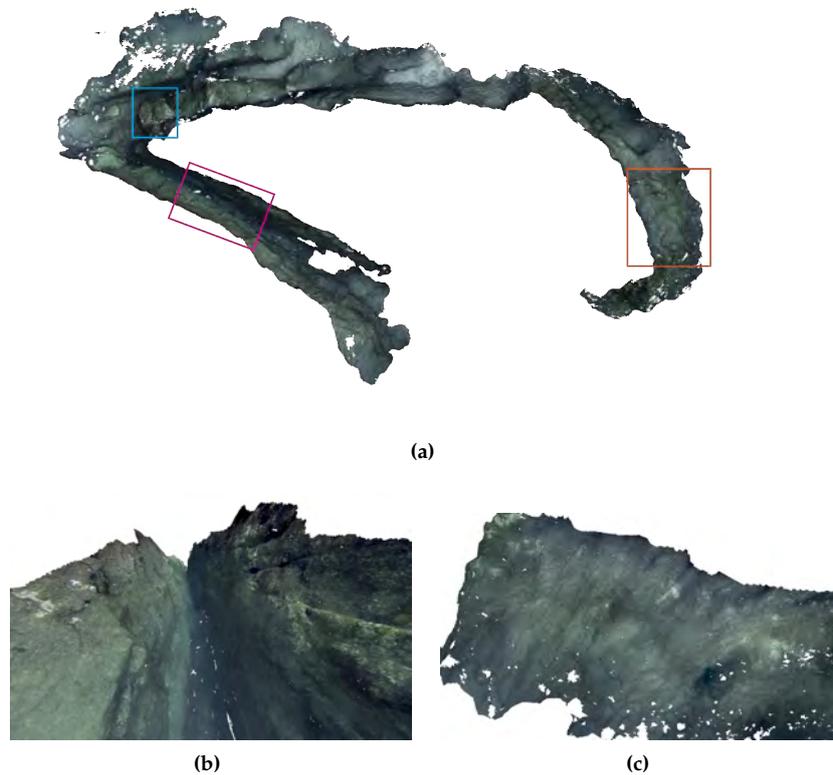

(b)   **(c)**

Figure 18. (**a**) top-down view of the textured 3D model with marked areas additionally depicted (magenta–Figure 18b, orange–Figure 18b and blue–Figure 19b); (**b**) generated view inside the underwater canyon; (**c**) generated view of the external side of the underwater rocky formation.

(a)

(b)

(c)
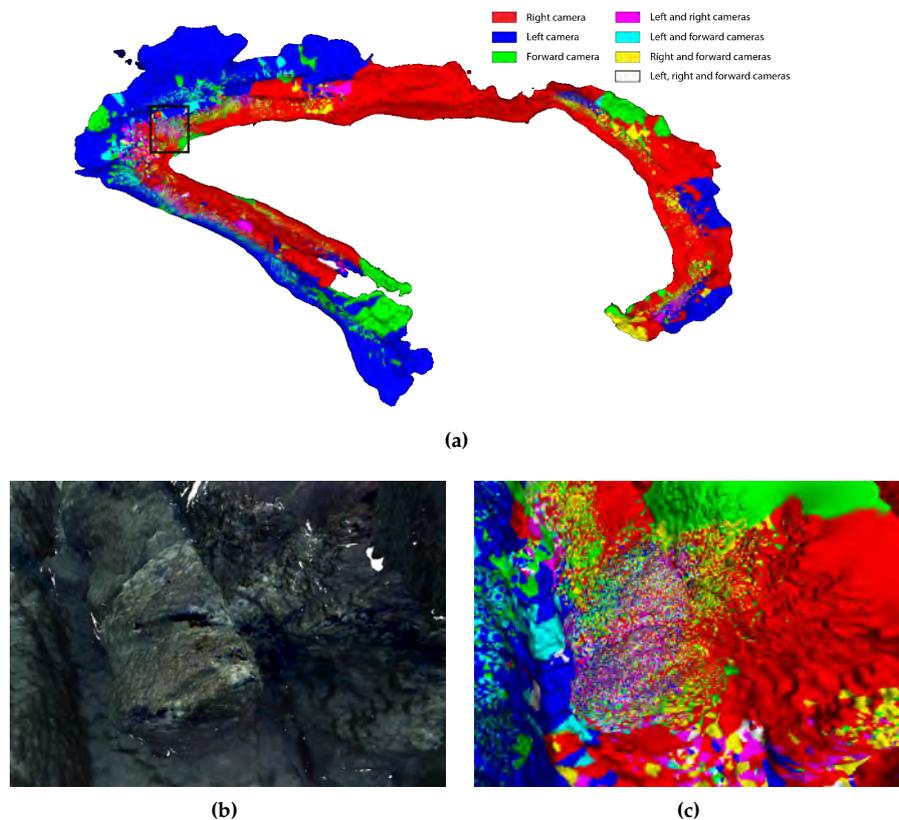
Figure 19. (**a**) color coded 3D reconstruction based on the cameras used in specific areas in top-down view; (**b**) generated view of a marked area in blue in Figure 19a using textured 3D model; (**c**) color coded view of the same view as Figure 19b.

Since the cameras are not connected or synchronized with the AUV, additional information from other sensors cannot be directly imposed in the reconstruction process. As a consequence, the problem solved by the SfM does not allow determination of the real scale of the reconstruction. In order to compare the vehicles and cameras trajectory estimated during the SfM, we manually estimated the scale of the reconstruction. Despite this, Figure 20 presents a visual comparison between them (green for the vehicle and red for the cameras). While both trajectories have a similar shape, it can be clearly observed how the one derived from the cameras is more realistic according to the rock observed in the surface (the rocky formation does not create a vertical wall, which means that the vehicle may have moved further from the visible part of the rock when navigating at 3 m deep), while the one estimated by the AUV's dead reckoning system seems to be colliding with the rock. This latter situation, as explained the previous section, is mainly due to the accumulation of errors in the navigation system.

A final and important observation in Figure 20 is the loss of visibility in the last part of the mission. Even though the path-planning pipeline attempts to guarantee the visibility of the inspected structure, the second query goal guided the vehicle towards the coordinate system origin (see Figure 16c), which clearly moved the vehicle away from the structure. However, even though such (visibility) discontinuity can be appreciated in the cameras' trajectory, it can also be observed how the reconstruction pipeline was able to properly match the subsequent images once the vehicle approached the canyon during the second travel. While in this particular case this prevented us from

successfully enforcing the loop closure, it does, however, demonstrate the pipeline's ability if the visibility conditions are met.

Figure 20. Vehicle's trayectory (green) calculated by its dead reckoning system and the camera's trajectory (red) estimated by the SfM, both overlapping a satellite image of the test scenario.

5. Conclusions

In this paper, we have presented a new end-to-end approach for autonomously mapping unknown underwater natural environments using AUV. To do so, we proposed a framework composed of two main functional pipelines. The first provides the AUV with the capability for creating an acoustic map online, while simultaneously planning collision-free paths. Such functionality is essential for safe navigation in unknown and potentially dangerous areas, and to maintain very short distance to the areas being imaged, as required by optical mapping underwater. Using the gathered image data, the second pipeline builds a photo-realistic 3D model, which can be used as base maps for environmental inspection and subsequent monitoring.

In recent previous work, both pipelines were independently tested in simulated and real-world non-natural (structured) environments. In order to thoroughly validate our approach, this paper presented results obtained in a challenging real-world natural scenario, in which the Sparus II AUV conducted several autonomous missions. The test scenario, containing an underwater canyon between two rocky formations, permitted us to demonstrate the extent of the capabilities of our approach. By successfully navigating through the natural environment, the AUV was able to acquire data subsequently used in the reconstruction of a complex textured 3D model of the area.

Our next effort will focus on using different perception modalities (such as multibeam sonar) to create online acoustic 3D maps, thus enabling missions at different depths. This will permit the path-planning pipeline to attempt maintaining the visibility not only with the inspected structure but also with the sea bottom if desired. Consequently, these kind of missions will allow for better exploitation of the reconstruction pipeline capabilities for representing complex 3D environments.

Abbreviations

The following abbreviations are used in this manuscript:

C-Space	configuration space
RRT	rapidly-exploring random tree
RRT*	asymptotic optimal RRT
1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
OMPL	open motion planning library
DFS	depth-first search
DOF	degrees of freedom
ROS	robot operating system
UUV	unmanned underwater vehicle
ROV	remotely operated vehicle
AUV	autonomous underwater vehicle
DVL	Doppler velocity log
IMU	inertial measurement unit
CIRS	underwater vision and robotics research center
COLA2	component oriented layer-based architecture for autonomy
UWSim	underwater simulator
SIFT	scale-invariant feature transform
SfM	structure from motion
DOG	difference of Gaussians
RANSAC	random sample consensus
AC-RANSAC	a contrario-RANSAC
GPU	graphics processing unit
CAD	computer-aided design

References

1. Dunbabin, M.; Marques, L. Robots for environmental monitoring: Significant advancements and applications. *IEEE Robot. Autom. Mag.* **2012**, *19*, 24–39.
2. Whitcomb, L. Underwater robotics: out of the research laboratory and into the field. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco, CA, USA, 24–28 April 2000; Volume 1, pp. 709–716.
3. Pathak, K.; Birk, A.; Vaskevicius, N. Plane-based registration of sonar data for underwater 3D mapping. In Proceedings of the 2010 IEEE/RSJ International Conference onIntelligent Robots and Systems (IROS), Taipei, Taiwan, 18–22 October 2010; pp. 4880–4885.
4. Hurtós, N.; Cufí, X.; Salvi, J. Calibration of optical camera coupled to acoustic multibeam for underwater 3D scene reconstruction. In Proceedings of the OCEANS 2010 IEEE-Sydney, Sydney, Australia, 24–27 May 2010; pp. 1–7.
5. Coiras, E.; Petillot, Y.; Lane, D.M. Multiresolution 3-D reconstruction from side-scan sonar images. *IEEE Trans. Image Process.* **2007**, *16*, 382–390.
6. Gracias, N.; Ridao, P.; Garcia, R.; Escartin, J.; L'Hour, M.; Cibecchini, F.; Campos, R.; Carreras, M.; Ribas, D.; Palomeras, N.; et al. Mapping the Moon: Using a lightweight AUV to survey the site of the 17th century ship 'La Lune'. In Proceedings of the 2013 MTS/IEEE OCEANS-Bergen, Bergen, Norway, 10–14 June 2013; pp. 1–8.
7. Campos, R.; Garcia, R.; Alliez, P.; Yvinec, M. A surface reconstruction method for in-detail underwater 3D optical mapping. *Int. J. Robot. Res.* **2014**, *34*, 64–89.
8. Massot-Campos, M.; Oliver-Codina, G. Optical Sensors and Methods for Underwater 3D Reconstruction. *Sensors* **2015**, *15*, 31525–31557.

9. Clarke, M.E.; Tolimieri, N.; Singh, H. Using the seabed AUV to assess populations of groundfish in untrawlable areas. In *The future of fisheries science in North America*; Springer: Dordrecht, The Netherlands, 2009; pp. 357–372.

10. Grasmueck, M.; Eberli, G.P.; Viggiano, D.A.; Correa, T.; Rathwell, G.; Luo, J. Autonomous underwater vehicle (AUV) mapping reveals coral mound distribution, morphology, and oceanography in deep water of the Straits of Florida. *Geophys. Res. Lett.* **2006**, *33*, doi:10.1029/2006GL027734.

11. Smale, D.A.; Kendrick, G.A.; Harvey, E.S.; Langlois, T.J.; Hovey, R.K.; Van Niel, K.P.; Waddington, K.I.; Bellchambers, L.M.; Pember, M.B.; Babcock, R.C.; et al. Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). *ICES J. Marine Sci. J. Cons.* **2012**, *69*, doi:10.1093/icesjms/fss082.

12. Yoerger, D.R.; Jakuba, M.; Bradley, A.M.; Bingham, B. Techniques for deep sea near bottom survey using an autonomous underwater vehicle. *Int. J. Robot. Res.* **2007**, *26*, 41–54.

13. Barreyre, T.; Escartín, J.; Garcia, R.; Cannat, M.; Mittelstaedt, E.; Prados, R. Structure, temporal evolution, and heat flux estimates from the Lucky Strike deep-sea hydrothermal field derived from seafloor image mosaics. *Geochem. Geophys. Geosyst.* **2012**, *13*, doi:10.1029/2011GC003990.

14. Yoerger, D.R.; Bradley, A.M.; Walden, B.B.; Cormier, M.H.; Ryan, W.B. Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot. In Proceedings of the ICRA'00 IEEE International Conference on Robotics and Automation, San Francisco, CA, USA, 24–28 April 2000; Volume 2, pp. 1787–1792.

15. Bruno, F.; Gallo, A.; De Filippo, F.; Muzzupappa, M.; Petriaggi, B.D.; Caputo, P. 3D documentation and monitoring of the experimental cleaning operations in the underwater archaeological site of Baia (Italy). In Proceedings of the IEEE Digital Heritage International Congress (DigitalHeritage), Marseille, France, 28 October–1 November 2013; Volume 1, pp. 105–112.

16. Eric, M.; Kovacic, R.; Berginc, G.; Pugelj, M.; Stopinšek, Z.; Solina, F. The impact of the latest 3D technologies on the documentation of underwater heritage sites. In Proceedings of the IEEE Digital Heritage International Congress (DigitalHeritage), Marseille, France, 28 October–1 November 2013; Volume 2, pp. 281–288.

17. Shihavuddin, A.; Gracias, N.; Garcia, R.; Gleason, A.C.R.; Gintert, B. Image-Based Coral Reef Classification and Thematic Mapping. *Remote Sens.* **2013**, *5*, 1809–1841.

18. Beijbom, O.; Edmunds, P.J.; Kline, D.I.; Mitchell, B.G.; Kriegman, D. Automated annotation of coral reef survey images. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1170–1177.

19. Shihavuddin, A.; Gracias, N.; Garcia, R.; Escartin, J.; Pedersen, R. Automated classification and thematic mapping of bacterial mats in the North Sea. In Proceedings of the 2013 MTS/IEEE OCEANS, Bergen, Norway, 10–14 June 2013.

20. Bewley, M.; Douillard, B.; Nourani-Vatani, N.; Friedman, A.; Pizarro, O.; Williams, S. Automated species detection: An experimental approach to kelp detection from sea-floor AUV images. In Proceedings of the Australasian Conference on Robotics & Automation, Wellington, New Zealand, 3–5 December 2012.

21. Bryson, M.; Johnson-Roberson, M.; Pizarro, O.; Williams, S. Automated registration for multi-year robotic surveys of marine benthic habitats. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 3344–3349.

22. Delaunoy, O.; Gracias, N.; Garcia, R. Towards Detecting Changes in Underwater Image Sequences. In Proceedings of the OCEANS 2008-MTS/IEEE Kobe Techno-Ocean, Kobe, Japan, 8-11 April 2008; pp. 1–8.

23. Galceran, E.; Campos, R.; Palomeras, N.; Ribas, D.; Carreras, M.; Ridao, P. Coverage Path Planning with Real-time Replanning and Surface Reconstruction for Inspection of Three-dimensional Underwater Structures using Autonomous Underwater Vehicles. *J. Field Robot.* **2014**, *32*, 952–983.

24. Mallios, A.; Ridao, P.; Ribas, D.; Carreras, M.; Camilli, R. Toward autonomous exploration in confined underwater environments. *J. Field Robot.* **2015**, *7*, doi:10.1002/rob.21640.

25. Hernández, J.D.; Vidal, E.; Vallicrosa, G.; Galceran, E.; Carreras, M. Online path planning for autonomous underwater vehicles in unknown environments. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 1152–1157.

26. Hernández, J.D.; Istenic, K.; Gracias, N.; García, R.; Ridao, P.; Carreras, M. Autonomous Seabed Inspection for Environmental Monitoring. In *ROBOT2015: Second Iberian Robotics Conference*; Reis, L.P., Moreira, A.P., Lima, P.U., Montano, L., Muñoz-Martinez, V., Eds.; Springer International Publishing: Lisbon, Portugal, 2016; pp. 27–39.

27. Hornung, A.; Wurm, K.M.; Bennewitz, M.; Stachniss, C.; Burgard, W. OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonom. Robots* **2013**, *34*, 189–206.

28. Karaman, S.; Frazzoli, E. Sampling-based Algorithms for Optimal Motion Planning. *Int. J. Robot. Res.* **2011**, *30*, 846–894.

29. LaValle, S.M.; Kuffner, J.J. Randomized Kinodynamic Planning. *Int. J. Robot. Res.* **2001**, *20*, 378–400.

30. Karaman, S.; Frazzoli, E. Incremental Sampling-based Algorithms for Optimal Motion Planning. In Proceedings of the Robotics: Science and Systems (RSS), Zaragoza, Spain, 27–30 June 2010.

31. Karaman, S.; Walter, M.R.; Perez, A.; Frazzoli, E.; Teller, S. Anytime Motion Planning using the RRT*. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 1478–1483.

32. Bekris, K.E.; Kavraki, L.E. Greedy but Safe Replanning under Kinodynamic Constraints. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Roma, Italy, 10–14 April 2007; pp. 704–710.

33. Bohlin, R.; Kavraki, L.E. Path planning using lazy PRM. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco, CA, USA, 24–28 April 2000; Volume 1, pp. 521–528.

34. Elibol, A.; Gracias, N.; Garcia, R. Fast topology estimation for image mosaicing using adaptive information thresholding. *Robot. Auton. syst.* **2013**, *61*, 125–136.

35. Prados, R.; Garcia, R.; Gracias, N.; Escartin, J.; Neumann, L. A novel blending technique for underwater gigamosaicing. *IEEE J. Ocean. Eng.* **2012**, *37*, 626–644.

36. Singh, H.; Howland, J.; Pizarro, O. Advances in large-area photomosaicking underwater. *IEEE J. Ocean. Eng.* **2004**, *29*, 872–886.

37. Szeliski, R. Image mosaicing for tele-reality applications. In Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; pp. 44–53.

38. Ferrer, J.; Elibol, A.; Delaunoy, O.; Gracias, N.; Garcia, R. Large-area photo-mosaics using global alignment and navigation data. In Proceedings of the MTS/IEEE OCEANS Conference, Vancouver, BC, Canada, 29 September–4 October 2007; pp. 1–9.

39. Nicosevici, T.; Gracias, N.; Negahdaripour, S.; Garcia, R. Efficient three-dimensional scene modeling and mosaicing. *J. Field Robot.* **2009**, *26*, 759–788.

40. Campos, R.; Gracias, N.; Ridao, P. Underwater Multi-Vehicle Trajectory Alignment and Mapping Using Acoustic and Optical Constraints. *Sensors* **2016**, *16*, 387, doi:10.3390/s16030387.

41. Garcia, R.; Campos, R.; Escartın, J. High-resolution 3D reconstruction of the seafloor for environmental monitoring and modelling. In Proceedings of the IROS Workshop on Robotics for Environmental Monitoring, San Francisco, CA, USA, 25–30 September 2011.

42. Hartley, R. I.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2004.

43. Cavan, N. Reconstruction of 3D Points from Uncalibrated Underwater Video. Ph.D. Thesis, University of Waterloo, Waterloo, ON, Canada, April 2011.

44. Bouguet, J.Y. *Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker Description of the Algorithm*; Intel Corporation: Santa Clara, CA, USA, 2001; Volume 5, p. 4.

45. Shi, J.; Tomasi, C. Good features to track. In Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94), Seattle, WA , USA, 21–23 June 1994; pp. 593–600.

46. Bryson, M.; Johnson-Roberson, M.; Pizarro, O.; Williams, S.B. True Color Correction of Autonomous Underwater Vehicle Imagery. *J. Field Robot.* **2015**, doi:10.1002/rob.21638.

47. Bianco, G.; Gallo, A.; Bruno, F.; Muzzupappa, M. A comparison between active and passive techniques for underwater 3D applications. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*; **2011**, *38*, 357–363.

48. Jaffe, J.S. Computer modeling and the design of optimal underwater imaging systems. *IEEE J. Ocean. Eng.* **1990**, *15*, 101–111.

49. Mobley, C.D. *Light and Water: Radiative Transfer in Natural Waters*; Academic Press: San Diego, USA, 1994.

50. Andono, P.N.; Purnama, I.; Hariadi, M. Underwater Image Enhancement Using Adaptive Filtering for Enhanced Sift-Based Image Matching. Available online: http://www.jatit.org/volumes/Vol51No3/7Vol51No3.pdf (accessed on 1 October 2015).

51. Ruderman, D.L.; Cronin, T.W.; Chiao, C.C. Statistics of cone responses to natural images: Implications for visual coding. *J. Opt. Soc. Am. A* **1998**, *15*, 2036–2045.

52. Treibitz, T.; Schechner, Y.Y.; Kunz, C.; Singh, H. Flat refractive geometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 51–65.

53. Gawlik, N. *3D Modelling of Underwater Archaeological Artefacts*; Institutt for bygg, anlegg og transport: Trondheim, Trondheim, 2014.

54. Kwon, Y.H. Object plane deformation due to refraction in two-dimensional underwater motion analysis. *J. Appl. Biomech.* **2010**, *15*, 396–403.

55. Wu, C. Critical configurations for radial distortion self-calibration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 25–32.

56. Bouguet, J.Y. Camera Calibration Toolbox for Matlab. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/ (accessed on 1 October 2015).

57. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In *Vision Algorithms: Theory and Practice*; Springer: Heidelberg, Germany, 1999; pp. 298–372.

58. Wu, C. SiftGPU: A GPU Implementation of Scale Invariant Feature Transform (SIFT). Available online: http://www.cs.unc.edu/~ccwu/siftgpu/ (accessed on 12 October 2015).

59. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

60. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.

61. 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.

62. Moisan, L.; Moulon, P.; Monasse, P. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Process. On Line* **2012**, *2*, 56–73.

63. La Bibliothèque OpenMVG: Open Source Multiple View Geometry. Available online: https://hal.archives-ouvertes.fr/hal-00829332/file/04.pdf (accessed on 21 September 2015).

64. Moulon, P.; Monasse, P.; Marlet, R. Adaptive structure from motion with a contrario model estimation. In *Computer Vision—ACCV 2012*; Springer: Heidelberg, Germany, 2012; pp. 257–270.

65. Remondino, F.; Spera, M.G.; Nocerino, E.; Menna, F.; Nex, F.; Gonizzi-Barsanti, S. Dense image matching: Comparisons and analyses. In Proceedings of the IEEE Digital Heritage International Congress (DigitalHeritage), Marseille , France, 28 October–1 November 2013; Volume 1, pp. 47–54.

66. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer Science & Business Media: London, UK, 2010.

67. Ahmadabadian, A.H.; Robson, S.; Boehm, J.; Shortis, M.; Wenzel, K.; Fritsch, D. A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. *ISPRS J. Photogramm. Remote Sens.* **2013**, *78*, 157–167.

68. Dall'Asta, E.; Roncella, R. A comparison of semiglobal and local dense matching algorithms for surface reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 187–194.

69. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.

70. Pierrot-Deseilligny, M.; Paparoditis, N. A multiresolution and Optimization-Based Image Matching Approach: An Application to Surface Reconstruction from SPOT5-HRS Stereo Imagery. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.4955&rep=rep1&type=pdf (accessed on 1 October 2015).

71. Pierrot-Deseilligny, M.; De Luca, L.; Remondino, F. Automated image-based procedures for accurate artifacts 3D modeling and orthoimage generation. *Geoinform. FCE CTU* **2011**, *6*, 291–299.

72. Pierrot-Deseilligny, M.P. MicMac, Un Logiciel Pour La Mise En Correspondance Automatique D'images Dans Le Contexte Géographique. Available online: http://logiciels.ign.fr/IMG/pdf/Bulletin-Info-IGPN-2007.pdf (accessed on 15 October 2015).

73. Roy, S.; Cox, I.J. A maximum-flow formulation of the n-camera stereo correspondence problem. In Proceedings of the 6th International Conference on Computer Vision, Bombay, India, 4–7 January 1998; pp. 492–499.

74. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson surface reconstruction. In Proceedings of the 4th Eurographics Symposium on Geometry Processing, Sardinia, Italy, 26–28 June 2006; Volume 7.

75. Kazhdan, M.; Hoppe, H. Screened poisson surface reconstruction. *ACM Trans. Graph. (TOG)* **2013**, *32*, 29, doi:10.1145/2487228.2487237.

76. Campos, R.; Garcia, R.; Nicosevici, T. Surface reconstruction methods for the recovery of 3D models from underwater interest areas. In Proceedings of the IEEE OCEANS, Santander, Spain, 6–9 June 2011.

77. Waechter, M.; Moehrle, N.; Goesele, M. Let there be color! Large-scale texturing of 3D reconstructions. In *Computer Vision–ECCV*; Springer: Cham, Switzerland, 2014; pp. 836–850.

78. Fuhrmann, S.; Langguth, F.; Moehrle, N.; Waechter, M.; Goesele, M. MVE—An image-based reconstruction environment. *Comput. Graph.* **2015**, *53*, 44–53.

79. Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. *ACM Trans. Graph. (TOG)* **2003**, 22, 313–318.

80. CIRS: Girona Underwater Vision and Robotics. Available online: http://cirs.udg.edu/ (accessed 2015–2016).

81. Palomeras, N.; El-Fakdi, A.; Carreras, M.; Ridao, P. COLA2: A Control Architecture for AUVs. *IEEE J. Ocean. Eng.* **2012**, *37*, 695–716.

82. Prats, M.; Perez, J.; Fernandez, J.J.; Sanz, P.J. An open source tool for simulation and supervision of underwater intervention missions. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 2577–2582.

83. Sucan, I.A.; Moll, M.; Kavraki, L.E. The Open Motion Planning Library. *IEEE Robot. Autom. Mag.* **2012**, *19*, 72–82.

84. Google Maps 2016. Sant Feliu de Guíxols, Girona, Spain 41°46′55.4″N, 3°02′55.1″E. Rocky Formation. Available online: https://www.google.es/maps/ (accessed on 2 May 2016).

3

Automatic Scale Estimation of Structure from Motion based 3D Models using Laser Scalers

I̶n this chapter we address the the problem of accurate scaling of 3D models created with images from monocular cameras in GPS-denied environments, such as in underwater applications. We propose two novel methods for scaling SfM-based 3D models using commonly available laser scalers. The first is a completely unrestricted method that allows the use of any laser setup, while the second method reduces the required knowledge of the setup's geometry by assuming the laser parallelism and equidistance of the laser origin to the camera. Together the methods encompass the majority of laser scaler configurations. The methods are extensively evaluated using data generated from a realistic 3D model as well as data collected during an oceanographic cruise in 2017.

Automatic scale estimation of structure from motion based 3D models using laser scalers in underwater scenarios

Klemen Istenič[a,b,*], Nuno Gracias[a], Aurélien Arnaubec[c], Javier Escartín[d], Rafael Garcia[a]

[a] Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, Edifici P-IV, Campus de Montilivi, 17071 Girona, Spain
[b] Coronis Computing, S.L., Science and Technological Park of UdG, Carrer Pic de Peguera, 15, 17003 Girona, Spain
[c] IFREMER, Ctr Mediterranee, Unité Syst. Marins, CS 20330, F-83507 La Seyne Sur Mer, France
[d] Université de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005 Paris, France

ABSTRACT

Improvements in structure-from-motion techniques are enabling many scientific fields to benefit from the routine creation of detailed 3D models. However, for a large number of applications, only a single camera is available for the image acquisition, due to cost or space constraints in the survey platforms. Monocular structure-from-motion raises the issue of properly estimating the scale of the 3D models, in order to later use those models for metrology. The scale can be determined from the presence of visible objects of known dimensions, or from information on the magnitude of the camera motion provided by other sensors, such as GPS.

This paper addresses the problem of accurately scaling 3D models created from monocular cameras in GPS-denied environments, such as in underwater applications. Motivated by the common availability of underwater laser scalers, we present two novel approaches which are suitable for different laser scaler configurations. A fully unconstrained method enables the use of arbitrary laser setups, while a partially constrained method reduces the need for calibration by only assuming parallelism on the laser beams and equidistance with the camera. The proposed methods have several advantages with respect to existing methods. By using the known geometry of the scene represented by the 3D model, along with some parameters of the laser scaler geometry, the need for laser alignment with the optical axis of the camera is eliminated. Furthermore, the extremely error-prone manual identification of image points on the 3D model, currently required in image-scaling methods, is dispensed with.

The performance of the methods and their applicability was evaluated both on data generated from a realistic 3D model and on data collected during an oceanographic cruise in 2017. Three separate laser configurations have been tested, encompassing nearly all possible laser setups, to evaluate the effects of terrain roughness, noise, camera perspective angle and camera-scene distance on the final estimates of scale. In the real scenario, the computation of 6 independent model scale estimates using our fully unconstrained approach, produced values with a standard deviation of 0.3%. By comparing the values to the only other possible method currently usable for this dataset, we showed that the consistency of scales obtained for individual lasers is much higher for our approach (0.6% compared to 4%).

1. Introduction

In increasing number of remote sensing applications photo-grammetry is used to obtain reliable geometric information about the environment. These optical-based reconstruction procedures, generally based on the Structure from Motion (SfM) approach, have gained significant popularity due to multiple factors. The improvements in both speed and robustness of many image processing techniques (Snavely et al., 2008; Remondino et al., 2008; Agarwal et al., 2009; Triggs et al., 1999) together with the increased computational capabilities of commonly available processing hardware, enable nowadays nearly black-box type of data processing, where there is little to no need for user intervention. The abundance of low cost cameras that can easily be mounted on a variety of vehicles, or used hand-held, has further

* Corresponding author at: Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, Edifici P-IV, Campus de Montilivi, 17071 Girona, Spain.

E-mail addresses: klemen.istenic@gmail.com (K. Istenič), ngracias@silver.udg.edu (N. Gracias), aurelien.arnaubec@ifremer.fr (A. Arnaubec), escartin@ipgp.fr (J. Escartín), rafael.garcia@udg.edu (R. Garcia).

Fig. 1. ROV VICTOR 6000 (IFREMER), used among other, in the SUBSAINTES 2017 cruise (doi:10.17600/17001000).

spearheaded the widespread application of these techniques in a variety of fields (e.g., Wallace et al., 2016; Javernick et al., 2014; Anderson and Gaston, 2013).

Concurrently, the field of underwater photogrammetry has also grown considerably with the availability of underwater vehicles. Whereas traditional aerial and terrestrial vehicles are increasingly equipped with single or multi-camera set-ups (e.g., stereo cameras, multi-camera systems), most underwater remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) that are nowadays used in scientific missions (e.g., VICTOR 6000 from IFREMER depicted in Fig. 1) have limited optical sensing capabilities. Common optical systems consist of a single main camera used by the ROV-pilot or, in the case of the larger workclass ROVs, also of additional cameras for maneuvering. As these are typically unsynchronized and have non-overlapping fields-of-view, they are not suited for stereo image processing. While multi-camera underwater metrology systems are starting to appear as commercially available products and services (Rovco, 2019; Comex, 2019), such systems are still too large and expensive for most ROV science applications. Nonetheless, the ability to produce accurate 3-dimensional (3D) models from monocular cameras despite the unfavorable properties of the water medium (i.e., light attenuation and scattering, among other effects) has given scientists unprecedented access to the underwater environment and its ecosystems, from shallow waters (Pizarro et al., 2017; Storlazzi et al., 2016; Rossi et al., 2019) to the deep ocean (Bingham et al., 2010; Escartín et al., 2016; Bodenmann et al., 2017).

Performing SfM based reconstruction using single camera imagery has an important limitation as it precludes obtaining a metric scale of the resulting model. The image formation process of projecting the 3D world onto 2-dimensional (2D) image planes obviously causes the loss of a dimension. When performing the reconstruction, this results in scale ambiguity, i.e. the estimated parameters of 3D structure and camera trajectory can be multiplied by an arbitrary factor and still give rise to the same image observations (Lourakis and Zabulis, 2013; Hartley and Zisserman, 2003). This also precludes or at least limits the possibility of conducting quantitative measurements based on geometric parameters (e.g., distances, areas, volumes, etc.) obtained from the models. To resolve the ambiguity, a general trend in aerial and terrestrial problems is to fuse the image measurements with other sensors (e.g., inertial navigation system (INS) (Spaenlehauer et al., 2017; Zhang and Singh, 2015) and Global Navigation Satellite System (GNSS) (SSoloviev and Venableoloviev and Venable, 2010; Mian et al., 2016)) or using ground control points (GCPs) (Eltner and Schneider, 2015; Mertes et al., 2017). These control points are extremely hard, if not impossible, to establish underwater, while the absorption of

electromagnetic waves in water prevents the use of GPS. Hence the scale is normally disambiguated either using INS (Sedlazeck et al., 2009; Pizarro et al., 2009; Campos et al., 2016) or through the introduction of known distances between points in the scene (Garcia et al., 2011). It is worth noting that reliable displacement information may not be available in smaller ROVs, since this normally requires a dedicated INS complemented with a Doppler Velocity Log (DVL) (Ribas et al., 2011) and there are rarely any known measurements readily available in real underwater scenarios. The scale is therefore often determined by placing objects with known dimensions (e.g., scaling cube (Cocito et al., 2003), locknuts (Kalacska et al., 2018), graduated bars (Neyer et al., 2018), etc.) into the scene. While such an approach does not require any additional equipment (with the exception of auxiliary objects), it does however involve their transport and placement, which can be challenging in deep-sea environments.

Alternatively, the distance between known points on the model can also be established from the projections of laser beams with known geometry (Robert et al., 2017; Bergmann et al., 2011; Tusting and Davis, 1992). The use of laser scalers to provide an absolute size reference in photographs is one of their most widespread applications (Tusting and Davis, 1992, 1993). Their initial use dates back to the late 1980s (Tusting and Davis, 1986; Caimi and Tusting, 1987). To compensate for the lack of knowledge about the scene and camera-scene distance, these methods require a perfect alignment of parallel lasers with the camera, planarity of the scene surface and perpendicularity between the camera and the scene. Comparing the spacing between two laser spots on the image and the known beam spacing, any measurement in the plane of the lasers, regardless of the camera-to-scene range, should be correctly estimated.

Seen as the most restrictive requirement, the necessity of perpendicularity between the optical axis of the camera and the plane of the scene has been addressed in various approaches with the introduction of additional lasers and sensors. Wakefield and Genin (1987) first introduced the idea of perspective grids to enable oblique camera views. Although being an improvement, the method imposed additional constraints on the camera-scene distance (altitude) and fixed inclination angle. To provide additional information about the camera-scene relationship, more lasers have also been added to the systems. A configuration consisting of three lasers, two aligned with the optical axis of the camera and a third laser oriented at an angle, has been described by Davis and Tusting (1991). It enables the estimation of range and size of objects from direct scaling of the position of the light spots on the image. An underwater photogrammetric system using several sensors to provide precision navigation for benthic surveys is described in Kocak et al. (2002, 2004). One of them, the ring laser gyroscope, made for measuring pitch/roll motions is integrated into a custom software package which establishes the scale reference from the projections of the three beam laser system. To enable the measurement of distance between any two points on the image, Pilgrim et al. (2000) presented a multi-laser approach. It gains information about the camera's inclination angle and distance to the scene by using four parallel lasers positioned equidistant from the camera center together with either a fifth laser set at an angle parallel to the bottom or a side pair, similar to the three-beam approach. The method works under the assumption of scene flatness and the restraint of the camera in either pan or tilt planes with respect to the sea bottom. A more versatile method capable of determining an arbitrary tilt of a surface was presented by Davis and Tusting (1991) which requires four parallel lasers aligned with the optical axis of the camera.

Due to the lack of a better approach, image-scaling methods are still commonly used for scaling 3D models, and therefore require not only that the image containing the projections of lasers be acquired in flat areas of the scene, but also complex laser alignment with the optical axis of the camera. In real scientific cruises these strict rigidity constraints can be nearly impossible to maintain, due to the frequent need to mount and dismount equipment, especially when the camera is not
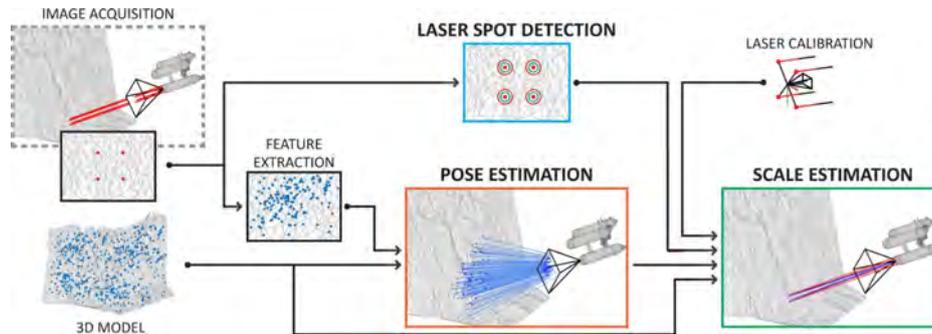
Fig. 2. Flowchart of the scale estimation process depicting three crucial steps in scale estimation: laser spot detection, pose estimation, and scale estimation.

rigidly attached to the laser scalers. As obtaining accurate geometrical information would entail repetitive calibration procedures, it significantly limits its usability. Furthermore, given that the image-scaling techniques only provide the estimated distance between points on an image, this information is not directly related to the model itself. In order to scale any model, a separate identification of these laser points has to be carried out on the model itself. As the identification of image points on the model is done manually, it is extremely error-prone and time consuming.

The main goal of this paper is to present two novel automated approaches to solve the scaling problem for SfM based 3D models, using commonly available laser scalers. The image information is exploited beyond the automatic location of laser spots, compensating for the known geometry of the laser scalers. The need for laser alignment with the optical axis, scene geometry or camera position are thus abolished together with the error-prone manual identification of 3D points on the model.

Each of the two proposed methods (i.e., fully unconstrained and partially constrained) is suitable for a different laser scaler configuration. While the fully unconstrained approach enables an arbitrary laser setup, the required rigidity between the lasers and the camera can be extremely limiting in real scenarios. To overcome this, we also present an alternative approach in which the required relation of the lasers to the camera is significantly relaxed at the cost of requiring the lasers to be parallel among themselves (but not necessarily with the optical axis). As the fully unconstrained method utilizes a fully-determined laser geometry, it is able to estimate the scale using a single laser, while the partial method requires a laser pair. Any additional laser measurements are used to further reduce the potential effect of noisy laser spot detections. These methods are considered universal, as they can be applied to standard imagery acquisitions, and are not linked to data acquired with specific sensors or hardware (e.g., stereo cameras). Hence, it is possible to process legacy data from previous missions acquired using different vehicles and imaging systems. However, as the novel methods require a description of a scene in a form of a 3D model, they cannot be utilized on moving objects (e.g., fish, benthic species, soft corals, etc.) or in highly dynamic environments.

The results of our methods are validated using a 3D model constructed using real underwater data and comparing them to the results which would have been obtained using an image-scaling method supporting an arbitrary tilt of the surface (Davis and Tusting, 1991). The effects of noise, camera perspective angle and camera-scene distance on our process and final estimates of scale are further analyzed. Finally, the results of using our method to scale a model reconstructed from data acquired during the SUBSAINTES 2017 cruise (doi:10.17600/17001000) (Escartín et al., 2017) are presented.

2. Scaling of SfM-based 3D models

Optical-based 3D models are produced using a set of images through a sequential series of steps. A sparse set of 3D points representing the general 3D geometry of the scene can be obtained by exploiting multiple projections of the same 3D point in overlapping images through the equations of projective geometry (Hartley and Zisserman, 2003). By extracting salient features and matching them across the image set, the 3D locations of these points (the structure) are estimated together with the camera parameters (the motion) through a technique called Structure from Motion (SfM). An accurate and highly detailed description of the scene is subsequently obtained through a multi-view stereo densification process followed by an estimation of a surface from the noisy point cloud. The final photo-realistic 3D model representation is achieved by finding a consistent high-quality texture through seamlessly mapping input images to a high-resolution triangle representation of the surface. If the imagery used in the process was acquired using one or more unsynchronized cameras, and no other auxiliary data is used, it is impossible to determine the correct scale of the model. Such a result can be visually pleasing but cannot be used for further scientific purposes where knowledge of the distances, areas and volumes is required. Therefore, a scale estimation step is vital in the reconstruction for scientific purposes.

Nowadays, the most common uses of laser scalers are based on the image-scaling multi-laser approaches introduced by Pilgrim et al. (2000) and Davis and Tusting (1991). The requirements associated with these methods, i.e. laser alignment with the optical axis and manual identification of the image points on the 3D models, while originally reasonable, are becoming constricting in an increasing number of situations in which data for photogrammetry can be collected.

In this section, we present two novel methods for scale estimation, namely the fully unconstrained method (FUM) and the partially constrained method (PCM), suitable for different laser-scaler configurations and scenarios. Both methods, based on computer vision techniques of image localization and ray casting, exploit the information acquired with an optical image in which the intersection of laser beams with the scene (laser spots) are visible. Both methods consist of three main steps, as depicted in Fig. 2. The two initial steps are identical in both methods. First, a laser detection method is required to determine the locations of laser spots on an image. Secondly, the pose of the camera (wrt. the 3D model), at the moment at which the image was acquired, is estimated through a feature-based localization process. These estimations are used in the third step, which differs between the two methods and depends on available laser configuration information. The scale of the model is finally computed after determining the 3D position of laser beams intersecting with the scene.

It is worth noting that our approaches are independent of the method used for detecting laser spots on the image. Laser spots can be selected either manually, through a simple method (e.g., color
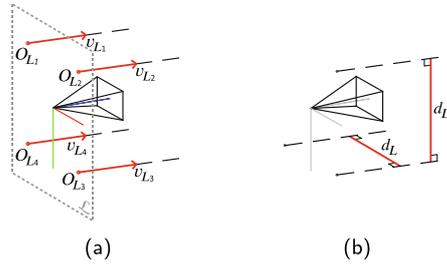
Fig. 3. (a) Fully- and (b) partially-calibrated setups of the measuring device (optical camera and separate lasers) with the required information marked in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thresholding) or even with a more complex approach such as machine learning (Rzhanov et al., 2005; Schoening et al., 2015).

2.1. Measuring device

The measuring setup required consists of two devices commonly used in underwater surveying using ROVs and AUVs: A laser scaler, which can contain a variable number of lasers, and a monocular optical camera. If the laser geometry (origins O_L and directions v_L) with respect to the optical axis of the camera are known, the setup is considered fully calibrated (Fig. 3a). The origins are defined as points on a plane $\mathcal{L}$, which is perpendicular to the optical axis of the camera and contains the optical center, while the directions are unit vectors expressed wrt. the camera's optical axis.

Depending on the circumstances (e.g., multiple dives involving mounting and dismounting of equipment with associated misalignments), the strict rigidity constraints between the lasers and the camera are very difficult to maintain, especially if the camera and the laser scaler are not rigidly coupled. As any change would thus entail a new calibration procedure, which may not be feasible in certain situations, we also present an alternative approach, in which laser pairs have to be parallel, along with the sole condition that the camera be equidistant to their origins (Fig. 3b). As there is no requirement of parallelism between the laser beams and the optical axis of the camera, this partially constrained approach permits alterations in the orientation between the camera and laser scaler, making it more suitable for scenarios with multiple mounting and dismounting operations, or situations in which an accurate calibration procedure is not possible or unavailable. These relaxed constraints render the system more usable in practice.

2.2. Pose estimation

The scale estimation process requires the knowledge of the camera pose $P = [R^T | - R^T t] \in \mathbf{SE}(3)$ defined as the projection from the world to the camera frame at the moment the image was taken. As these images contain laser spots, they do not reflect the real appearance of the environment and are, as such, considered undesirable in the 3D reconstruction process, specially in the densification and texture mapping steps. To estimate the poses, images can potentially still be included in the SfM step (and excluded from rest) or, alternatively, the pose can be computed through a separate feature-based image localization method presented here.

Salient 2D features extracted from the image, are matched with a full set of features associated with the model's sparse set of 3D points. Feature detection and matching procedures can be adjusted for each specific dataset, and do not influence the scale estimation process, as long as it is possible to produce successful pairs of 3D-2D observations ($\mathcal{F} = \{X_k, x_j\}$). Such matches are then exploited to obtain an initial estimate of camera extrinsic parameters P (and possible camera intrinsics K). In cases in which the camera has been calibrated, the solution is

obtained by solving a minimal case ($n = 3$) of the Perspectiven-Point (PnP) problem (Ke and Roumeliotis, 2017), while alternatively a Direct Linear Transform (DLT) (Hartley and Zisserman, 2003) algorithm can be used. As feature observations are noisy and might contain outliers, the process must be carried out in conjunction with a robust estimation method A Contrario Ransac (AC-RANSAC) (Moisan et al., 2012). Initial parameter values are subsequently refined through a non-linear optimization. Using Bundle Adjustment (BA) the re-projection error of known (and fixed) 3D points and their 2D observation is minimized:

$$\min_{P,K} \sum_{\mathcal{F}} \|x_j - \text{proj}(K, P, \mathbf{X}_k)\|^2. \tag{1}$$

2.3. Scale estimation

The scale of a 3D model is obtained as the ratio between a known quantity m and its model based estimate $\hat{m}$:

$$s = \frac{m}{\hat{m}}. \tag{2}$$

Using the location of recorded and detected laser spots x_L and previously estimated parameters of the camera $\{K, P\}$, it is possible to predict the geometry of the laser scaler which produced the recorded results. Given that the prediction is based on the 3D model, it is directly affected by the scale of the model and can therefore be used to determine it. Depending on the availability of information about the geometry of the lasers and the camera, we can either use the distance between the laser origins and camera's optical center (FUM) or the perpendicular distance between the two parallel beams (PCM).

2.3.1. Fully unconstrained method

When the complete laser geometry (origins O_L and directions v_L) are known, the position from where the lasers must be emitted $\hat{O}_L$ in order to produce the observed result can be determined regardless of potential non-parallelism between the lasers (Fig. 4). The position of origin of each laser can be estimated independently by exploiting the known direction of the laser beam and the determined position of the laser beam intersection with the scene X_L. As this point is seen on the image, the actual 3D point X_L had to be in the line-of-sight of the camera and can therefore be deduced using a ray casting procedure. The location is computed by finding the first surface of the 3D model which is intersected by a ray originating in the camera center and passing through the location of the detected laser spot on the image. Subsequently, to obtain the location of the origin, the point X_L expressed in the camera frame is back-projected according to a known direction of the beam v_L onto the plane $\mathcal{L}$ (Eq. (3)). Once known, the scale can be determined by comparing the displacement $\hat{m}_L = \|\hat{O}_L\|$ with its *a priori* known value m_L.

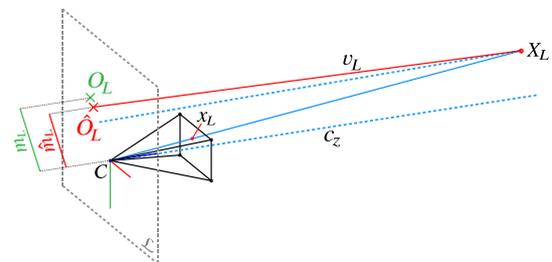**Fig. 4.** Scale estimation procedure using the fully unconstrained approach, based on the 3D model and optical image depicting the laser beam projection of the laser intersection with the scene. The displacement of the predicted laser origin $\hat{m}_L$, obtained by projecting the 3D point X_L onto the plane $\mathcal{L}$ according to the known direction of the laser beam v_L, is compared to its known value m_L.
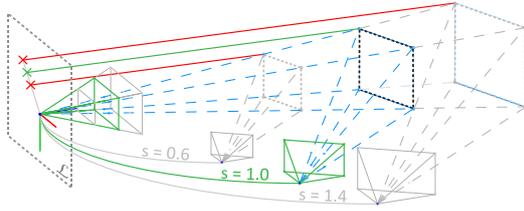
Fig. 5. The effect of scale ambiguity on a 3D reconstruction and its projection on images. The structure and motion can be multiplied by an arbitrary factor (incorrect – light gray, correct – black) and still produce projections in the same positions on the images. The error in scale can be determined by comparing the predicted locations of the laser origin (incorrect – red, correct – green cross). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\widehat{O}_L = PX_L - \frac{PX_L \cdot c_z}{v_L \cdot c_z} v_L, \tag{3}$$

where c_z represents the optical axis of the camera.

Fig. 5 depicts the effect different scaling factors (affecting the 3D model) have on the displacement of the predicted laser origin (represented as cross on plane $\mathcal{L}$). Due to the scale ambiguity, all variations of the model (light gray dotted line), are valid solutions of the 3D reconstruction process. As shown, only the laser intersection point (blue) obtained using the model with the true scale (dark gray dotted line) produces the correct prediction of the laser origin on the plane $\mathcal{L}$ (green cross) by using known directions of the laser beams.

2.3.2. Partially constrained method

While fully unconstrained method enables the use of an arbitrary laser setup, the required rigidity between the lasers and the camera can be extremely limiting in certain real scenarios. To alleviate this, we present an alternative approach, in which the required relation between the camera and the lasers is significantly less rigid. This approach only requires the two lasers to be parallel and equidistant from the camera. As opposed to the existing image-scaling methods, the lasers do not have to be aligned with the optical axis of the camera. The scale of the model is therefore estimated by comparing a known perpendicular distance between the two parallel beams to the one estimated from the image and the model $\widehat{d}_L$ (Fig. 6). To overcome the fact that the direction of the parallel beams wrt. the camera is not known, we exploit the knowledge that the lasers are equidistant to the camera and approximate the direction with the direction of the vector connecting the camera center and the midpoint between the two points of laser beam intersections with the model X_{L_1} and X_{L_2}. Since it is reasonable to expect the camera-scene distance to be significantly greater than its difference measured at the two points, this approximation leads to a negligible error. Similarly to the FUM, the location of laser beam intersections with the scene X_{L_1} and X_{L_2} are determined through a ray casting procedure and are affected by the same scale as the model and therefore

affect the final estimated distance $\widehat{d}_L$ by the same factor:

$$\cos\alpha = \frac{v_{1,2} \cdot v_{CM}}{|v_{1,2}||v_{CM}|}, \tag{4}$$

$$\widehat{d}_L = \sin\alpha \cdot |v_{1,2}|, \tag{5}$$

where $v_{1,2}$ represents the vector between scene points X_{L_1} and X_{L_2} and v_{CM} the vector connecting camera center with the middle point X_M.

2.4. Camera – laser calibration

Depending on the method used for estimating the scale, several parameters describing the geometric arrangement between the camera and the laser scaler need to be estimated. To obtain a fully calibrated system, both origins and orientations of the laser beams with respect to the optical camera have to be known. Normally, the calibration procedure consists of the acquisition of images with clearly visible laser-surface intersections for which the distances between the camera and surface are known or can be easily computed (e.g., using a checkerboard pattern). A set of points lying on the laser beam is thus obtained by expressing the 3D positions of these intersections in the camera coordinate system. Given that the spread of distances at which the data is collected is sufficient, the direction of the laser beam can be confidently estimated through a line fitting procedure minimizing the sum of squared perpendicular distances between the 3D points and the laser beam line. To avoid any potentially erroneous 3D points affecting the final calibration, a robust estimation method such as RANSAC can be utilized. Finally, the origin of the laser is determined by computing the intersection between the now estimated direction of the laser beam and plane $\mathcal{L}$. It is important to note that, in underwater scenarios, significant refraction can occur at the air-acrylic-water interface of the laser housing. This effect has to be considered in the calibration procedure, either mathematically or by performing the data acquisition underwater.

Alternatively, the partially constrained method requires the knowledge of the distance between the parallel pair of lasers and that the camera center is equidistant to the laser origins (without the need for the knowledge of its value). For most cases, commercially available laser scalers have laser beams that are adequately parallel. On the contrary, ensuring that the camera center is equidistant to the beams is more challenging, unless one is using a purposely designed mounting bracket for the camera/laser system. Having this in mind, it is important to be aware of a potential error in the accuracy of the scale estimation caused by not having the camera equidistant to the laser origins.

The direct consequence of such error is the inaccuracy induced in the estimation of the vector v_{CM} used for approximating the direction of the parallel lasers. Given that the magnitude of the misalignment error will always be disproportionately small compared to the camera-scene distance (mm/cm vs. m), the effect on the estimated direction and subsequently on the final scale estimation will be negligible small as seen in Fig. 7. The analysis of the induced error due to the camera center misalignment up to 5 cm showed that the error increases with
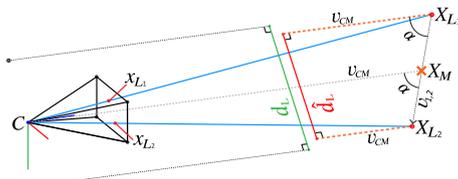

Fig. 6. Scale estimation procedure using the partially constrained method, based on the 3D model and optical image depicting the laser beam projections of the laser intersections with the scene. The direction of the parallel laser beams (wrt. the camera) is approximated with the vector v_{CM} and used to compute the perpendicular distance between the predicted laser beams originating at the point of laser intersensions with the scene (X_{L_1} and X_{L_2}).
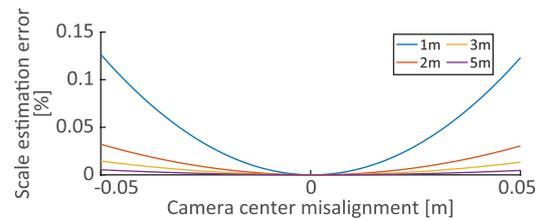

Fig. 7. Error in the estimation of scale using partially constrained method due to the misalignment of camera center with the laser pair.
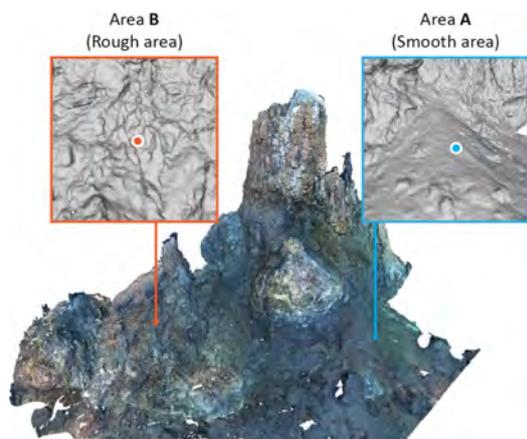
Fig. 8. 3D model of an underwater hydrothermal vent (Eiffel Tower at Lucky Strike vent field, Mid-Atlantic Ridge) used for model reconstruction evaluation at two marked areas. Data acquired during the 2015 MOMARSAT cruise (doi:10.17600/15000200).

decreased camera-scene distance as well as the error is small (less than 0.15% in the worst case). To ensure the error is in fact due to the mis-alignment, the camera was positioned parallel to a flat surface.

3. Results

To assess the applicability and theoretical accuracy of the two proposed approaches (PCM and FUM), tests were performed on both real and simulated scenario datasets. To validate the performance using different laser configurations and acquisition conditions, we have used a real 3D model built using underwater imagery, as depicted in Fig. 8. Various laser measurements were generated as they would have been captured during an ROV survey. As the absolute scale of the model is not precisely known, for the purpose of this evaluation, it was assumed that the model and its scale are correct. Therefore, the performance can be evaluated by comparing the deviations of the estimated scales with the assumed (imposed) correct value of the scale of the model ($s = 1$). This allowed us to confirm the correctness of our approaches, as well as analyze the effects of various types and levels of noise have on the estimation.

Given our goal of developing methods usable in real world scenarios, three distinct laser configurations were devised (Fig. 9)) to test the performance:

(A) Lasers are parallel and aligned with the optical axis of the camera;
(B) Lasers are parallel and positioned equidistant from the camera center, but not aligned with the optical axis;
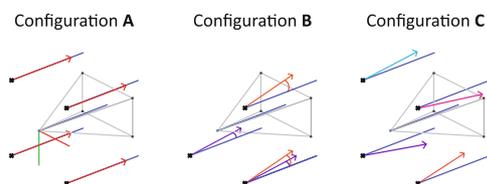(C) Lasers have arbitrary positions and directions.

Fig. 9. Various laser configurations used in evaluation: (*A*) Optical axis aligned laser beams; (*B*) Pair-wise parallel laser pairs; (*C*) Lasers with arbitrary origins and orientations. Blue lines represent the optical axis, and the remaining lines depict lasers which are parallel among themselves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To illustrate the advantages of our proposed methods in comparison to commonly used image-scaling approaches, the approach by Davis and Tusting (1991) was additionally evaluated, as one of the most versatile methods. The procedure requires four parallel lasers aligned with the optical axis of the camera as well as assuming scene flatness. By exploiting the known spacing between the laser spots on the image and the known displacement of laser origins from the optical center of the camera, distances between various points on the image can be computed for an arbitrary tilt and pan of the camera. As only laser configuration *A* meets the requirements of their method, and other configurations cause dramatic and unpredictable errors, we limit the reporting of the results for Davis and Tusting's approach to laser configuration *A*. Another commonly used method presented by Pilgrim et al. (2000) was not evaluated, as this method requires the restriction of the pose of the camera in either pan or tilt with respect to the scene, which can only be a reasonable restriction if the scene is flat (e.g., the sea bottom), which is almost never the case in models reconstructed using SfM.

3.1. Data

The generation of image and laser data as they would have been recorded in real scenarios enabled us to simulate different perspective angles and camera-scene distances, and analyze their effects on the resulting estimations of scales. The measuring system consisted of an ideal pinhole camera and laser scalers. Both the intrinsic parameters of the camera and the refraction of the lasers occurring at the air-housing-water interface were correctly modelled in the calibration. A real 3D model depicted in Fig. 8 was used in this simulation. The 3D chimney was reconstructed from 908 images of an underwater vent field at the deep-sea Lucky Strike area, collected during the MOMARSAT 2015 cruise (doi:10.17600/15000200). The model covers an area of approximately 200 m² with a height range of ~13 m. Assuming the 3D model has a correct scale, we can compute the location of laser spots and feature points as they would appear on the images taken from different poses and according to the pre-determined laser configurations. The number of feature points has been selected to reflect an average number of successfully matched features per image in underwater scenarios ($n = 1500$). To mimic the various perspective angles of the camera, we generate views for which the image plane is not only perpendicular to the surface normal (at the point viewed by the principal point of the camera), but also at a wide range of angles. In total 289 different views were created from different combinations of pitch and roll angles deviating from between $-40°$ and $40°$ in $5°$ steps (Fig. 10). If not specified differently, the camera-scene distance (i.e., distance between the camera center and the point of interest on the surface) has been kept constant at 3 m; based on our experience, this is a reasonable assumption for a typical ROV survey of the scene in this type of environments.

The lasers have been positioned according to the configurations envisioned in different scenarios (Fig. 9). In configuration *A*, the lasers
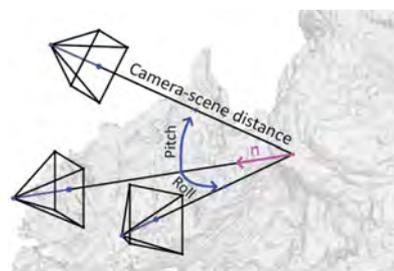

Fig. 10. Definition of perspective angles and camera-scene distance used in the generation of the evaluation data.
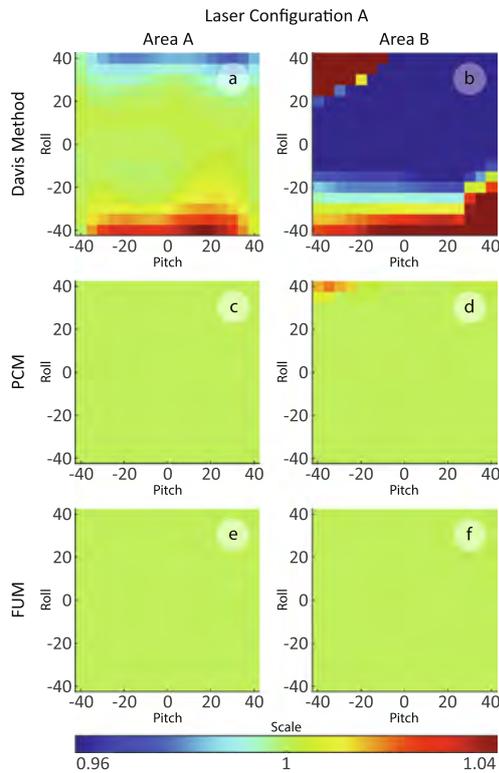
Fig. 11. Estimated scales of the model at a smooth (area A) and rough area (area B) with various perspective angles and constant camera-scene distance ($d = 3$ m) using Davis and Tusting (1991), partially constrained (PCM) and fully unconstrained method (FUM). Lasers were aligned with the optical axis (configuration A).

have been positioned at an equidistance of 10 cm from the camera center. For configuration B two pairs of lasers, with a 10 cm perpendicular distance between the beams, have been used, positioned vertically and horizontally. The pairs are perfectly parallel but not aligned with the optical axis of the camera. Each of the pairs has been used independently to test the two most common scenarios, with laser scalers positioned either below or at the side of the camera. As both produced similar results we only present the results for the horizontal pair.

Finally, the configuration C reflects a real laser configuration used during the 2017 SUBSAINTES cruise (doi:10. 17600/17001000) (Escartín et al., 2017). The laser set-up in the ROV VICTOR (IFREMER) used for image acquisition during this cruise was slightly misaligned, while the laser origins are placed at an approximately equal distance of 16.5 cm with slight rotation around the z-axis of the camera.

3.2. Terrain roughness

We first compare the results of estimated scales on two different types of terrain (smooth – *Area A* and rough – *Area B*) acquired from a variety of perspective angles and laser configurations. Fig. 11 presents the results obtained using laser configuration *A* and with our two proposed methods (FUM and PCM) as well as with the Davis approach.

Comparing the errors among the methods, we notice that the Davis and Tusting method is capable of estimating the correct scale only if the flatness assumption is only slightly violated, i.e. the area is nearly flat and the perspective angle is not too large (Fig. 11a). As that is not the case on rough terrain (Fig. 11b), the estimated scale varies significantly with different perspective angles, confirming the strong dependency of

this method on scene geometry. On the other hand, our two methods correctly compensate for any changes in the viewing angle and terrain roughness. The laser direction approximation assumed in PCM does, however, cause a slight error – up to 1.5% in extreme cases (e.g., rough terrain and large perspective angle – Fig. 11d), a situation in which the camera-scene distance discrepancy between the two laser points is strongly boosted. The fact that the scale is correctly estimated in all cases, clearly shows the ability of the FUM to correctly compensate for the effects of terrain roughness and perspective angle (Figs. 11e and 11f). Additionally, it is important to emphasize again that image-scaling methods require an additional association between the image points and the model in order to be able to estimate the scale. In our tests, we assumed perfect association, which is nearly impossible to achieve as it is a manual error-prone process. The actual results in real cases are therefore expected to be even worse.

In scenarios in which the lasers are not perfectly aligned with the camera (i.e., laser configurations *B* and *C*), the image-scaling methods become unusable as the errors increase dramatically and unpredictably. For this reason, we only present the results of our proposed methods (FUM and PCM) for the remaining two configurations. Similarly, we limit the presented results to the rough terrain, as the methods will perform better (or equally) on flat areas.

As seen in Fig. 12a and c, both of our methods obtain good results with a laser configuration *B*, in which the lasers are mounted parallel to each other. As in the previous cases, the partial method exhibits slight errors due to the assumed laser direction approximation. Analysis of data collected using laser configuration *C*, shows that the partial method fails, with results strongly affected by the irregularities in the parallelism. Instead, the full method (Fig. 12d) correctly compensates these irregularities and yields correct results.

3.3. Laser direction approximation

To illustrate the influence that a difference of camera-scene distances (measured at the two points hit by the laser beams) has on the results of the partial method at various distances, we estimated the scale using $10,000$ randomly-selected points across the model (Fig. 13a). For each point, the camera has been positioned at a distance
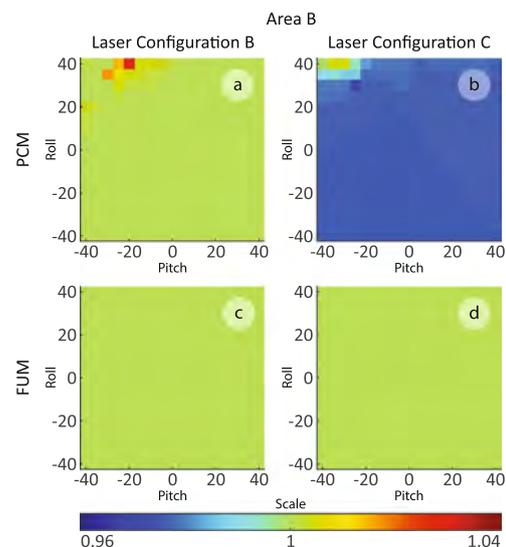

Fig. 12. Estimated scales of the model at a rough area (area B) with various perspective angles and constant camera-scene distance ($d = 3$ m) using partially constrained (PCM) and fully unconstrained method (FUM). Lasers were in configuration B and C.
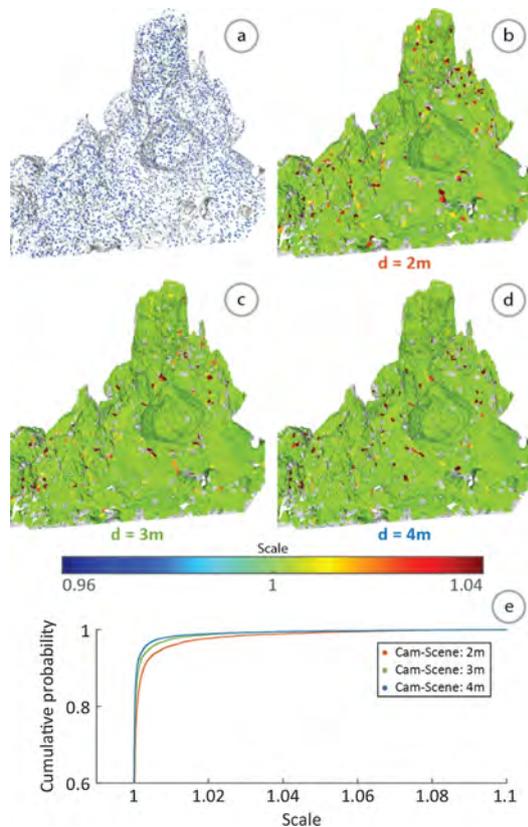
Fig. 13. (a) 10,000 random points used for estimating the scale across the model; (b–d) Estimated model scales at various camera-scene distances with laser configuration B using a partially constrained method; (e) Cummulative probability distribution of estimated scales.

Fig. 14. Estimated model scale (using partially constrained method at 10,000 random points) with respect to the difference of the camera-scene distances measured at the two points of laser beam intersection with the model color coded by the average camera-scene distance per image. Narrower parabolas at shorter distances indicate that the accuracy of the partially constrained method is more affected by the difference of the camera-scene distances at the two laser points than at longer distances.

d in the direction of the normal of the surface. Results obtained at three distances (2 m, 3 m and 4 m), illustrated in Fig. 13b–d, show that the error decreases with increasing distance of the camera (i.e., larger d). This is especially visible in rougher areas, such as the top of the hydrothermal vent and the areas near previously mentioned area B. In those areas, larger discrepancies between the camera-scene distances measured at the detected laser spots are expected, as the probability of laser beams hitting different parts of the model is much higher. Therefore the result indicates that the increased camera-scene distance decreases the effect of difference of distances on the accuracy of the results. We also document the cumulative distribution functions of these estimated scales obtained at various camera-scene distances (Fig. 13e), from which it is noticeable that a higher percentage of points with scales closer to the anticipated value of 1.0 is obtained the further the camera is from the scene.

The relation between the average camera-scene distance and its difference measured at the two points of laser beam-scene intersections can be clearly observed in Fig. 14, which shows the estimated scale vs. the difference of distances, with color coded average camera-scene distance. As expected, the error in the estimation grows with the increase in the distance discrepancies between the points. Furthermore, we can see that the increase follows a parabola-shaped functions determined by the camera-scene distance. Short distances define a narrow parabola, and cause an increase in the error that is larger than that for longer distances. This indicates that a distance discrepancy between the two points of laser – scene intersections causes a greater error in the result when the camera is near the scene. The shape and steepness of
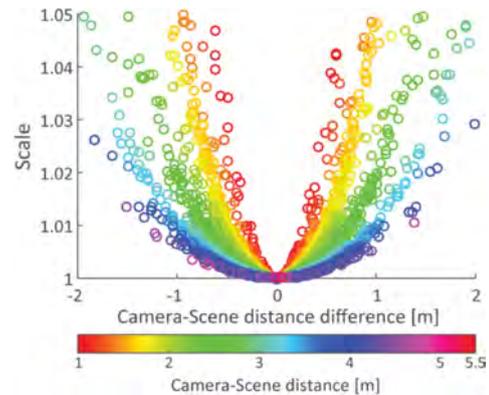
the parabolas is dependent on the displacement of the lasers from the camera origin, as well as their orientation with respect to the optical axis of the camera.

3.4. Noise

As collected data is never noise-free, we performed an additional analysis to evaluate the effects that the expected noise in feature and laser spot detection have on the scale estimation process. The various values for noise assumed in laser spot detection intend to represent the dispersion and absorption of the laser beam in the water medium. The greater the absorption and/or dispersion, the less certain the detection of the laser positions will be. The experiment was performed on *area B* of the model, with camera angles ranging from $-15°$ to $15°$ in pitch and roll; the range of view geometries which give consistent results in the ideal scenario (Fig. 11). The observation distributions were modelled by assuming multivariate Gaussian distributions with dimension-independent noise for both feature and laser spot detections. For 2D features, the values were set matching those normally obtained in underwater scenarios (Garcia and Gracias, 2011) $\sigma_f = \{0.5px, 1.0px\}$, while laser detection noise was defined by assuming 95% accuracy of peak detection within one or two pixels $\sigma_l = \{0.25px, 0.5px\}$. As feature matches themselves are normally corrupted with a certain level of outliers, we have also performed experiments with various inlier/outlier ratios (Campos et al., 2015) $r = \{0\%, 10\%, 20\%\}$. Each of the tests were repeated 500 times.

The resulting distributions of estimated scales with parallel and free laser configurations (i.e., configurations B and C) are presented in Table 1 with a subset of the results shown in Fig. 15. Given that the FUM requires only a single laser to obtain a scale estimate, results from separate lasers were fused by computing their average. The effect of such averaging can be identified in Table 1, where the results for a single laser (FUM – single) are shown side by side with the final averaged result (FUM – all).

As expected, the uncertainty of estimated scales increases with the increasing noisiness of the laser detections, as each estimation is directly influenced by displacements in laser spot positions. Comparison of these results show that with noisy data the PCM method performs better than the FUM with a single laser point, but worse when multiple laser points are used instead. This occurs due to the averaging of independent scale estimates. As each laser produces a result that is

Table 1
The results obtained with various methods (PCM, FUM) with different levels of noise induced into the location of detected features and laser spots.

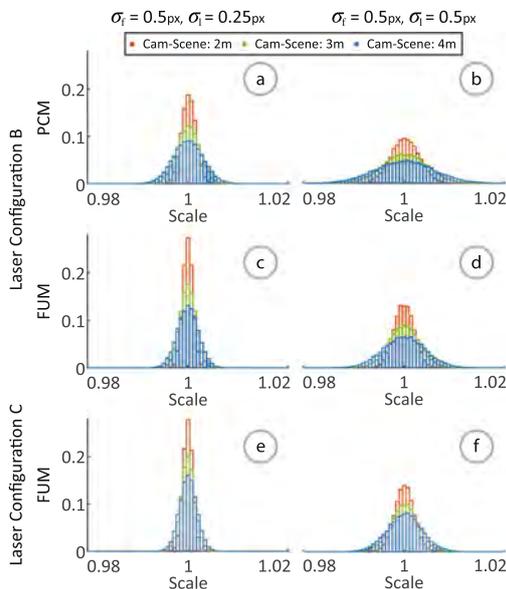| Cam-Scene	Configuration B		Configuration C	
distance [m]	PCM	FUM – all	FUM – single	FUM – all
$\sigma_f = 0.5$, $\sigma_l = 0.25$				
2	1.0 ± 0.0014	1.0 ± 0.0010	1.0 ± 0.0019	1.0 ± 0.0010
3	1.0 ± 0.0022	1.0 ± 0.0015	1.0 ± 0.0028	1.0 ± 0.0014
4	1.0 ± 0.0030	1.0 ± 0.0021	1.0 ± 0.0034	1.0 ± 0.0017
$\sigma_f = 1.0$, $\sigma_l = 0.25$				
2	1.0 ± 0.0014	1.0 ± 0.0010	1.0 ± 0.0019	1.0 ± 0.0010
3	1.0 ± 0.0022	1.0 ± 0.0015	1.0 ± 0.0028	1.0 ± 0.0014
4	1.0 ± 0.0030	1.0 ± 0.0021	1.0 ± 0.0034	1.0 ± 0.0017
$\sigma_f = 0.5$, $\sigma_l = 0.5$				
2	1.0 ± 0.0028	1.0 ± 0.0020	1.0 ± 0.0038	1.0 ± 0.0020
3	1.0 ± 0.0044	1.0 ± 0.0031	1.0 ± 0.0056	1.0 ± 0.0028
4	1.0 ± 0.0059	1.0 ± 0.0042	1.0 ± 0.0069	1.0 ± 0.0034

Fig. 15. Distributions of estimated model scales with partially constrained and fully unconstrained methods at various noise levels induced into the location of detected features and laser spots. The results obtained at different camera-scene distances are depicted with (2 m – red; 3 m – green; 4 m – blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

independently affected by noise, the subsequent averaging reduces its effect.

To some extent this can also be observed in the partial method with the simultaneous use of two laser points, which explains the improved results over the full method with the single laser. It is also clear that the uncertainty of the scaling estimate also increases with the camera-scene distance, which is expected as errors on the image are magnified when projected further from the camera.

In contrast, the noise affecting the feature points used in the pose estimation, does not significantly influence the final scaling results. This is due to the use of BA in the pose optimization, which is a maximum likelihood estimator when the image error is zero-mean and normally distributed, as is the case in our tests. Similarly, the effects of outliers are mitigated by the use of a robust estimation method AC-RANSAC

Fig. 16. ROV VICTOR 6000 (IFREMER) with enlarged camera and laser scaler system.

(Moisan et al., 2012). As the outliers do not follow a specific pattern, the iterative procedure successfully identifies and removes spurious matches, and hence the final estimate is unaffected. It is important to note that while the results obtained might indicate a method which is extremely robust to any discrepancy in the feature points, the approach is still vulnerable to (a) outliers that obey the estimated geometric model; to (b) the possibility of having a set of feature points which can be explained with multiple camera poses, or to both (a) and (b). However, this vulnerability can be reduced to a level that is not of practical concern, by ensuring that the set of features is well spread throughout the image.

3.5. Real scenario

The fully unconstrained method was used on a real dataset collected during the SUBSAINTES cruise (doi: 10.17600/ 17001000). Throughout the cruise, extensive seafloor imagery was collected using the ROV VICTOR 6000 (IFREMER) (Michel et al., 2003) with a mounted monocular camera (Sony FCB-H11 with corrective optics and dome port), and a laser scaler with four laser beams positioned around the camera (Fig. 16). The intrinsic parameters of the camera were determined using a standard calibration procedure (Bouguet, 2008) assuming a pinhole model together with the 3rd degree radial distortion model. Once calibrated, the camera parameters were kept constant through the entire acquisition process.

One of the main goals of this cruise is to identify, map, and measure indicators of displacement at the seafloor associated with a recent submarine earthquake (Escartín et al., 2016) that occurred in the French Antilles, offshore from Les Saintes Islands in 2004 (Feuillet et al., 2004). These traces are visible in outcrops of an active submarine fault scarp at depths of up to ~1000 m below sea level, that has been systematically mapped and surveyed. Imagery was used to obtain ~30 3D models, that will ultimately be used to conduct measurements of displacement associated with the 2004 earthquake. Therefore proper scaling is required to enable accurate geological measurements.

The 3D models have been reconstructed using an adapted 3D reconstruction procedure consisting of multiple open-source solutions (OpenMVG (Moulon et al., 2018; Moulon et al., 2013), OpenMVS (Shen, 2013; Jancosek and Pajdla, 2014), MVS-Texturing (Waechter et al., 2014)) as described in (Hernández et al., 2016). Fig. 17 depicts one such model, named FPA, which has been reconstructed from a total of 218 images with a resolution of 1920 × 1080. This particular outcrop was already imaged during a prior cruise (ODEMAR, doi:10.17600/ 13030070) (Escartín et al., 2016).

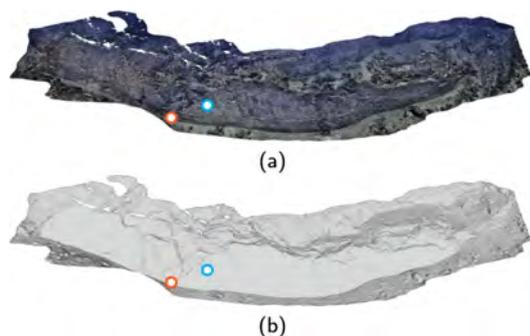As the FPA model was reconstructed using only optical images

Fig. 17. (a) Textured and (b) triangle mesh representations of FPA 3D model, with marked areas of evaluation.
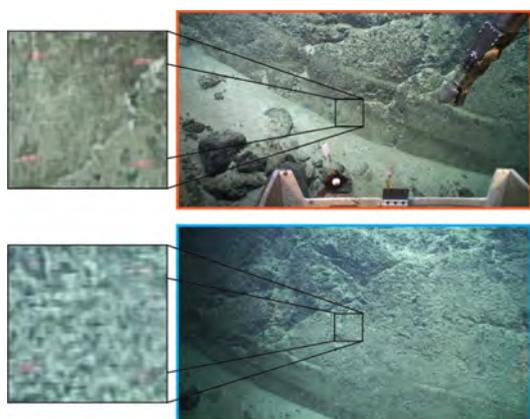

Fig. 18. Example of images from the two areas of evaluation with visible laser projections on the scene.

acquired by a monocular camera, the proper scale of the model can be obtained using images containing laser beams projected on the surface of the scene by applying one of our proposed methods. During the SUBSAINTES cruise, such images were collected in addition to the ones already used in the reconstruction process. Six images with clearly noticeable laser spots (Fig. 18) have been selected from the center of the 3D model, at two different locations as indicated in Fig. 18. The images were collected at camera-scene distances of approximately 3 m and 4 m respectively while keeping the camera intrinsic parameters constant and equal to the ones used in the acquisition process. Subsequently, the laser spot's locations have been marked manually (with the guidance of simple color thresholding) with an expected error that was on average between 1 px and 2 px. Due to multiple changes in the vehicle payload throughout the cruise, the lasers became misaligned and therefore a fully unconstrained method was used to obtain the scale of the model.

Given that the setup consisted of four lasers, the FUM method computed four independent estimates of the model's scale per image. As we have shown in the previous experiments, averaging these independent results further reduces the effects of errors in the detection processes, leading to a better constrained final solution. The scaling results for each of the 6 selected images are presented in Table 2 and Fig. 19. In this figure, the scale estimates obtained for each laser beam are depicted as circles, while the final estimate per image is marked with a black cross (x). The average of all the values obtained is additionally shown by a red dashed line.

The average value of the scale of the FPA model estimated per image was 0.237 ± 0.0008 which represents 0.3% of the scale value. The obtained result means that each unit in the current model is equal to

Table 2

Estimated FPA model's scale using fully unconstrained method and a simplistic direct 3D approach. Reported numbers represent the ratio between the model's unit and a meter – each measurement has to be multiplied with the inverse of the ratio to obtain a metric result.

	Cam-Scene distance [m]	FUM (per laser)				FUM (all)	Direct 3D (all)
		L_1	L_2	L_3	L_4		
1	3.05	0.234	0.239	0.237	0.236	0.237 ± 0.002	0.235 ± 0.009
2	3.06	0.236	0.239	0.236	0.238	0.237 ± 0.002	0.236 ± 0.008
3	3.05	0.237	0.237	0.235	0.236	0.236 ± 0.001	0.235 ± 0.008
4	3.90	0.239	0.241	0.236	0.236	0.238 ± 0.003	0.236 ± 0.013
5	3.91	0.238	0.239	0.237	0.234	0.237 ± 0.002	0.236 ± 0.013
6	3.60	0.238	0.236	0.236	0.233	0.236 ± 0.002	0.234 ± 0.010
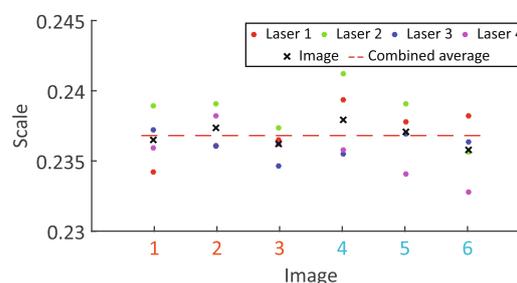
Fig. 19. Estimated scale factors for FPA model, per laser and per image, using our fully unconstrained method. Colour of image numbers (x axis) corresponds to locations shown in Fig. 18. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.237 m or alternatively, the model has to be scaled by a factor 4.22 to obtain a metric result. This implies that on a measurement of 25 cm on a model with lateral and vertical dimensions of 33 m by 10 m is expected to have an uncertainty, due to scaling error introduced by the proposed method, of approximately 1 mm. It is important to note that this estimate is based on the assumption of adequate calibration of the camera and of the camera-laser system. Given that these methods estimate the scale from image information, it is impossible to decouple the error reported by the method with the actual error of the model without the validation with the external measurements. A sign of such problems would be disproportionately large variations of scale estimates from individual lasers within the same image, as each result would be affected differently by the erroneous calibration. Comparing the deviations of scale estimates for image sets 1-3 and 4-6, the correlation between increasing camera-scene distance and increased uncertainty is apparent and consistent with previous result from generated data.

The analyses of scaling deviations computed for each laser with respect to the final estimated scale per image (Fig. 20) shows that independent evaluations deviate by about 0.6% with a maximum
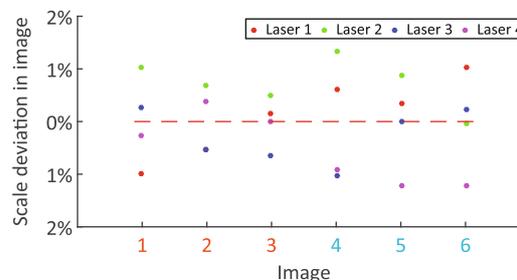
Fig. 20. Deviation of estimated FPA model's scales, using our fully unconstrained method FUM, and for each laser in each image.

deviation of 1.3% for laser 2 in image 4. These results are again in agreement with the results previously computed with the validation data on the hydrothermal vent in Fig. 8.

To further show the robustness and usefulness of our approach, we compare our results to the ones that would have been obtained if our method was not available. As the non-alignment of lasers with the optical axis would have prevented the use of both image-scaling methods (Pilgrim et al. (2000) and Davis and Tusting (1991)), the only option available would have been a manual and somewhat simplistic approach still widely used in laser photogrammetry (Kocak et al., 2004; E. Rowe and Dawson, 2008; Robert et al., 2017; Pilgrim et al., 2000). This involves manual identification of laser beam intersection points with the scene on the 3D model, and assuming pair-wise Euclidean distances to be the actual distances between the laser pairs. In order to compare our results with the best possible outcome of this simplistic approach, we determined the points on the model using a ray-casting technique, effectively completely eliminating the extremely error-prone human step. The results averaged over 4 laser pairs are presented in the last column of Table 2 (Direct 3D). We can see that the results of different laser pairs are much more incoherent (4.3% deviation compared to 0.6% in the case of fully unconstrained method). We also note that the results of such a simplistic method are extremely dependent on the perspective angle of the camera, and the degree of misalignment of the lasers, as well as errors induced by manual point selection. As shown with the validation tests, our fully unconstrained method remains unaffected.

4. Conclusions

This paper introduced two novel methods for automatic scaling of SfM-based 3D reconstructions using laser scalers, methods that are applicable for routine underwater surveys with ROVs or AUVs. Both methods were validated using a series of generated datasets based on an underwater 3D model derived from submarine field imagery, and showed its applicability in real scenarios using a dataset collected during a recent cruise (SUBSAINTES 2017).

The two approaches presented here, namely the fully and partially constrained methods, overcome a multitude of restrictions imposed by prior laser photogrammetry methods (e.g., laser alignment with the optical axis of the camera, perpendicularity of laser beams with the scene). These methods, within the step of pose estimation, also remove the need for manual identification of identical points on the image and 3D model, an extremely time-consuming and error-prone processing step.

Each of the two methods is designed to address the different types of laser setup, encompassing the variety of most commonly used setups in real underwater scenarios. The fully unconstrained method is applicable to arbitrary laser setups, with known geometric relations between the camera and the lasers. The ability to compensate for any misalignments enables accurate scaling in a wider variety of circumstances, such as the manipulation of equipment between surveys during a cruise and precluding strict parallelism. We thus propose a partially constrained method, which significantly reduces the camera-laser rigidity constraints, that may be otherwise too restrictive in real scenarios. This approach requires parallel lasers but alleviates the need for a time-consuming calibration process. The partially constrained method can thus be used to accurately and automatically scale 3D models built with data acquired using ROVs, including smaller shallow-water ones. Pre-calibrated underwater laser scalers are readily available nowadays, and need only to be placed near the optical camera.

To robustly validate the performance of the methods, a real 3D model of an underwater hydrodynamic vent was used to generate laser and image information as it would have been obtained from various laser configurations, camera viewing angles and camera-scene distances. We tested our methods with three laser configurations (i.e., aligned with the optical axis of the camera; parallel but misaligned with

the optical axis; and freely oriented) which can account for nearly all possible laser setups in real seafloor surveying situations using ROVs and AUVs. The initial evaluation was performed on two different types of terrain (smooth and rough), and demonstrated the advantages provided by the two proposed approaches relative to previously used image-scaling methods. Our methods can be used in the field, with misaligned or freely oriented lasers, and with extreme camera angles during image acquisitions, reaching up to 40° in both pitch and roll.

While the fully unconstrained method yielded robust results under all tested circumstances, the partially constrained method was affected by a slight error (2.9% in the most extreme case) due to the approximation used for determining the laser direction. We further analyzed the effect of the approximation by evaluating 10, 000 randomly selected points. We demonstrate that scaling errors depend on the difference of camera-scene distances between the two points of laser beam – scene intersection, and that this effect decreases with an increasing camera-scene distance. The consequences of inevitable noise in feature and laser-spot detection uncertainty were also examined, together with the effects of potential errors in feature matching (outliers). Due to the specificity of the algorithms used, the noise and potential outliers in the feature detection and matching process did not have a significant effect on the results, while the noise induced in the position of laser spots did directly influence the estimations. As expected, increases in camera-scene distance results in higher errors in the estimation, as the displacements are magnified with distance. Additionally we compared the results obtained from a single laser measurement with the average obtained from all and demonstrated that such fusion further reduces the effects of noise.

It is important to acknowledge that the achieved accuracy in the simulated scenario should be regarded as theoretical accuracy, given that the camera and camera-laser calibrations used in generating the data were ideal. Errors in the calibration will affect the final result, in such a way that, without external validation (using known measurements) it becomes impossible to decouple the error reported by the method and the actual error of the model. We note that a sign of such problems would be large deviations of scale estimated obtained by independent lasers/laser pairs within the same image.

Finally we report on the application of the fully unconstrained method to determine the scale of a model built using images from a geologic outcrop, recorded during the SUBSAINTES cruise. Six images with clearly visible laser spots have been selected from two different model locations, and used to independently determine the scale of the model. The average scale estimated using our fully unconstrained method was 0.237 with the standard deviation of 0.3% between the results from various images. The average deviation of estimated scales by independent lasers was 0.6% with the maximum deviation of 1.3%. We also documented data showing that images acquired at a longer camera-scene distance exhibited bigger deviations of estimated scales, as predicted from the validation test results.

The results of our two methods were also compared to those that would have been obtained without the availability of our method. Due to laser non-alignment with the optical axis of the camera, the only approach possible would be a somewhat simplistic method which involves manual identification of laser intersection points with the 3D model, and assumes that the pair-wise Euclidean distances are the actual distances between the laser pairs. To predict the best possible outcome, we automatically determined these correspondences, alleviating any additionally induced errors. The results from the simplistic scale method show a much more significant deviation than that of our method (4.3% vs. 0.6%, respectively). Based on our results we also stress that the results of such simplistic methods are extremely dependent on the perspective angle of the camera and the degree of misalignment of the lasers, which is not the case for our fully unconstrained method. Finally, these methods can be used universally as they are based on standard sensors available for ROVs and AUVs (cameras and laser scalers), do not require any dedicated hardware, and can be applied to

legacy data.

Although the presented methods are designed to be independent of the laser spot detection approach used, we showed that its performance directly influences the scale estimation accuracy. In the reported results, we identified the location of the spots manually albeit with the help of simple color thresholding. While relatively accurate, this manual process is time consuming. An effort is currently ongoing aimed at automating the detection of the laser spots, which will facilitate the ability to perform scale estimation on a larger number of images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

References

Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R., 2009. Building rome in a day. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 72–79. https://doi.org/10.1109/ICCV.2009.5459148.

Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Front. Ecol. Environ. 11, 138–146. https://doi.org/10.1890/120150.

Bergmann, M., Langwald, N., Ontrup, J., Soltwedel, T., Schewe, I., Klages, M., Nattkemper, T.W., 2011. Megafaunal assemblages from two shelf stations west of svalbard. Mar. Biol. Res. 7, 525–539. https://doi.org/10.1080/17451000.2010.535834.

Bingham, B., Foley, B., Singh, H., Camilli, R., Delaporta, K., Eustice, R., Mallios, A., Mindell, D., Roman, C., Sakellariou, D., 2010. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. J. Field Robot. 27, 702–717. https://doi.org/10.1002/rob.20350.

Bodenmann, A., Thornton, B., Ura, T., 2017. Generation of high-resolution three-dimensional reconstructions of the seafloor in color using a single camera and structured light. J. Field Robot. 34, 833–851. https://doi.org/10.1002/rob.21682.

Bouguet, J.Y., 2008. Camera calibration toolbox for matlab (2008). http://www.vision.caltech.edu/bouguetj/calib_doc1080. Accessed: January 22 2018.

Caimi, F.M., Tusting, R.F., 1987. Application of lasers to ocean research and image recording systems. In: Proceedings of the International Conference on LASERS, STS Press McLean, Virginia. pp. 518–524.

Campos, R., Garcia, R., Alliez, P., Yvinec, M., 2015. A surface reconstruction method for in-detail underwater 3d optical mapping. Int. J. Robot. Res. 34, 64–89.

Campos, R., Gracias, N., Ridao, P., 2016. Underwater multi-vehicle trajectory alignment and mapping using acoustic and optical constraints. Sensors 16, 387. https://doi.org/10.3390/s16030387.

Cocito, S., Sgorbini, S., Peirano, A., Valle, M., 2003. 3-d reconstruction of biological objects using underwater video technique and image processing. J. Exp. Mar. Biol. Ecol. 297, 57–70. https://doi.org/10.1016/S0022-0981(03)00369-1.

Comex.SA, Orus 3d subsea photogrammetry. https://comex.fr/en/orus3d. (accessed: September 6 2019).

Davis, D., Tusting, R., 1991. Quantitative Benthic Photography using Laser Calibrations. Undersea World, San Diego, California.

E. Rowe, L., Dawson, S., 2008. Laser photogrammetry to determine dorsal fin size in a population of bottlenose dolphins from doubtful sound, New Zealand. Aust. J. Zool. 56, 239–248. https://doi.org/10.1071/ZO08051.

Eltner, A., Schneider, D., 2015. Analysis of different methods for 3d reconstruction of natural surfaces from parallel-axes uav images. Photogram. Rec. 30, 279–299. https://doi.org/10.1111/phor.12115.

Escartín, J., Le Friant, A., Feuillet, N., 2017. Subsaintes cruise report, n/o l'atalante - rov victor - auv asterx. https://campagnes.flotteoceanographique.fr/campagnes/17001000/, doi:https://doi.org/10.17600/17001000.

Escartín, J., Leclerc, F., Olive, J.A., Mevel, C., Cannat, M., Petersen, S., Augustin, N., Feuillet, N., Deplus, C., Bezos, A., et al., 2016. First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history. Earth Planet. Sci. Lett. 450, 96–107. https://doi.org/10.1016/j.epsl.2016.06.024.

Feuillet, N., Beauducel, F., Jacques, E., Tapponnier, P., Delouis, B., Bazin, S., Vallée, M., King, G., 2004. The mw= 6.3, november 21, 2004, les saintes earthquake (guadeloupe): Tectonic setting, slip model and static stress changes. J. Geophys. Res.: Solid Earth 116. https://doi.org/10.1029/2011JB008310.

Garcia, R., Campos, R., Escartín, J., 2011. High-resolution 3d reconstruction of the seafloor for environmental monitoring and modelling. In: 2011 IEEE/RSJ International Conference on Proc. Intelligent Robots and Systems (IROS).

Garcia, R., Gracias, N., 2011. Detection of interest points in turbid underwater images. In: OCEANS 2011 IEEE-Spain, IEEE. pp. 1–9.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press, New York, NY, USA. https://doi.org/10.1017/CBO9780511811685.001.

Hernández, J.D., Istenič̌ K., Gracias, N., Palomeras, N., Campos, R., Vidal, E., Garcia, R., Carreras, M., 2016. Autonomous underwater navigation and optical mapping in unknown natural environments. Sensors 16, 1174. https://doi.org/10.3390/s16081174.

Jancosek, M., Pajdla, T., 2014. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. Int. Scholarly Res. Notices 2014. https://doi.org/10.1155/2014/798595.

Javernick, L., Brasington, J., Caruso, B., 2014. Modeling the topography of shallow braided rivers using structure-from-motion photogrammetry. Geomorphology 213, 166–182. https://doi.org/10.1016/j.geomorph.2014.01.006.

Kalacska, M., Lucanus, O., Sousa, L., Vieira, T., Arroyo-Mora, J., 2018. Freshwater fish habitat complexity mapping using above and underwater structure-from-motion photogrammetry. Remote Sensing 10, 1912. https://doi.org/10.3390/rs10121912.

Ke, T., Roumeliotis, S.I., 2017. An efficient algebraic solution to the perspective-three-point problem. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7225–7233.

Kocak, D.M., Caimi, F.M., Jagielo, T.H., Kloske, J., 2002. Laser projection photogrammetry and video system for quantification and mensuration. In: OCEANS '02 MTS/IEEE, pp. 1569–1574 vol 3. doi:https://doi.org/10.1109/OCEANS.2002.1191869.

Kocak, D.M., Jagielo, T.H., Wallace, F., Kloske, J., 2004. Remote sensing using laser projection photogrammetry for underwater surveys. In: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, vol 2., pp. 1451–1454 doi:https://doi.org/10.1109/IGARSS.2004.1368693.

Lourakis, M., Zabulis, X., 2013. Accurate scale factor estimation in 3d reconstruction. In: International Conference on Computer Analysis of Images and Patterns, Springer. pp. 498–506.

Mertes, J., Zant, C., Gulley, J., Thomsen, T., 2017. Rapid, quantitative assessment of submerged cultural resource degradation using repeat video surveys and structure from motion. J. Maritime Archaeol. 12, 91–107. https://doi.org/10.1007/s11457-017-9172-0.

Mian, O., Lutes, J., Lipa, G., Hutton, J., Gavelle, E., Borghini, S., 2016. Accuracy assessment of direct georeferencing for photogrammetric applications on small unmanned aerial platforms. Int. Arch. Photogramm. Remote Sensing Spatial Inf. Sci. 40, 77. https://doi.org/10.5194/isprs-archives-XL-3-W4-77-2016.

Michel, J.L., Klages, M., Barriga, F.J., Fouquet, Y., Sibuet, M., Sarradin, P.M., Siméoni, P., Drogou, J.F., et al., 2003. Victor 6000: design, utilization and first improvements. In: The Thirteenth International Offshore and Polar Engineering Conference, International Society of Offshore and Polar Engineers.

Moisan, L., Moulon, P., Monasse, P., 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. Image Process. On Line 2, 56–73. https://doi.org/10.5201/ipol.2012.mmm-oh.

Moulon, P., Monasse, P., Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3248–3255. https://doi.org/10.1109/ICCV.2013.403.

Moulon, P., Monasse, P., Marlet, R., Others, Openmvg. an open multiple view geometry library. https://github.com/openMVG/openMVG. Accessed: January 11 2018.

Neyer, F., Nocerino, E., Gruen, A., 2018. Monitoring coral growth-the dichotomy between underwater photogrammetry and geodetic control network. Int. Arch. Photogramm. Remote Sensing Spatial Inf. Sci. 42, 2. https://doi.org/10.5194/isprs-archives-XLII-2-759-2018.

Pilgrim, D.A., Parry, D.M., Jones, M.B., Kendall, M.A., 2000. Rov image scaling with laser spot patterns. Underwater Technol. 24, 93–103. https://doi.org/10.3723/175605400783259684.

Pizarro, O., Eustice, R.M., Singh, H., 2009. Large area 3-d reconstructions from underwater optical surveys. IEEE J. Oceanic Eng. 34, 150–169. https://doi.org/10.1109/JOE.2009.2016071.

Pizarro, O., Friedman, A., Bryson, M., Williams, S.B., Madin, J., 2017. A simple, fast, and repeatable survey method for underwater visual 3d benthic mapping and monitoring. Ecol. Evol. 7, 1770–1782. https://doi.org/10.1002/ece3.2701.

Remondino, F., El-Hakim, S.F., Gruen, A., Zhang, L., 2008. Turning images into 3-d models. IEEE Signal Process. Mag. 25, 55–65. https://doi.org/10.1109/MSP.2008.923093.

Ribas, D., Palomeras, N., Ridao, P., Carreras, M., Mallios, A., 2011. Girona 500 auv: From survey to intervention. IEEE/ASME Trans. Mechatronics 17, 46–53.

Robert, K., Huvenne, V.A., Georgiopoulou, A., Jones, D.O., Marsh, L., Carter, G.D., Chaumillon, L., 2017. New approaches to high-resolution mapping of marine vertical

structures. Sci. Rep. 7, 9005.

Rossi, P., Castagnetti, C., Capra, A., Brooks, A., Mancini, F., 2019. Detecting change in coral reef 3d structure using underwater photogrammetry: critical issues and performance metrics. Appl. Geomatics 1–15. https://doi.org/10.1007/s12518-019-00263-w.

Rovco, 2019. High definition 3d visualisation. https://www.rovco.com/technology/3d-photogrammetry. Accessed: September 6 2019.

Rzhanov, Y., Mamaenko, A., Yoklavich, M., 2005. Uvsd: software for detection of color underwater features. In: Proceedings of OCEANS 2005 MTS/IEEE, vol. 3, pp. 2189–2192. doi:https://doi.org/10.1109/OCEANS.2005.1640089.

Schoening, T., Kuhn, T., Bergmann, M., Nattkemper, T.W., 2015. Delphi—fast and adaptive computational laser point detection and visual footprint quantification for arbitrary underwater image collections. Front. Mar. Sci. 2, 20. https://doi.org/10.3389/fmars.2015.00020.

Sedlazeck, A., Koser, K., Koch, R., 2009. 3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions. In: OCEANS 2009-EUROPE, pp. 1–10. doi:https://doi.org/10.1109/OCEANSE.2009.5278305.

Shen, S., 2013. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. IEEE Trans. Image Process. 22, 1901–1914. https://doi.org/10.1109/TIP.2013.2237921.

Snavely, N., Seitz, S.M., Szeliski, R., 2008. Modeling the world from internet photo collections. Int. J. Comput. Vision 80, 189–210. https://doi.org/10.1007/s11263-007-0107-3.

Soloviev, A., Venable, D., 2010. Integration of gps and vision measurements for navigation in gps challenged environments. In: IEEE/ION Position, Location and Navigation Symposium, pp. 826–833. doi:https://doi.org/10.1109/PLANS.2010.5507322.

Spaenlehauer, A., Fremont, V., Sekercioglu, Y.A., Fantoni, I., 2017. A loosely-coupled approach for metric scale estimation in monocular vision-inertial systems. In: 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp. 137–143. https://doi.org/10.1109/MFI.2017.8170419.

Storlazzi, C.D., Dartnell, P., Hatcher, G.A., Gibbs, A.E., 2016. End of the chain? rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (sfm) technology. Coral Reefs 35, 889–894. https://doi.org/10.1007/s00338-016-1462-8.

Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 1999. Bundle adjustment – a modern synthesis. In: Vision Algorithms: Theory and Practice. Springer, Berlin, Germany, pp. 298–372. https://doi.org/10.1007/3-540-44480-7_21.

Tusting, R., Davis, D., 1986. Non-conventional techniques for sampling and collecting marine organisms. In: Proceedings of the Pacific Congress on Marine Technology, PACON'86, pp. 12–18.

Tusting, R.F., Davis, D., 1993. Improved methods for visual and photographic benthic surveys.

Tusting, R.F., Davis, D.L., 1992. Laser systems and structured illumination for quantitative undersea imaging. Mar. Technol. Soc. J. 26, 5–12.

Waechter, M., Moehrle, N., Goesele, M., 2014. Let there be color! large-scale texturing of 3d reconstructions, in: Computer Vision–ECCV. Springer, Berlin, Germany, pp. 836–850. doi:https://doi.org/10.1007/978-3-319-10602-1_54.

Wakefield, W.W., Genin, A., 1987. The use of a canadian (perspective) grid in deep-sea photography. Deep Sea Research Part A. Oceanogr. Res. Pap. 34, 469–478. https://doi.org/10.1016/0198-0149(87)90148-8.

Wallace, L., Lucieer, A., Malenovsky, Z., Turner, D., Vopenka, P., 2016. Assessment of forest structure using two uav techniques: a comparison of airborne laser scanning and structure from motion (sfm) point clouds. Forests 7. https://doi.org/10.3390/f7030062.

Zhang, J., Singh, S., 2015. Visual-inertial combined odometry system for aerial vehicles. J. Field Robot. 32, 1043–1055. https://doi.org/10.1002/rob.21599.

Klemen Istenič: received his Diploma degree in Computer Science from the University of Ljubljana, Slovenia in 2013 and a joint M.Sc. degree with distinction in Computer Vision and Robotics (European Master ViBOT) from Heriott-Watt University, UK, University of Girona, Spain, and the University of Burgundy, France in 2015. He is currently pursuing a Ph.D. in Technology at the University of Girona as a member of the Underwater Robotics Research Center (CIRS), part of the Computer Vision and Robotics Institute, as well as a member of an European Academy for Marine and Underwater Robotics (EU FP7 Marie Curie ITN network no 608096 – Robocademy). His research focuses on 3D mapping, color restoration and change detection using optical data in underwater scenarios.

Dr. Nuno Gracias: was awarded the Ph.D. degree in 2003 from the Technical University of Lisbon, Portugal. From 2004 to 2006 he was a post-doctorate fellow at the University of Miami. Since 2006 he has been a member of the Computer Vision and Robotics Group (ViCOROB) of the University of Girona. His research interests span the areas of underwater optical mapping, and navigation and guidance of autonomous underwater robots, image processing and classification. Dr. Gracias has authored more than 80 articles in peer-review journals and scientific conferences, and co-supervised 3 PhD and 8 MSc theses. He is adjunct faculty at the department of Marine Geosciences of the University of Miami, and member of the editorial board of the Journal of Intelligent and Robotic Systems.

Dr. Aurélien Arnaubec: received his Ph.D degree from the Universé Aix-Marseille III,Marseille, France, in 2012. He did his Ph.D. degree at the French Aerospace Laboratory, Office National d'Etudes et Recherches Aérospatiales, (ONERA), Salon Cedex Air, France, and in the Physics and Image Processing Group, Fresnel Institute, Marseille, France where his main research interests was radar imaging and statistical signal processing for remote sensing. Then from 2012 to 2014 he did a postdoc at Ifremer, la Seyne sur Mer, France, where he developed image processing techniques for optical mapping. He has now joined Ifremer PRAO team (Positionning, Robotic, Acoustic and Optics) where he is in charge of all underwater optical systems and image processing software.

Dr. Javier Escartín: received his PhD from the MIT/WHOI Joint Program in 1996 (USA). He is now a CNRS Senior Research Scientist at the Institute de Physique du Globe de Paris – Université de Paris, where since 2017 he leads the Marine Geosciences Group. He is also Associated CNRS Professor at the ENS of Paris. His research focuses on deep-sea exploration of the seafloor to understand geological processes, such as tectonism, volcanism, and hydrothermal activity. For this research he employs acoustic and optical mapping of the seafloor acquired with deep-sea vehicles.

Dr. Rafael Garcia: received a M.S. degree in computer engineering from the Universitat Autonoma de Barcelona in 1994, and a Ph.D. in Computer Engineering from the Universitat de Girona (Spain) in 2001. He is the founder and director of the Underwater Vision Lab (UVL), within the Computer Vision and Robotics Group in the Department of Computer Architecture at the University of Girona. His main research interests are underwater robotics and computer vision. He is particularly interested in understanding how to make underwater vehicles sense their environment in order to carry out autonomous surveys. He has published more than 190 technical contributions, including journal papers and conference proceedings. Dr. Garcia is a Member of the IEEE.

4

Scale Accuracy Evaluation of Image-Based 3D Reconstruction Strategies Using Laser Photogrammetry

F OLLOWING the work presented in the previous chapter, we incorporate the proposed scaling methods into a Monte Carlo (MC) simulation capable of assessing the confidence levels for each of the scale estimates. First, we present a comprehensive overview of the most commonly used 3D reconstruction and navigation fusion strategies for underwater applications. Then our automatic scale accuracy framework is utilized to perform a systematic analysis of the impact of reconstruction strategies on the creation of 3D models on a larger scale in GPS-denied environments with monocular camera images. Two distinct surveys from the *SUBSAINTES* scientific cruise are used. For the models, which were built with multiobjective BA navigation fusion strategy, an L-curve analysis is performed to determine the optimal weight between competing objectives of the optimization, while the potential offset in navigation between different dives is studied in a representative experiment. In order to facilitate the detection of even strongly attenuated laser beams, we also present a new automatic laser spot detection method, the main innovation of which is the estimation of uncertainties based on the recovery of characteristic shapes of laser spots.

Article

Scale Accuracy Evaluation of Image-Based 3D Reconstruction Strategies Using Laser Photogrammetry

Klemen Istenič [1,2,*] ⓘ, **Nuno Gracias** [1] ⓘ, **Aurélien Arnaubec** [3], **Javier Escartín** [4] ⓘ and **Rafael Garcia** [1] ⓘ

[1]   Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, Edifici P-IV, Campus de Montilivi, 17071 Girona, Spain; ngracias@silver.udg.edu (N.G.); rafael.garcia@udg.edu (R.G.)

[2]   Coronis Computing, S.L., Science and Technological Park of UdG, Carrer Pic de Peguera, 15, 17003 Girona, Spain

[3]   IFREMER, Centre Méditerranée, Unité Syst. Marins, CS 20330, F-83507 La Seyne Sur Mer, France; aurelien.arnaubec@ifremer.fr

[4]   Université de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005 Paris, France; escartin@ipgp.fr

*   Correspondence: klemen.istenic@gmail.com

check for updates

Abstract: Rapid developments in the field of underwater photogrammetry have given scientists the ability to produce accurate 3-dimensional (3D) models which are now increasingly used in the representation and study of local areas of interest. This paper addresses the lack of systematic analysis of 3D reconstruction and navigation fusion strategies, as well as associated error evaluation of models produced at larger scales in GPS-denied environments using a monocular camera (often in deep sea scenarios). Based on our prior work on automatic scale estimation of Structure from Motion (SfM)-based 3D models using laser scalers, an automatic scale accuracy framework is presented. The confidence level for each of the scale error estimates is independently assessed through the propagation of the uncertainties associated with image features and laser spot detections using a Monte Carlo simulation. The number of iterations used in the simulation was validated through the analysis of the final estimate behavior. To facilitate the detection and uncertainty estimation of even greatly attenuated laser beams, an automatic laser spot detection method was developed, with the main novelty of estimating the uncertainties based on the recovered characteristic shapes of laser spots with radially decreasing intensities. The effects of four different reconstruction strategies resulting from the combinations of Incremental/Global SfM, and the a priori and a posteriori use of navigation data were analyzed using two distinct survey scenarios captured during the SUBSAINTES 2017 cruise (doi: 10.17600/17001000). The study demonstrates that surveys with multiple overlaps of nonsequential images result in a nearly identical solution regardless of the strategy (SfM or navigation fusion), while surveys with weakly connected sequentially acquired images are prone to produce broad-scale deformation (doming effect) when navigation is not included in the optimization. Thus the scenarios with complex survey patterns substantially benefit from using multiobjective BA navigation fusion. The errors in models, produced by the most appropriate strategy, were estimated at around 1% in the central parts and always inferior to 5% on the extremities. The effects of combining data from multiple surveys were also evaluated. The introduction of additional vectors in the optimization of multisurvey problems successfully accounted for offset changes present in the underwater USBL-based navigation data, and thus minimize the effect of contradicting navigation priors. Our results also illustrate the importance of collecting a multitude of evaluation data at different locations and moments during the survey.

1. Introduction

Accurate and detailed 3D models of the environment are now an essential tool in different scientific and applied fields, such as geology, biology, engineering, archaeology, among others. With advancements in photographic equipment and improvements in image processing and computational capabilities of computers, optical cameras are now widely used due to their low cost, ease of use, and sufficient accuracy of the resulting models for their scientific exploitation. The application of traditional aerial and terrestrial photogrammetry has greatly expanded in recent years, with commercial and custom-build camera systems and software solutions enabling a nearly black-box type of data processing (e.g., the works by the authors of [1–4]).

These rapid developments have also significantly benefited the field of underwater photogrammetry. The ability to produce accurate 3D models from monocular cameras under unfavorable properties of the water medium (i.e., light attenuation and scattering, among other effects) [5], and advancements of unmanned underwater vehicles have given scientists unprecedented access to image the seafloor and its ecosystems from shallow waters to the deep ocean [6–9]. Optical seafloor imagery is now routinely acquired with deep sea vehicles, and often associated with other geophysical data (acoustic backscatter and multibeam bathymetry) and water column measurements (temperature, salinity, and chemical composition). High-resolution 3D models with associated textures are, thus, increasingly used in the representation and study of local areas of interest. However, most remotely operated vehicles (ROVs) or autonomous underwater vehicles (AUVs) that are currently used in science missions have limited optical sensing capabilities, commonly comprising a main camera used by the ROV-pilot, while larger workclass ROVs have additional cameras for maneuvering. Due to the nature of projective geometry, performing 3D reconstruction using only optical imagery acquired by monocular cameras results in a 3D model which is defined only up to scale, meaning that the unit in the model is not necessary a standard unit such as a meter [10]. In order to correctly disambiguate the scale, it is essential to use additional information in the process of model building. Predominantly, solutions in subaerial applications are based on the fusion of image measurement with robust and dependable satellite references, such as Global Navigation Satellite System (GNSS) [11–13], or ground control pointss (GCPs) [14–16], due to their accuracy and ease of integration. On the contrary, the water medium not only hinders the possibility of accurately establishing the control points, but also prevents the use of global positioning system (GPS) due to the absorption of electromagnetic waves. Hence the scale is normally disambiguated either using a combination of acoustic positioning (e.g., Ultra-Short BaseLine (USBL)) and inertial navigation system (INS) [17–19], or through the introduction of known distances between points in the scene [20].

In shallow water environments, i.e., accessible by divers, researchers have often placed auxiliary objects (such as a scaling cube [21], locknuts [22], graduated bars [23], etc.) into the scene, and used the knowledge of their dimensions to scale the model a posteriori. Such approaches, while applicable in certain scenarios, are limited to the use in small-scale reconstructions (e.g., a few tens of square meters), and in shallow water environments, due to the challenges in transporting and placing objects in deep sea environments. Similarly, laser scalers have been used since the late 1980s, projecting parallel laser beams onto the scene to estimate the scale of the observed area, given the known geometric setup of the lasers. Until recently, lasers have been mostly used in image-scaling methods, for measurements within individual images (e.g., Pilgrim et al. [24] and Davis and Tusting [25]). To provide proper scaling, we have recently proposed two novel approaches [26], namely, a fully-unconstrained (FUM) and a partially-constrained method (PCM), to automatically estimate 3D model scale using a single

optical image with identifiable laser projections. The proposed methods alleviate numerous restrictions imposed by earlier laser photogrammetry methods (e.g., laser alignment with the optical axis of the camera, perpendicularity of lasers with the scene), and remove the need for manual identification of identical points on the image and 3D model. The main drawback of these methods is the need for purposeful acquisition of images with laser projections, with the required additional acquisition time.

Alternatively, the model scaling can be disambiguated with known metric vehicle displacements (i.e., position and orientation from acoustic positioning, Doppler systems, and depth sensors [19,27,28]). As this information is recorded throughout the mission, such data are normally available for arbitrary segments even if they have not been identified as interesting beforehand. The classic range-and-bearing position estimates from acoustic-based navigation, such as USBL, have an uncertainty that increases with increasing range (i.e., depth) in addition to possible loss of communication (navigation gaps). Consequently, the scale information is inferred from data which is often noisy, poorly resolved, or both. Hence the quality of the final dataset is contingent on the strategy used in the fusion of image and navigation information. Depending on the approach, the relative ambiguity can cause scale drift, i.e., a variation of scale along the model, causing distortions [29]. Furthermore, building of large 3D models may require fusion of imagery acquired in multiple surveys. This merging often results in conflicting information from different dives, and affects preferentially areas of overlap between surveys, negatively impacting the measurements on the model (distances, areas, angles).

The need to validate the accuracy of image-based 3D models has soared as the development of both the hardware and the techniques enabled the use of standard imaging systems as a viable alternative to more complex and dedicated reconstruction techniques (e.g., structured light). Numerous evaluations of this accuracy are available for aerial and terrestrial 3D models (e.g., the works by the authors of [2,30–32]). Environmental conditions and limitations of underwater image acquisition preclude their transposition to underwater image acquisition and, to date, most underwater accuracy studies use known 3D models providing reference measurements. This leads to marine scientists nowadays being constantly faced with the dilemma of selecting appropriate analyses that could potentially be performed on the data derived from the reconstructed 3D models.

Early studies [21,33–38] evaluated the accuracy of small-scale reconstructions (mainly on coral colonies), comparing model-based and laboratory-based volume and surface areas for specific corals. More recently, auxiliary objects (e.g., locknuts [22], graduated bars [23], special frames [39,40], and diver weights [41]) have been used to avoid removal of objects from the environment. Reported inaccuracies range from 0.85% to 17%, while more recent methods achieve errors as low as 2–3% [22,41]. Diver-based measurements and the placement of multiple objects at the seafloor both restrict the use of these methods in shallow water or experimental environments, and hinder such approaches in deep sea environments (e.g., scientific cruises), where reference-less evaluation is needed instead, which has been performed in only a few experiments.

Ferrari et al. [38] evaluated their reconstruction method on a medium-size reef area (400 m) and a 2 km long reef transect. Maximum heights of several quadrants within the model were compared to in situ measurements, coupled with an estimation of structural complexity (rugosity). The stated inaccuracies in reef height were $18 \pm 2\%$. This study split larger transects into approx 10 m long sections to reduce potential drift, and hence model distortion. Similarly, Gonzales et al. [42] reported 15% error in rugosity estimates from stereo imaging and compared them with results from a standard chain-tape method, along a 2 km long transect. To the best of our knowledge, no other scale accuracy estimate of submarine large-area models has been published. Furthermore, although laser scalers are often used for qualitative visual scaling, they have never been used to evaluate the accuracy of underwater 3D models.

Objectives

Although a growing body of literature supports the belief that underwater image-based 3D reconstruction is a highly efficient and accurate method at small spatial extents, there is a clear absence

of scale accuracy analyses of models produced at larger scales (often in deep sea scenarios). Validation of 3D reconstruction methods and associated error evaluation are, thus, required for large underwater scenes and to allow the quantitative measurements (distances and volumes, orientations, etc.) required for scientific and technical studies.

The main goal of this paper is to present and use an automatic scale accuracy estimation framework, applicable to models reconstructed from optical imagery and associated navigation data. We evaluate various reconstruction strategies, often used in research and industrial ROV deep sea surveys.

First, we present several methods of 3D reconstruction using underwater vehicle navigation, to provide both scaling and an absolute geographic reference. Most commonly, SfM uses either an incremental or a global strategy, while the vehicle navigation may be considered a priori as part of the optimization process, or a posteriori after full 3D model construction. Here, we compare four different strategies resulting from combination of Incremental/Global SfM and the a priori and a posteriori use of navigation data. We discuss the impact of each strategy in the final 3D model accuracy.

Second, the four methods are evaluated to identify which one is best suited to generate 3D models that combine data from multiple surveys, as this is often required under certain surveying scenarios. Navigation data from different surveys may have significant offsets at the same location (x, y, z, rotation), show noise differences, or both. The changes between different acquisitions of a single scene are taken into account differently by each 3D reconstruction strategy.

Third, prior approaches, recently presented by Istenič et al. [26], to estimating model scale using laser scalers, namely FUM and PCM methods, are augmented with Monte Carlo simulations to evaluate the uncertainty of the obtained scale estimates. Furthermore, the results are compared to the kinds of estimates commonly used and suffering from parallax error.

Fourth, an automatic laser detection and uncertainty estimation method is presented. Accurate analyses require a multitude of reliable measurements spread across the 3D model, whose manual annotation is extremely labor-intensive, error-prone, and time-consuming, when not nearly impossible. Unlike previous detection methods, our method detects the centers of laser beams by considering the texture of the scene, and then determines their uncertainty, which, to the best of our knowledge, has not been presented in the literature hitherto.

With the data from the SUBSAINTES 2017 cruise (doi: 10.17600/17001000; [43]) we evaluate the advantages and drawbacks of the different strategies to construct underwater 3D models, while providing quantitative error estimates. As indicated above, these methods are universal as they are not linked to data acquired using specific sensors (e.g., laser systems and stereo cameras), and can be applied to standard imagery acquired with underwater ROVs. Hence, it is possible to process legacy data from prior cruises and with different vehicles and/or imaging systems. Finally, we discuss the best practices for conducting optical surveys, based on the nature of targets and the characteristics of the underwater vehicle and sensors.

2. Image-Based Underwater 3D Reconstruction

Textured 3D models result from a set of sequential processing steps (Figure 1). As scene geometry is computed entirely from the optical imagery, the end result directly depends on image quality and an adequate survey strategy. Compared to subaerial imagery, the unfavorable properties of the water medium (i.e., light attenuation and scattering effects) [5] cause blurring of details, low image contrast, and distance-dependent color shifts [44]. As such, acquisition is conducted at close range, thus limiting the observation area of any single image, while significantly increasing the amount of data collected and processed. The distance from the camera to the scene is often defined by a combination of several factors, such as the visibility, amount of available light, terrain roughness, and maneuverability of the imaging platform. As some of these factors may change from place to place and over time; it is common to adjust the distance during acquisition. The survey speed is also affected by several factors. Commonly, the survey speed is adjusted as a function of the distance to the scene

and of the image exposure time, in order to keep motion blurriness to minimum levels (often less than 2 pixels). Typically, the survey speed is in the order of a 1/4 of the distance to the scene, per second.
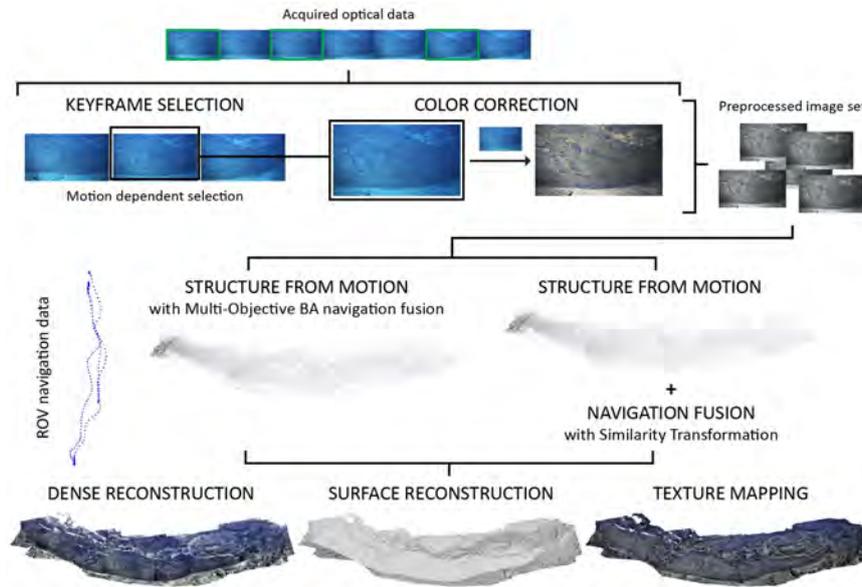

Figure 1. Flowchart of a 3D reconstruction process for underwater applications.

2.1. Preprocessing

Keyframe selection is hence important preprocessing step, used to remove unnecessary redundancy (i.e., images taken from very similar poses). Selecting too many images may represent an unnecessary increase of processing time, whereas the selection of too few images may result in missing observations and prevent the reconstruction of a single complete model. The commonly used strategy for image selection based on constant time intervals (e.g., selecting a frame every *n*-th second) is often not suitable; surveys with significant changes in speed and/or distance to the scene can lead to over- or underselection of images. Instead, we use an approach with implicit detection of frames with similar vantage points [45] through estimates of feature displacements between consecutive frames (e.g., Lucas–Kanade tracking algorithm [46]). For sufficiently dense image sets (e.g., video acquisitions), sharpness may be used for further selection (e.g., variance of Laplacian [47]).

Additionally, color correction can be used to counteract the degradation effects of water. Minimizing the effects of the water medium not only benefits human perception and interpretation of the scene, but also improves the quality and quantity of successful feature matches between image pairs [48], thus increasing the quality of the final model. Accurate color information recovery depends on knowledge of the physical image formulation process model which is rarely available in its completeness. Alternatively, color enhancing methods (e.g., Bianco et al. [49] used in our tests) can remove the attenuation effects, as well as the color cast introduced by an unknown illuminant (Figure 2).

| (a) | (b) | (c) | (d) |

Figure 2. (**a**) Original underwater (UW) image. (**b**) Chromatic components (α, β) of the estimated local illuminant. (**c**) White balanced image. (**d**) Final enhanced image.

2.2. Sparse Reconstruction

A concise set of preselected images is then used to jointly estimate the sparse 3D geometry of the scene (set of 3D points) and the motion of the camera (trajectory) through a technique called Structure from Motion (SfM). The inherent scale ambiguity of the reconstructed 3D structure and camera motion from a set of images taken by a monocular camera is addressed by either using the vehicle navigation a priori as part of the optimization process (multiobjective BA), or a posteriori through an alignment with the reconstructed camera path using a similarity transformation.

As the structure and motion parameters are inferred entirely from multiple projections of the same 3D point in overlapping images, the robustness of detection, and matching of feature points across the image set is important. In the evaluated approach, the features are detected as accelerated KAZE (AKAZE) [50], and described using Modified-SURF descriptors [51]. To avoid an empirical selection of the inlier/outlier threshold in the geometric filtering procedure (e.g., fundamental/essential matrix [10]), a parameter-free A Contrario Ransac (AC-RANSAC) [52], implemented in openMVG library [53], is used. The approach automatically determines the threshold and model meaningfulness by a statistical balance between the tightness of fitting of data and the number of the inliers.

Due to the nonlinearity in the projection process, a nonlinear optimization, Bundle Adjustment (BA), is required, with the final solution obtained by formulating a nonlinear least squares (NLS) problem. The cost function to be minimized is normally an image-based error, consisting of the sum of squared re-projection errors, defined as the distances between the 2-dimensional (2D) feature observations of the 3D point and their corresponding projections onto the images. Intrinsic camera parameters, if a priori known, can be excluded from the optimization, leading to lowering the complexity of the problem and thus improving the results. The problem can be efficiently solved using iterative methods such as Levenberg-Marquardt (LM) [54], which, however, only guarantees finding a local minimum of the optimizing function. This makes it extremely sensitive to the initial parameter estimate, leading to different strategies proposed for their initialization broadly classified as either: incremental or global.

Incremental SfM expands model reconstruction one image at a time, allowing for a gradual estimation of parameters for the newly added points and cameras. After each addition, intermediate BA can be performed to propagate and minimize the error of intermediate reconstructions. Incremental approaches are widely used, given that the intermediate partial reconstructions enable a more robust detection of outliers, and thus decrease the chance of convergence to a wrong local minimum. However, when no prior information about the scene is available, the initialization step of decomposing the fundamental/essential matrix is critical, as a poor selection of the seed pair of images can quickly force the optimization to a nonrecoverable state. Furthermore, as the method inherently gives disproportionate weight to images used at the beginning of the process, it can result in error accumulation. This may produce significant drift and fail to reconstruct the scene in the form of a single connected model. In our tests, the method of Moulon et al. [53,55] was used with a contrario model estimation.

Global SfM considers instead the entire problem at once, with full BA performed only at the end. To alleviate the lack of partial reconstructions, that identifies possible outliers, the parameter initialization is split into two sequential steps (i.e., rotation and translation estimation), the first one being more robust to a small number of outliers. This mitigates the need for intermediate nonlinear optimizations, as camera and scene points are estimated simultaneously in a single iteration. It also ensures an equal treatment of all the images, and, consequently, equal distribution of errors. These methods rely on averaging relative rotations and translations, thus requiring images to have overlap with multiple other images, to ensure meaningful constraints and mutual information. As a consequence, the reconstruction from a sparsely connected set of images will result in distorted or even multiple disconnected components. In our test, the method of Moulon et al. [53,56] is used.

2.3. Navigation Fusion

The result of the sparse reconstruction process is thus expressed as a set of 3D points $\mathcal{X} = \{X_k \in \mathbb{R}^3 \mid k = 1 \ldots L\}$; camera motion is defined with the following set of projection matrices, $\mathcal{P} = \{P_i = \left[\boldsymbol{R}_i^T \mid -\boldsymbol{R}_i^T \boldsymbol{t}_i \right] \mid i = 1 \ldots N\}$, where $P_i \in \mathbf{SE(3)}$ defines the projection from world to camera frame, and the following set of intrinsic camera parameters, $\mathcal{K} = \{K_i \mid i = 1 \ldots N\}$. As the joint parameter estimation is an inherently ill-conditioned problem, when estimated from a set of images acquired by a single camera, the solution is determined only up to an unknown scale [10]. The estimated parameters can be multiplied by an arbitrary factor, resulting in an equal projection of the structure on the images. A metric solution thus requires known measurements [20] or metric vehicle displacements (navigation/inertial priors) [19,27,28]. Depending on the availability of synchronization between the camera and the navigation, priors $\mathcal{C} = \{C_i \mid i = 1 \ldots N\}$ extracted from the ROV/AUV's navigation, can either be used in a multisensor fusion approach or to align the reconstructed camera path via a similarity transformation.

2.3.1. Multiobjective BA

When navigation priors are available for a significant proportion of images, then this information can be incorporated in the optimization through a multisensor fusion approach. When the measurement noises are not known, the problem of appropriate weighting of different objectives arises, as each of the sensor mean squared error (MSE) does not share the same unit neither the same significance [57].

Most commonly, there is no unique solution that would simultaneously optimize both the re-projection ($\mathcal{E}_{i,k(v)} = \left\| x_{i,k} - \mathrm{proj}(K_i, P_i, \mathbf{X}_k) \right\|^2$) and navigation fit errors ($\mathcal{E}_{i(n)} = T_i - C_i$). Instead there exists a hypersurface of Pareto optimal solutions, where one of the objective functions can only be improved by degrading the other [58]. Such solution space can be defined as a weighted compound function of the two objectives [57]. Assuming that both re-projection and navigation fit errors are independent and Gaussian, it is statistically optimal to weight the errors by their variance [59,60]:

$$\mathcal{E} = \frac{1}{M\sigma_v^2} \sum_{i,k} \left\| \mathcal{E}_{i,k(v)} \right\|^2 + \frac{1}{N\sigma_n^2} \sum_i \left\| \mathcal{E}_{i(n)} \right\|^2, \tag{1}$$

which can be rewritten as

$$\mathcal{E} = \frac{1}{M} \sum_{i,k} \left\| \mathcal{E}_{i,k(v)} \right\|^2 + \frac{1}{N} \lambda^2 \sum_i \left\| \mathcal{E}_{i(n)} \right\|^2, \tag{2}$$

where $\lambda = \sigma_v / \sigma_n$ indicates the ratio between the two covariances, representing the noise variance of each sensor measurement and M and N are the number of re-projection and navigation prior terms.

In such cases, the weighting can be selected empirically or through automatic weight-determining methods. For bi-objective optimizations, Michot et al. [57] have shown that the L-Curve criterion is the preferred selection method. This criterion is based on plotting the trade-off between the cost of the objectives using different weights, represented in log–log space. This plot has a typical L-curve shape, with two prominent segments. Each term dominating a segment (flat and vertical part) is used to detect the "corner" separating the two, essentially identifying a neutral objective dominance. The associated weight is considered to be the optimal, and representative of the ratio between the covariances of the sensors. Lying between two nearly flat segments, it can be easily identified as the point with maximum curvature.

2.3.2. Similarity Transformation

Alternatively, the navigation data can be used in an a posteriori step of rescaling and georeferencing. A similarity transformation, which minimizes the sum of differences between the reconstructed camera poses and their navigation priors, is applied to the reconstructed model. Depending on the survey pattern, this method can be used even in cases when the camera is not

synchronized with the navigation data. If the reconstructed path can be unambiguously matched to the path given by the navigation data, then the associations between the cameras and navigation poses can be determined through finding the closest points between the paths.

2.4. Dense, Surface and Texture Reconstruction

To better describe the scene geometry, a denser point cloud representation is computed using the method of Shen [61]. For each image reconstructed in SfM, a depth-map is computed, and subsequently refined to enforce consistency over neighboring views. These depth maps are merged into a single (dense) set of 3D points, where points with high photometric inconsistencies are removed to ensure the visibility constraints.

The final steps towards obtaining a final photo-realistic 3D model require estimating both a surface and high-quality texture to be pasted upon such surface. As underwater reconstructions are inevitably affected by noise and outliers [5], a method is used [62] to compute the most probable surface, by modeling the surface as an interface between the free and full space as opposed to directly using the input points. The reconstruction is completed by estimating the texture with a two step method [63]. The method prioritizes near, well-focused and orthogonal high-resolution views, as well as similar adjacent patches. Texture inconsistencies are mitigated by an additional photo-consistency check. Finally, any significant color discontinuities between neighboring regions are addressed by per-vertex-based globally optimal luminance correction as well as with Poisson image editing [64].

3. Model Evaluation Framework

Estimating the scale accuracy of 3D models reconstructed from underwater optical imagery and robot navigation data is of paramount importance since the input data is often noisy and erroneous. The noisy data commonly leads to inaccurate scale estimates and noticeable variations of scale within the model itself, which precludes use of such models for their intended research applications. Real underwater scenarios usually lack elements of known sizes that could be readily used as size references to evaluate the accuracy of 3D models. However, laser scalers are frequently used during underwater image collection to project laser beams onto the scene and can be used to provide such size reference.

The framework we will describe builds upon two recently introduced methods [26] for scale estimation of SfM-based 3D models using laser scalers. We extend the scale estimation process by including it into a Monte Carlo (MC) simulation, where we propagate the uncertainties associated with the image features and laser spot detections through the estimation process.

As the evaluated models are built with metric information (e.g., the vehicle navigation data, dimensions of auxiliary objects), their scale is expected to be consistent with the scale provided by the laser scaler (s_L). Therefore, any deviation from the expected scale value ($s = 1.0$) can be regarded as an inaccuracy of the scale of the model (ϵ_s). The error can be used to represent the percentage by which any spatial measurement using the model will be affected:

$$\epsilon_s = s_L - 1.0 = \frac{m}{\hat{m}} - 1.0 \,, \tag{3}$$

where m and $\hat{m}$ represent a known metric quantity and its model based estimate.

3.1. Scale Estimation

The two methods, namely, the fully-unconstrained method (FUM) and the partially-constrained method (PCM), are both suitable for different laser scaler configurations. FUM permits an arbitrary position and orientation for each of the lasers in the laser scaler, at the expense of requiring a full a priori knowledge of their geometry relative to the camera (Figure 3a). On the other hand, the laser-camera constraints are significantly reduced when using the PCM method. The laser origins have to be equidistant from the camera center and laser pairs have to be parallel (Figure 3b). However, in

contrast to prior image scaling methods [24,25], the lasers do not have to be aligned with the optical axis of the camera.
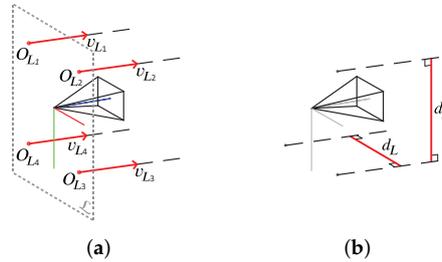

Figure 3. (**a**) Fully- and (**b**) partially-calibrated setup consisting of an optical camera and lasers, with the required information marked in red.

Both methods exploit images with visible intersections of the laser beams with the scene, not just the simple location of the laser spots. The model scale is estimated through a three step process: laser detection, pose estimation and scale estimation (Figure 4). The two initial steps are identical in both methods: First, a laser detection method determines the locations of laser spots on an image; second, the pose of the camera (wrt. the 3D model) at the time of image acquisition is estimated through a feature-based localization process.
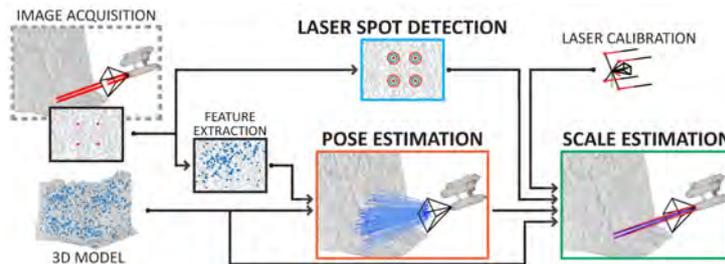

Figure 4. Flowchart of the scale estimation process depicting three crucial steps in scale estimation: laser spot detection, pose estimation, and scale estimation.

The initial camera-extrinsic values (and optionally also camera-intrinsics) are obtained by solving an Perspective-n-Point (PnP) problem [65] using 3D–2D feature pairs. Each pair connects an individual image feature and a feature associated with the sparse set of points representing the model. As these observations and matches are expected to be noisy and can contain outliers, the process is performed in conjunction with a robust estimation method A-Contrario Ransac (AC-RANSAC) [52]. The estimate is further refined through a nonlinear optimization (BA), minimizing the re-projection error of known (and fixed) 3D points and their 2D observation on the image.

The camera pose and location of the laser spots are lastly used either to estimate the position of the laser origin, so as to produce the recorded result (FUM), or else to estimate the perpendicular distance between the two parallel laser beams (PCM). As these predictions are based on the 3D model, they are directly affected by its scale, and can therefore be used to determine it through a comparison with a priori known values. As shown through an extensive evaluation in our previous work, both FUM and PCM can be used to estimate model scale regardless of the camera view angle, camera–scene distance, or terrain roughness [26]. After the application of a maximum likelihood estimator (BA) and a robust estimation method (AC-RANSAC), the final scale estimation is minimally affected by noise in the detection of feature positions and the presence of outlier matches.

In the fully-unconstrained method (Figure 5a), knowledge of the complete laser geometry is used (origins, O_L, and directions, v_L) to determine the position of laser emission $\hat{O}_L$, and then produce the

results observed on the image (Equation (4)). The laser origins $\hat{O}_L$ are predicted by projecting 3D points X_L, representing the location of laser beam intersections with the model, using a known direction of the beam v_L. As the points X_L must be visible to the camera, i.e., be in the line-of-sight of the camera, their positions can be deduced by a ray-casting procedure using a ray starting in the camera center and passing through the laser spot x_L detected in the image. The final scale estimate can then be determined by comparing the displacement of the $\hat{m}_L = \|\hat{O}_L\|$ with its a priori known value $\|O_L\|$.

$$\hat{O}_L = PX_L - \frac{PX_L \cdot c_z}{v_L \cdot c_z} v_L \,, \tag{4}$$

where P is defined as the projection from world to camera frame and c_z represents the optical axis of the camera.

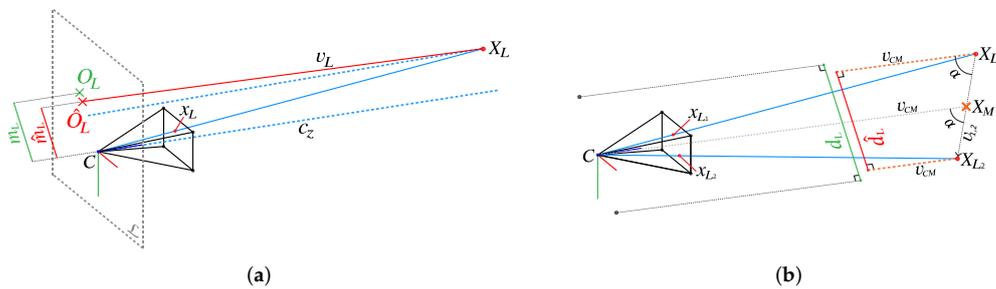(**a**)   (**b**)

Figure 5. Scale estimation using (**a**) fully-unconstrained and (**b**) partially-constrained approach, based on the 3D model and optical image depicting the laser beam projection on the scene intersection with the scene.

Alternatively, the partially-constrained method (Figure 5b) can be used when laser pairs are parallel but with unknown relation to the camera. As opposed to other image scaling methods, laser alignment with the optical axis of the camera is not required, allowing its application to numerous scenarios in which strict rigidity between camera and lasers is undetermined or not maintained (e.g., legacy data). To overcome the lack of information on the direction of laser beams with respect to the camera, the equidistance between the laser origins and the camera center is exploited. Laser beam direction is thus approximated with the direction of the vector connecting the camera center and the middle point between the two points of lasers intersections with the model v_{CM}. As we have shown in our previous work [26], this approximation can lead to small scaling errors only in the most extreme cases where the distance discrepancy between two points on the model is disproportionately large compared to the camera–scene distance. As underwater surveys are always conducted at sufficiently large safety distances, this scenario is absent in underwater reconstructions.

3.2. Uncertainty Estimation

Uncertainty characterization of each scale estimate is crucial for quantitative studies (precise measurement of distances and volumes, orientations, etc.), as required in marine science studies where accurate metrology is essential (such as in geology, biology, engineering, archaeology and others). The effect of uncertainties in the input values on the final estimate is evaluated using a MC simulation method. The propagation of errors through the process is modeled by repeated computations of the same quantities, while statistically sampling the input values based on their probability distributions. The final uncertainty estimate in scale is derived from the independently computed values.

Figure 6 depicts the complete MC simulation designed to compute the probability distribution of an estimated scale error, computed from multiple laser observations in an image. We assume that the sparse 3D model points, associated with the 2D features in the localization process, are constant, and thus noise free. On the other hand, uncertainty in the imaging process and feature detection

is characterized using the re-projection error obtained by the localization process. We also account for the plausible uncertainty in the laser calibration and laser spot detection, with each laser being considered independently.
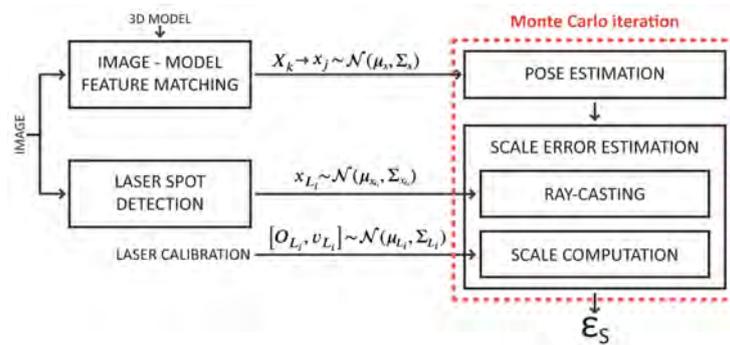

Figure 6. Monte Carlo simulation scheme used for propagating input uncertainties through the process of scale error estimation.

4. Laser Spot Detection

The accurate quantification of scale errors affecting 3D models derived from imagery requires numerous reliable measurements that have to be distributed throughout the model. As scale estimates are obtained by exploiting the knowledge of laser spot positions on the images, the quantity and quality of such detections directly determine the number of useful scale estimates. Furthermore, to properly estimate the confidence levels of such estimated scales, the uncertainty of the laser spot detections needs to be known.

The laser beam center is commonly considered to be the point with the highest intensity in the laser spot, as the luminosity of laser spots normally overpowers the texture of the scene. However, due to the properties of the water medium, the laser light can be significantly attenuated on its path to the surface before being reflected back to the camera. In such cases, the final intensity of the beam reaching the camera might be overly influenced by the texture at the point of the impact (Figure 7). As such, performing manual accurate annotations of laser spots tends to be extremely challenging and labor intensive, and even impossible in certain cases.

Figure 7. Example of image used for scale error evaluation with enlarged laser area.

Considerable attention has been given to the development of the image processing components of laser scanners, namely on laser line detection [66,67], while the automatic detection of laser dots from underwater laser scalers has only been addressed in a few studies. Rzhanov et al. [68] developed a toolbox (The Underwater Video Spot Detector (UVSD)), with a semiautomatic algorithm based on a Support Vector Machine (SVM) classifier. Training of this classifier requires user-provided detections. Although the algorithm can provide a segmented area of the laser dot, this information

is not used for uncertainty evaluation. More recently, the authors of [69] presented a web-based, adaptive learning laser point detection for benthic images. The process comprises a training step using k-means clustering on color features, followed by a detection step based on a k-nearest-neighbor (kNN) classifier. From this training on laser point patterns the algorithm deals with a wide range of input data, such as the cases of having lasers of different wavelengths, or acquisitions under different visibility conditions. However, neither the uncertainty in laser point detection nor the laser line calibration are addressed by this method.

To overcome the lack of tools capable of detecting and estimating the uncertainty in laser spot detection while still producing robust and accurate detections, we propose a new automatic laser detection method. To mitigate the effect of laser attenuation on the detection accuracy, scene texture is considered while estimating the laser beam center. Monte Carlo simulation is used to estimate the uncertainty of detections, considering the uncertainty of image intensities.

4.1. Detection

To determine laser spot positions on any image, the first step is a restriction of the search area to a patch where visible laser spots are expected (Figure 8a). Although not compulsory, this restriction minimizes false detections and reduces computational complexity and cost. The predicted area may be determined from the general pose of the lasers with respect to the camera, and from the range of distances to the scene.
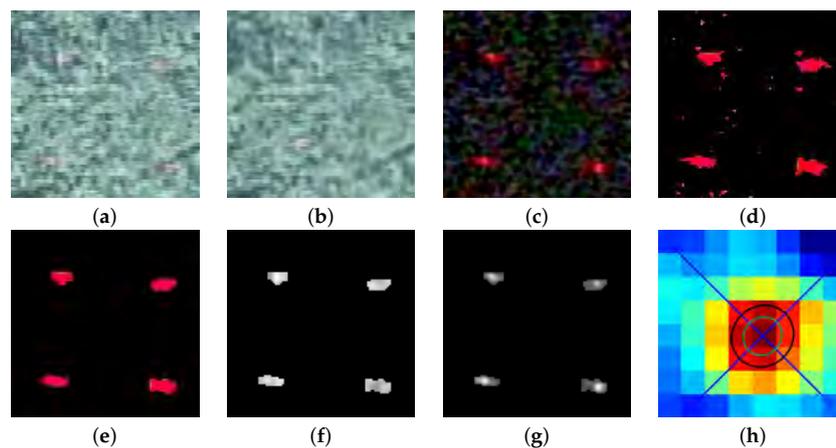

Figure 8. Laser spot detection: (**a**) predicted ROI of original image, (**b**) aligned auxiliary patch, (**c**) ROI after the removal of texture information (intensity ×5), (**d**) potential laser pixels after color thresholding, (**e**) filtered laser pixels, (**f**,**g**) estimated laser beam luminosity without/with texture removal, and (**h**) detected laser spot with detection uncertainty.

An auxiliary image is used to obtain a pixel-wise aligned description of the texture in the patch. This additional image is assumed to be acquired at a similar distance and with laser spots either absent or in different positions. This ensures visually similar texture information at the positions of the laser spots. This requirement is easily achievable for video acquisitions, as minor changes in camera pose sufficiently change the positions of the lasers. Alternatively, if still images are acquired, in addition to each image with visible laser spots, an additional image taken from a slightly different pose or in the absence of laser projections has to be recorded. The appropriate auxiliary patch is determined using normalized cross correlation in Fourier domain [70] using the original patch and the auxiliary image. The patch is further refined using a homography transformation estimated by enhanced correlation coefficient maximization [71] (Figure 8b). Potential discrepancies caused by the changes in the environment between the acquisitions of the two images, are further reduced using histogram matching. Once estimated, the texture is removed from the original patch to

reduce the impact of the texture on the laser beam spots. A low-pass filter further reduces noise and the effect of other artifacts (e.g., image compression), before detection using color thresholding (e.g., red color) in the HSV (Hue, Saturation, and Value) color space (Figure 8d). Pixels with low saturation values are discarded as hue cannot be reliably computed. The remaining pixels are further filtered using mathematical morphology (opening operation). The final laser spots are selected by connected-component analysis (Figure 8e).

Once the effects of the scene texture have been eliminated, the highest intensity point may be assigned to the laser beam center. In our procedure, the beam luminosity is characterized by the V channel of the HSV image representation. Figure 8f,g depicts the estimate of the laser beam luminosity with and without texture removal. Our proposed texture removal step clearly recovers the characteristic shape of the beam, with radially decreasing intensity from the center. Fitting a 2D Gaussian distribution to each laser spot allows us to estimate the center of the beam, assuming a 95% probability that the center falls within the top 20% of the luminance values (Figure 8h).

4.2. Uncertainty

Given that the estimation of the laser center is based on color information, it is important to consider the effect of image noise. Depending on the particularities of the image set, image noise is the result of the combined effects of sensor noise, image compression, and motion blur, among others. In our approach, the image noise is characterized by comparing the same area in two images taken within a short time interval (e.g., time between two consecutive frames), where the sensed difference can be safely attributed to image noise rather than an actual change in the environment.

For a randomly selected image from dataset FPA, the relation between assumed image noise (pixel-wise difference of intensities) and pixel intensities per color channel is depicted in Figure 9a, with the histogram of differences shown in Figure 9b. The results clearly illustrate a lack of correlation between image noise and pixel intensity levels or color channels as well as the fact that the noise can be well described by a Gaussian distribution. Furthermore, the analysis of 52 images from both datasets (FPA and AUTT28), acquired at a wide range of camera–scene distances and locations, in which the image noise was approximated by Gaussian distribution, indicating that the distribution of noise remains bounded regardless of the dataset or camera–scene distance (Figure 9c). While it is worth noting that the motion blur will be increasingly noticeable in images acquired at closer ranges, the analyzed range of distances is representative of the ones used for the evaluation presented in the following sections.
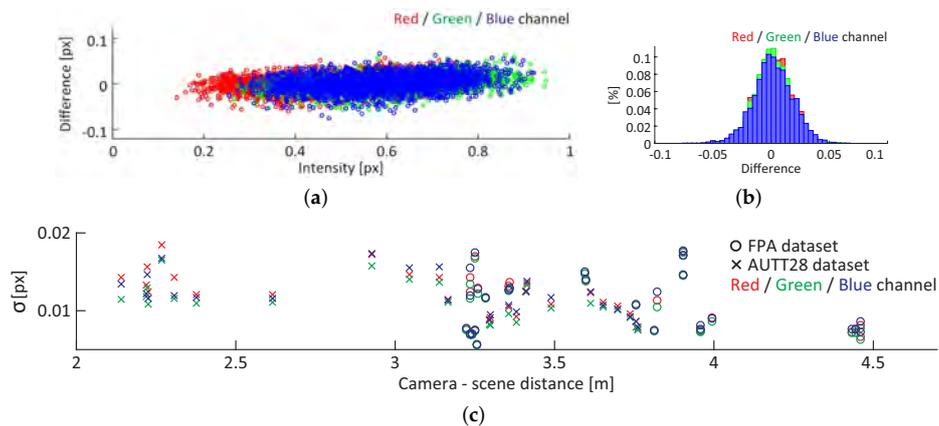

Figure 9. Characterization of image noise using an image from the FPA dataset: (**a**) image noise vs. pixel intensity, (**b**) distribution of noise per color channel and (**c**) standard deviation of estimated image noise for a series of images from multiple datasets and camera–scene distances, color-coded by the color channel.

To obtain a final estimate of confidence levels of detection, the uncertainty of image intensities are propagated through the laser detection process using MC simulation. At each iteration the noise is added independently to each pixel before the described laser spots detection. The iterations yield a set of independent detections, each characterized by a Gaussian distribution (Figure 8h). The final laser spot detection is subsequently obtained by determining an equivalent Gaussian using the Unscented Transform [72] to make this process less computationally expensive. If the laser is not detected in more than 80% of iterations, the detection is considered unstable and discarded. A set of laser spot detections obtained by a MC simulation is shown in Figure 10, together with the final joint estimation. Red and green ellipses represent 66% and 95% confidence levels for independent detections, while blue and cyan indicate the final (combined) uncertainty.
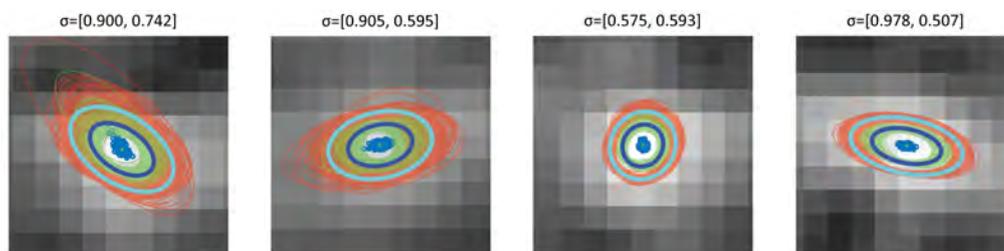

Figure 10. Examples of detected laser spot with uncertainty estimated through MC simulation for image shown in Figure 8. Individual detections and uncertainties are depicted with blue dots and red/green ellipses, while final uncertainty estimate is blue and cyan.

5. Dataset

During the SUBSAINTES 2017 cruise (doi: 10.17600/17001000) [43] extensive seafloor imagery was acquired with the ROV VICTOR 6000 (IFREMER) [73]. The cruise targeted tectonic and volcanic features off Les Saintes Islands (French Antilles), at the same location as that of the model published in an earlier study [9], and derived from imagery of the ODEMAR cruise (doi: 10.17600/13030070). One of the main goals of this cruise was to study geological features associated with a recent earthquake—measuring the associated displacement along a fault rupture—while expanding a preliminary study that presented a first 3D model where this kind of measurement was performed [9]. To achieve this, the imagery was acquired at more than 30 different sites along ~ 20 km, at the base of a submarine fault scarp. This is, therefore, one of the largest sets of image-derived underwater 3D models acquired with deep sea vehicles to date.

The ROV recorded HD video with a monocular camera (Sony FCB-H11 camera with corrective optics and dome port) at 30Hz, and with a resolution of 1920×1080 px (Figure 11). Intrinsic camera parameters were determined using a standard calibration procedure [74] assuming a pinhole model with the 3rd degree radial distortion model. The calibration data was collected underwater using a checkerboard of 12×8 squares, with identical optics and camera parameters as those later used throughout the entire acquisition process. The final root mean square (RMS) re-projection error of the calibration was (0.34 px, 0.30 px). Although small changes due to vibrations, temperature variation, etc. could occur, these changes are considered too small to significantly affect the final result.

Onboard navigation systems included a Doppler velocity log (Teledyne Marine Workhorse Navigator), fibre-optic gyrocompass (iXblue Octans), depth sensor (Paroscientific Digiquartz), and a long-range USBL acoustic positioning system (iXblue Posidonia) with a nominal accuracy of about 1% of the depth. As the camera was positioned on a pan-and-tilt module lacking synchronization with the navigation data, only the ROV position can be reliably exploited.
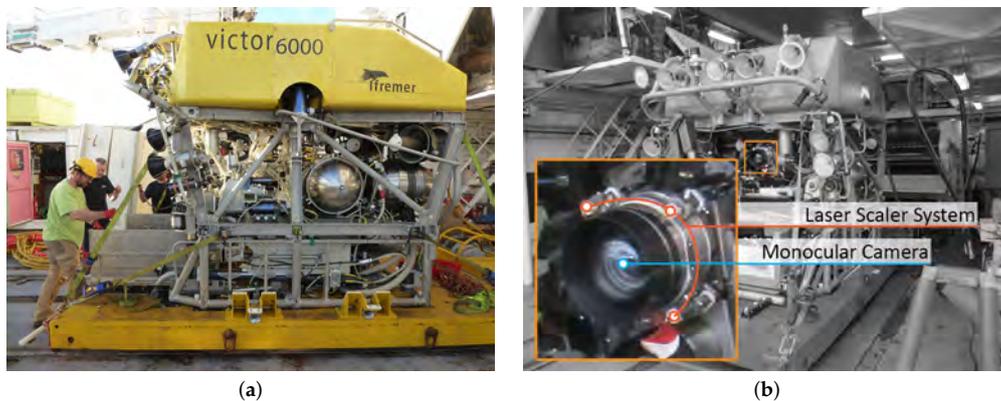
(**a**)   (**b**)

Figure 11. (**a**) ROV VICTOR 6000 (IFREMER) [73]. (**b**) Enlarged camera and laser system.

To date, 3D models have been built at more than 30 geological outcrops throughout the SUBSAINTES study area. Models vary in length between ∼10 m and ∼300 m horizontally, and extend vertically up to 30 m. Here we select two out of the 30 models (FPA and AUTT28), representative both of different survey patterns and spatial extents and complexity. Concurrently, evaluation data were collected with the same optical camera centered around a laser scaler consisting of four laser beams. For both selected datasets, numerous laser observations were collected, ensuring data spanning throughout the whole area. This enabled us to properly quantify the potential scale drifts within the models.

5.1. FPA

The first model (named FPA), extends laterally 33 m and 10 m vertically, and corresponds to a subvertical fault outcrop at a water depth of 1075 m. The associated imagery was acquired in a 10 min 51 s video recording during a single ROV dive (VICTOR dive 654). To fully survey the outcrop, the ROV conducted multiple passes over the same area. In total, 218 images were selected and successfully processed to obtain the final model shown in Figure 12. The final RMS re-projection errors of BA using different strategies are reported in Table 1. As expected, the optimizations using solely visual information and incremental approach are able to achieve lower re-projection errors, which, however, is not sufficient proof of an accurate reconstruction.

Figure 12. Textured 3D model of FPA area.

Table 1. Final RMS re-projection error (in pixels) achieved for the reconstructions of the FPA and AUTT28 models using all four strategies.

	FPA	AUTT28
Global SfM w/Similarity T.	0.42	0.44
Incremental SfM w/Similarity T.	0.37	0.36
Global SfM w/multiobjective BA	0.49	0.74
Incremental SfM w/multiobjective BA	0.44	0.56

5.2. AUTT28

The second model (named AUTT28), shown in Figure 13, is larger and required a more complex surveying scenario, as is often encountered in real oceanographic cruises. Initially, the planned area of interest was recorded during VICTOR dive 654. Following a preliminary onboard analysis of the data, a vertical extension of the model was required, which was subsequently surveyed during VICTOR dive 658. This second survey also partially overlapped with the prior dive, with overlapping images acquired at a closer range and thus providing higher textural detail. The survey also included a long ROV pass with the camera nearly parallel to the vertical fault outcrop, an extremely undesirable imaging setup. This second 3D model is the largest constructed in this area, covering a sub-vertical fault scarp spanning over 300 m laterally and 10 m vertically, with an additional section of approximately 30 m in height from a vertical ROV travel. This model is thus well suited to evaluate scaling errors associated with drift as it includes several complexities (survey strategy and geometry, multiple dives, extensive length and size of the outcrop). After keyframe selection, 821 images were used out of a combined 1 h 28 min and 19 s of video imagery to obtain reconstructions with the RMS re-projection error, as reported in Table 1.

Figure 13. Textured 3D model of AUTT28 area.

5.2.1. Multiobjective BA Weight Selection

Models built with a priori navigation fusion through the multiobjective BA strategy require a weight selection which represents the ratio between re-projection and navigation fit errors. As uncertainties of the two quantities are in different units and, more importantly, not precisely known, this selection must be done either empirically or automatically. Due to the tedious and potentially ambiguous trial-and-error approach of empirical selection, the weight was determined using L-Curve analysis.

The curve, shown in Figure 14a, uses the FPA dataset and 100 BA repetitions with weights λ ranging from 0.18 to 18. As predicted, the shape of the curve resembles an "L", with two dominant parts. The point of maximum curvature is determined to identify the weight with which neither objective has dominance (Figure 14b). As noise levels of the camera and navigation sensors do not significantly change between the acquisition of different datasets, the same optimal weight $\lambda = \mathbf{2.325}$ was used in all our multiobjective optimizations.

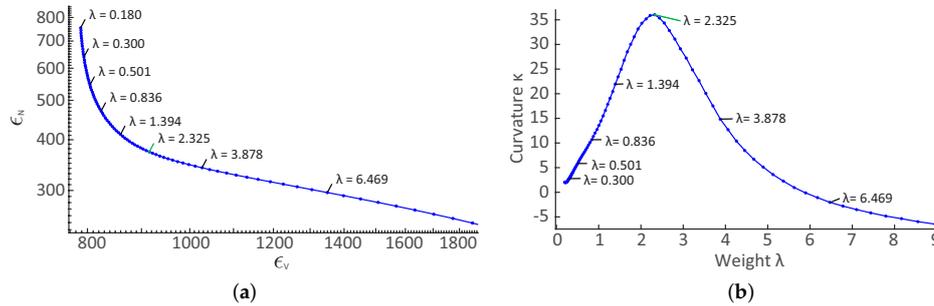**Figure 14.** (**a**) L-Curve for FPA dataset. (**b**) Curvature of L-Curve (shown on a smaller segment of weights for bigger clarity).

Given the heuristic nature of the optimal weight determination, it is important to evaluate the effects of the selection uncertainty on the final results. Figure 15 depicts the maximum difference in scale between the reconstructions computed using different weights and the reconstruction obtained with the optimal weight. The maximum expected scale differences were determined by comparing the Euclidean distances between the cameras of various reconstructions. The scale of the model is not expected to change if the ratio between the position of cameras does not change. The results show that the scale difference increases for an approximately 1% with the increment or decrement of λ by a value of one. Given that the optimal λ can be determined with the uncertainty of less than 0.1, as illustrated in Figure 14b, it can be assumed that the uncertainty in the determination of optimal weight has no significant effect on the final result.
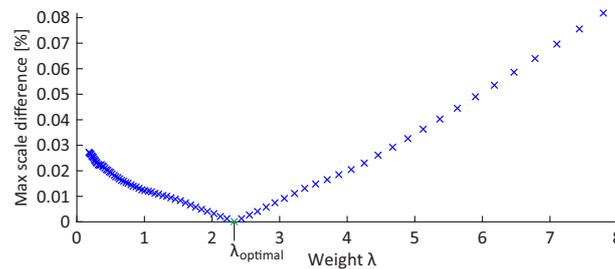

Figure 15. Maximum expected scale difference compared to the results obtained using the optimal weight λ (shown over a smaller range of weights for clarity).

5.2.2. Multisurvey Data

As is often the case in real cruise scenarios, the data for the AUTT28 model was acquired in multiple dives (Figure 16). When combining the data, it is important to consider the consequences of the merger. Optical imagery can be simply combined, given the short period of time between the two dives, in which no significant changes are expected to occur in the scene. In contrast, the merging of navigation data may be challenging; ROV navigation is computed using smoothed USBL and pressure sensor data, with expected errors in acoustic positioning being ~1% deep. As data was collected at roughly 1000 m depth, the expected nominal errors are ~ 10 m, or more in areas of poor acoustic conditions (e.g., close to vertical scarps casting acoustic shadows or reverberating acoustic pings). These errors, however, do not represent the relative uncertainty between nearby poses, but rather a general bias of the collected data for a given dive. Although constant within each dive, the errors can differ between the dives over the same area, and are problematic when data from multiple dives are fused. Models built with data from a single dive will only be affected by a small error in georeferencing, while multisurvey optimization may have to deal with contradicting navigation priors; images taken

81

from identical positions would have different acoustic positions, with offsets of the order of several meters or more.

This is overcome by introducing an additional parameter to be estimated, in the form of a 3D vector for each additional dive, representing the difference between USBL-induced offsets. Each vector is estimated simultaneously with the rest of the parameters in the SfM. For the case of AUTT28, the offset between the dives 654 and 658 was estimated to be $-2.53\,\text{m}$, $1.64\,\text{m}$, and $-0.02\,\text{m})$ in the x (E-W), y (N-S), and z (depth) directions, respectively. The disproportionately smaller *z* offset is due to the fact that the pressure sensor yields inter-dive discrepancies that are orders of magnitude smaller than the USBL positions.
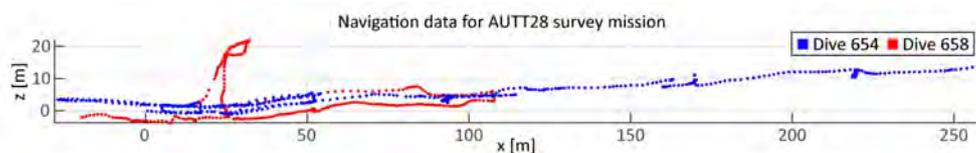

Figure 16. Navigation data for AUTT28 model merged from multiple dives (654-blue; 658-red).

5.3. Laser Calibration

Normally, the calibration process consists of the initial acquisition of images containing clearly visible laser beam intersections with a surface at a range of known or easily determined distances (e.g., using a checkerboard pattern). The 3D position of intersections, expressed relative to the camera, can be subsequently computed through the exploitation of the known 2D image positions and aforementioned camera–scene distances. Given a set of such 3D positions spread over a sufficient range of distances, the direction of the laser can be computed through a line-fitting procedure. Finally, the origin is determined as the point where the laser line intersect the image plane. Given the significant refraction at the air–acrylic–water interfaces of the laser housing, the images used in the calibration process must be collected under water.

In our case, the evaluation data was collected during multiple dives separated by several days, and with camera and lasers being mounted and dismounted several times. While the laser-scaler mounting brackets ensured that the laser origins remained constant, the laser directions with respect to the camera changed slightly with each installation. Due to operational constraints on the vessel, it was not possible to collect dedicated calibration data before each dive. However, given that the origins of the lasers are known a priori and remained fixed throughout the cruise, the only unknown in our setup is the inter-dive variation of the laser directions (relative to the camera and with respect to each other). The fact that independent laser directions do not encapsulate scale information (only the change of direction on an arbitrary unit) enables us to overcome the lack of dedicated calibration images and alternatively determine the set of points lying on each laser beam using images collected over the same area for which a 3D model has been constructed.

As our interest is only in the laser directions, the points used in the calibration can be affected by an arbitrary unknown scale factor, as long as this factor is constant for all of the points. Therefore, it is important to avoid models with scale drift, or to use data from multiple models with different scales. For each of the images used in the calibration, the camera was localized with respect to the model by solving an PnP problem [65] as in the FUM and PCM and additionally refined through BA. Each of the individual laser points were then determined by a ray-cast process and expressed in the camera coordinate system, before the direction of each of the lasers was determined by line-fitting. To maximize the conditioning of line-fitting, the selection of a model with the widest distance range of such intersection points and the smallest scale drift is important. This is the case for the AUTT28 model built using Global SfM and multiobjective BA, selected here. The global nature of the SfM and internal fusion of navigation data is predicted to most efficiently reduce a potential scale drift. As noisy laser detections are used to obtain the 3D points utilized in the calibration, laser spot

uncertainties were propagated to obtain the associated uncertainty of the estimated laser direction. A MC simulation with a 1000 repetitions was used. Together with the a priori known origin of the laser, this calibration provides us with all the information needed to perform scale estimation using the fully-unconstrained method.

The evaluation data were collected on dives 653, 654, and 658. As no camera and laser dismounting/mounting occurred between dives 653 and 654, there are two distinct laser setups: one for dives 653 and 654 and one for dive 658. Figure 17a depicts all laser intersections with the scene (for both AUTT28 and FPA models), as well as the calibration results, projected onto an image plane. Intersections detected in 3D model AUTT28 are depicted in black, and those from 3D model FPA are shown in orange. Similarly, the squares and circles represent dives 653/654 and dive 658, respectively. The projections of the final laser beam estimations are presented as solid and dotted lines. The figure shows a good fit of estimated laser beams with the projections of the intersections, both in the AUTT28 and FPA models.
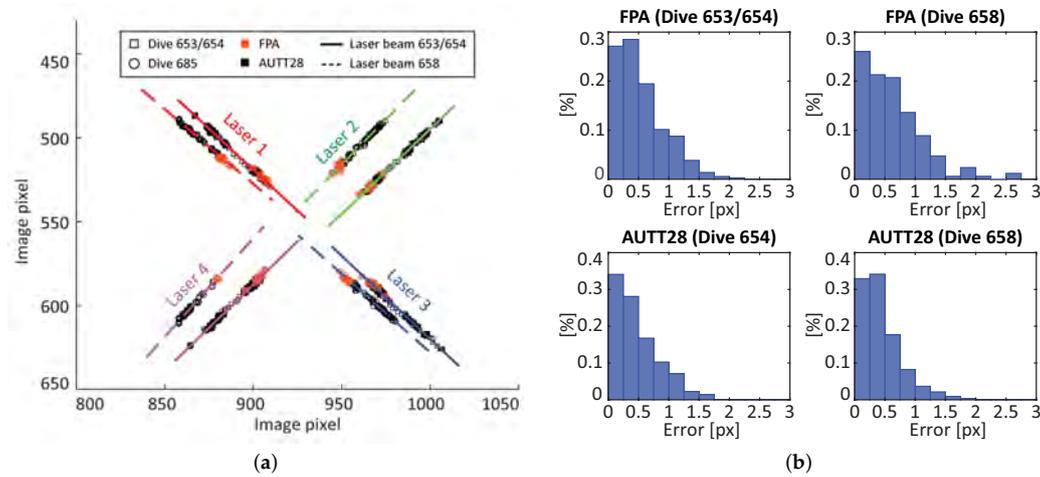

Figure 17. (**a**) Calibration results for dives 653/654 and 658. Solid and dotted lines represent the projections of estimated laser beams on the image plane, while projected laser intersections with the scene are depicted as squares/circles. (**b**) Histogram of perpendicular distances between the projected laser beams and projected laser intersections for the two datasets and dives.

The distributions of calibration errors, measured as perpendicular distances between the calibrated laser beams and the projected laser intersections with the scene, are depicted in Figure 17b, and the RMS errors are listed in Table 2. The adequate fit to the vast majority of AUTT28 points (RMS error < 0.6px) shows that the model used in the calibration had no significant scale drift. Furthermore, the fitting of the FPA related points (RMS error < 0.8px), which were not used in the calibration and are affected by a different scale factor, confirms that the calibration of laser directions is independent of the 3D model used, as well as of different scalings. The broad spread of the black points relative to the orange ones also confirms that the choice of the AUTT28 over the FPA model was adequate for this analysis. Lastly, it is worth reiterating that the data from all the models cannot be combined for calibration, as they are affected by different scale factors.

Table 2. RMS calibration errors (in pixels) measured as perpendicular distances between the projected laser beams and laser intersections with the scene for the two datasets and dives.

Dataset	FPA		AUTT28	
Dive	653/654	658	654	658
RMSE [px]	0.66	0.79	0.60	0.55

6. Results

As the accuracy of the measurements performed in 3D models for the purposes of quantitative studies (precise measurement of distances and volumes, etc.) depends on the strategy used for image-based 3D reconstruction, in addition to data quality itself, four of the most widely used approaches were evaluated:

A) Incremental SfM with a posteriori navigation fusion.
B) Global SfM with a posteriori navigation fusion.
C) Incremental SfM with multiobjective BA navigation fusion.
D) Global SfM with multiobjective BA navigation fusion.

The models for each of the two datasets (FPA and AUTT28) were built using each of the four strategies, and subsequently evaluated on multiple segments spread across the observed area. Using the model evaluation framework and laser spot detection method presented above, the scale accuracy and its associated uncertainties were automatically estimated using more than 550 images. To minimize the effects of possible false laser spot detections, only images with at least two confidently detected laser points were used. Furthermore, any images exhibiting excessive variation of the estimated scale between the individual lasers were discarded, as scale can be assumed to be locally constant.

6.1. Scale Accuracy Estimation

During accuracy evaluation, the scale error ϵ_s is estimated for each image independently. The final per-image scale error and its uncertainty are estimated through a MC simulation, with input variables (features, laser spot locations and laser calibration) sampled according to their probability distributions. The repeated computation with noisy data thus results in an equal number of final scale error estimates per laser. Figure 18 shows one example of such an estimation, together with the selected intermediate results of the evaluation process. As each MC iteration encapsulates the complete evaluation process (image localization, ray-casting, origin estimation, and scale error evaluation), intermediate distributions presented in Figure 18 are only shown for illustration, and are not used as distribution assumptions in the process itself.
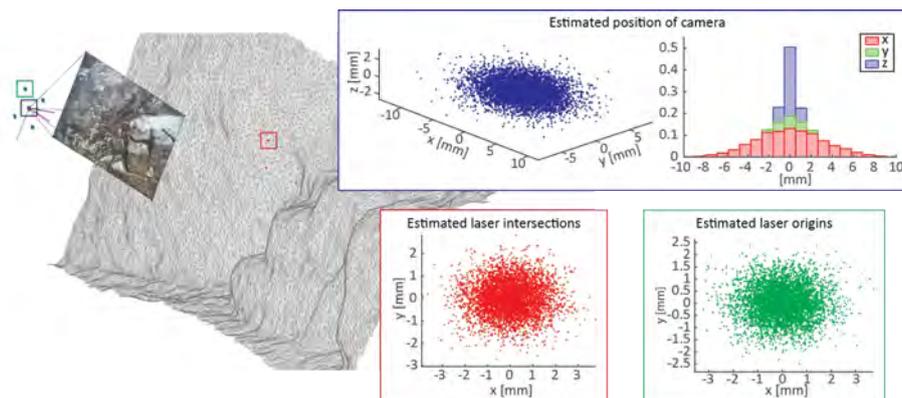

Figure 18. Intermediate results of a scale estimation procedure.

To satisfactorily represent the complexity of the process, 5000 iterations were used for each estimation. Figure 19 shows the evolution of the estimated scale error with associated uncertainty with increasing number of samples. After 500 iterations, the errors exhibit only minor fluctuations, and after 1500 iterations there is no noticeable difference. Hence, our selection of 5000 iterations is more than adequate to encapsulate the distribution of noise.

To demonstrate the advantages of our fully-unconstrained approach compared to previously available methods or our partially-constrained method, scale estimates obtained for each laser/laser pair are compared. Given the nonalignment of lasers with the optical axis of the camera, the majority of previous image-scaling methods (e.g., [24,25]) are not applicable. The only available option is thus a simplified approach where the Euclidean distance between a pair of 3D points (laser intersection points with the scene) is assumed to be the actual distance between the laser pair.

Results using different lasers (Figure 20) show that the FUM method produces the most consistent results. This is expected as the estimation process considers both individual laser directions and the geometry of the scene. The effect of scene geometry is clear when Figure 20a,b are compared. The slightly slanted angle together with the uneven geometry of the scene causes a large variation in the scale error estimates by the individual laser pairs. Similarly, the comparison of Figure 20b,c shows the effect of an inaccurate assumption of laser parallelism. This error depends on the camera–scene distance as shown in Figure 21. It is likely that the overestimation of laser pair 3–4 and the underestimation of other laser pairs can be explained by the use of oversimplified laser geometry. To validate this assumption, the results of the partially-constrained method were corrected by the expected errors (at $d = 2m$) induced by disregarding nonparallelism of laser beams (Figure 20d). While the result is nearly identical to that from a FUM method (Figure 20c), we note that the scale error in Figure 20c is computed for each laser individually, while the partially-constrained method considers laser pairs instead, and hence there are minor discrepancies.
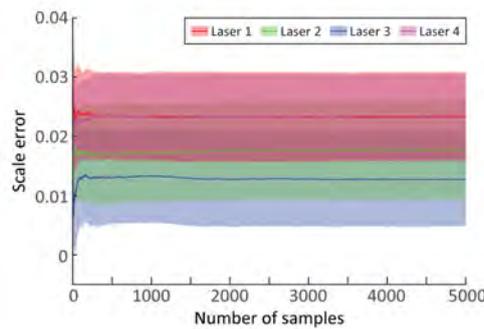

Figure 19. Evolution of scale error estimate with increasing number of MC iterations.
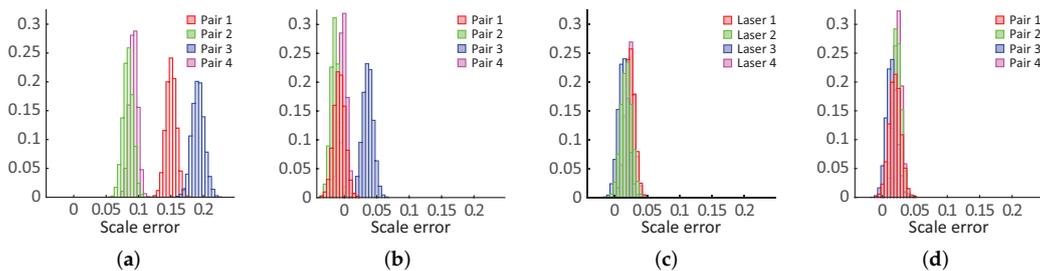

Figure 20. Estimated scale error per laser using different methods of computation: (**a**) Simplistic, (**b**) partially-constrained method, (**c**) fully-unconstrained method, and (**d**) partially-constrained method corrected for errors induced by nonparallelism of laser beams.
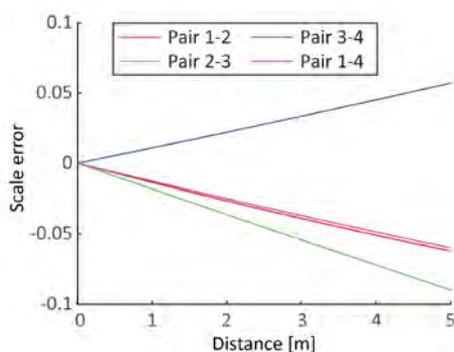
Figure 21. Error induced in scale error estimate due to disregarding nonparallelism of laser beams.

6.2. FPA

The accuracy of the FPA model was analyzed using 148 images (432 lasers). To represent the results concisely, measurements are grouped into 7 segments based on their model position (Figure 22 and Table 3). To ensure that the scale of the model did not vary within each segment, the maximum distance of any laser detection to the assigned segment center was set to 1 m.
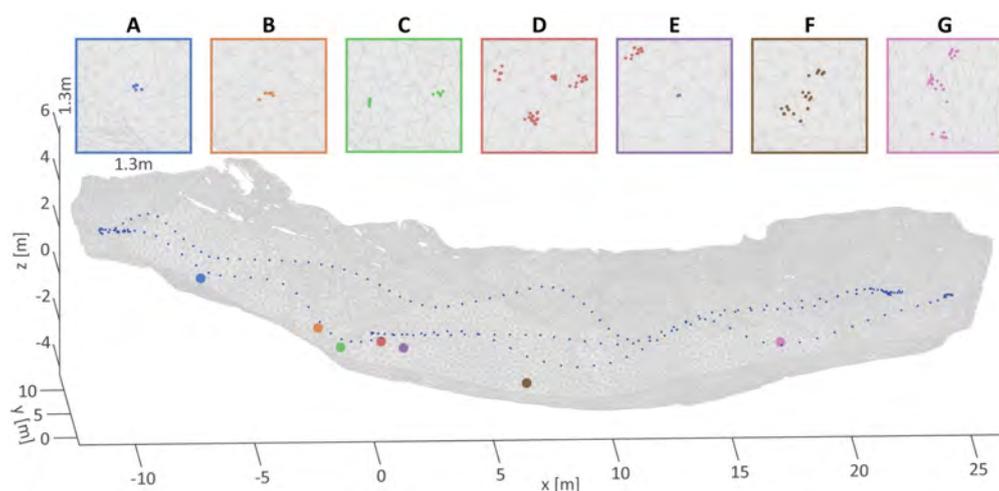

Figure 22. 3D reconstruction with the distribution of laser observations per segment for FPA area.

Table 3. Distribution of laser observations per segment for FPA area.

	A	B	C	D	E	F	G
# Images	12	6	39	40	12	24	15
# Lasers	29	12	137	103	33	53	36
Laser distance (min/max) [m]	3.23/3.29	4.44/4.46	3.03/3.50	3.58/4.01	3.59/3.61	3.19/3.37	3.19/3.79

FPA covers a relatively small area, and is imaged with multiple passes, thus providing redundancy that promotes model accuracy. Hence, it is expected to have only minor variations in scale error between areas. Figure 23 depicts the distribution of estimated scale errors for all four methods of 3D model construction. The comparison of results (Table 4) shows that accuracy does not significantly differ between them. The scale error varies between −1% and −5% with estimated uncertainties of approximately ±3%. The highest errors occur at the borders of the model. As expected, uncertainty is

closely related to the camera–scene distance, as small uncertainties in the laser direction translate to larger discrepancies at larger distances.
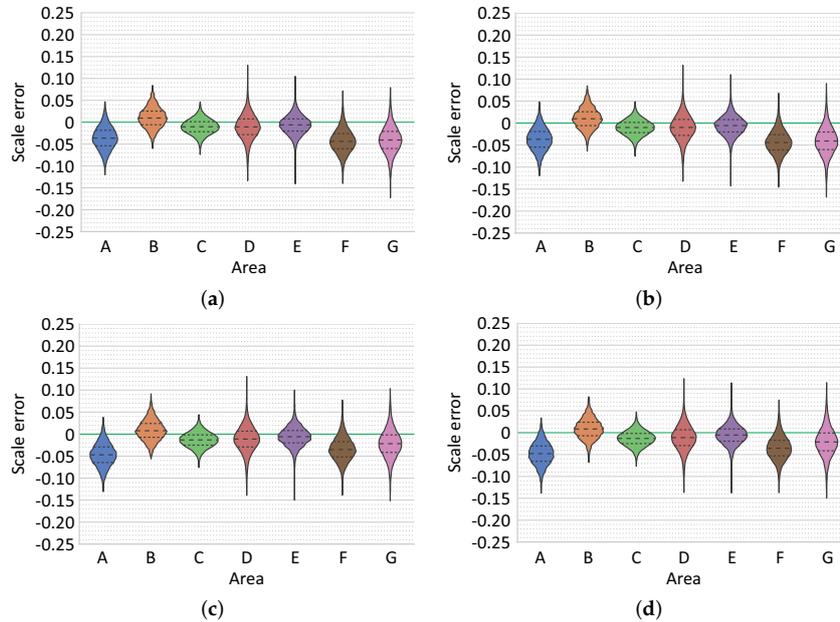

Figure 23. Estimated scale errors per segment for model FPA: (**a**/**b**) Global/Incremental SfM with similarity transformation navigation fusion. (**c**/**d**) Global/Incremental SfM with multiobjective BA navigation fusion.

Table 4. Estimated scale errors (%) per segment for different reconstructions of FPA area (values represent mean value with standard deviation).

	A	B	C	D	E	F	G
Global SfM w/Similarity T.	-3.6 ± 2.9	0.9 ± 3.2	-1.1 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-4.0 ± 2.9	-4.0 ± 3.4
Incremental SfM w/Similarity T.	-3.6 ± 2.9	0.9 ± 3.2	-1.0 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-4.0 ± 2.9	-4.0 ± 3.5
Global SfM w/multiobjective BA	-4.7 ± 2.8	0.7 ± 3.2	-1.3 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-3.1 ± 2.9	-2.2 ± 3.5
Incremental SfM w/multiobjective BA	-4.7 ± 2.9	0.7 ± 3.2	-1.3 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-3.2 ± 2.9	-2.1 ± 3.5

6.3. AUTT28

For model AUTT28, the evaluation data (images containing projected laser spots) were gathered during VICTOR dives 654 and 658, after the video acquisition of data used for 3D model creation. A total of 432 images with 1378 laser measurements were selected and grouped into 6 distinct sections throughout the 3D model, as shown in Table 5 and Figure 24. Dive 654 covered a longer vertical path (blue dots), while dive 658 (red dots) surveyed an additional horizontal segment together with parts of the area already viewed using dive 654. The higher density of red points indicates that the ROV observed the scene at a closer range during dive 658, requiring a higher number of images to obtain the necessary overlap compared to dive 654.

Table 5. Distribution of laser observations per segment for AUTT28 area.

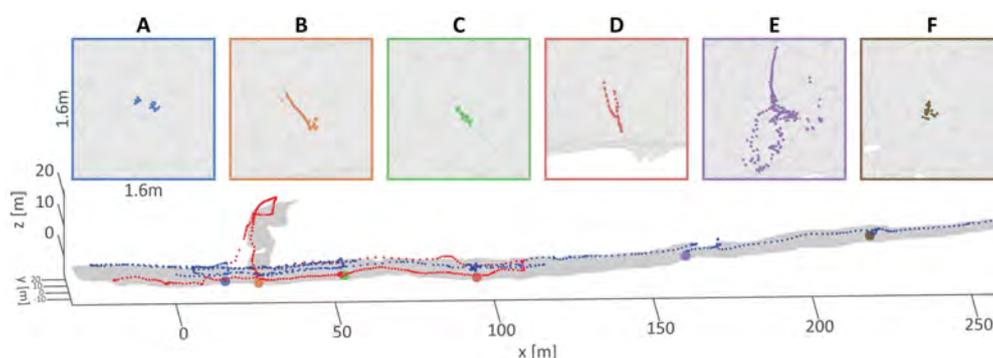|	A	B	C	D	E	F
# Images	30	47	20	46	261	28
# Lasers	97	169	51	165	812	84
Laser distance (min/max) [m]	1.95/2.25	2.13/2.67	2.90/3.36	3.21/3.89	1.70/4.14	3.63/3.79

Figure 24. 3D reconstruction with the distribution of laser observations per segment for AUTT28 area. Blue and red dots correspond to VICTOR dives 654 and 658.

The comparison of results (Table 6) shows that the models built using a posteriori navigation fusion (Figure 25a,b) are significantly impacted by scale drift ($\sim$15%), and that this impact is nearly identical regardless of the use of global or incremental SfM approaches. The gradual scale sliding observed is caused by inherent scale ambiguity of the two-view image pair geometry when BA is solely dependent on visual information. While this might not have been as obvious in the previous case, the long single pass of the camera, as performed in dive 654, introduces in this particular model numerous consecutive two-view image pairs, magnifying the scale drift. As shown in Figure 25c,d, additional constraints in the BA (e.g., navigation data) reduce ambiguity and, ultimately, nearly eliminate scale drift. Overall, scale error of the model built with global SfM using multiobjective BA is less than 1% with nearly zero scale drift, while a model built with incremental SfM approach showed a 2% scale drift along its 300 m length. It should be noted that the observed difference in scale estimates are within the uncertainty levels of the estimations, and therefore inconclusive.
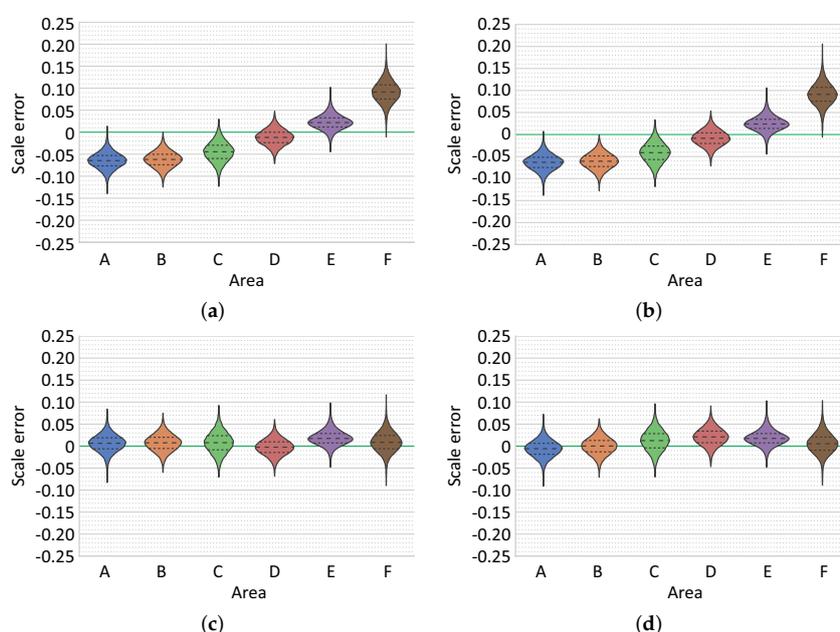

Figure 25. Estimated scale errors per segment for model AUTT28: (**a**/**b**) Global/Incremental SfM with similarity transformation navigation fusion. (**c**/**d**) Global/Incremental SfM with multiobjective BA navigation fusion.

Table 6. Estimated scale errors (%) per segment for different reconstructions of AUTT28 area (values represent mean value with standard deviation).

	A	**B**	**C**	**D**	**E**	**F**
Global SfM w/ Similarity T.	-6.4 ± 2.3	-6.3 ± 1.9	-4.5 ± 2.4	-1.1 ± 2.0	2.1 ± 2.2	9.2 ± 2.9
Incremental SfM w/Similarity T.	-6.3 ± 2.3	-6.1 ± 1.9	-4.1 ± 2.5	-0.8 ± 2.0	2.3 ± 2.3	9.3 ± 2.8
Global SfM w/multiobjective BA	0.7 ± 2.4	0.7 ± 2.0	0.8 ± 2.6	-0.2 ± 2.0	1.7 ± 2.3	0.9 ± 2.6
Incremental SfM w/multiobjective BA	-0.6 ± 2.4	-0.1 ± 2.1	1.2 ± 2.6	2.0 ± 2.2	1.7 ± 2.3	0.6 ± 2.7

6.3.1. Multiobjective BA vs. Similarity Transformation Navigation Fusion

The effects of different navigation fusion strategies are further demonstrated through the comparison of two reconstructions obtained using Global SfM with multiobjective BA and with similarity transformation (Figure 26). The reconstructions diverge on the outer parts of the model, consistent with a "doming" effect. A broad-scale systematic deformation produces a reconstruction that appears as a rounded-vault-distortion of a flat surface. This effect is a result of applying a rigorous re-projection error minimization to a loosely interconnected longer sequence of images taken from a nearly parallel direction, combined with slight inaccuracies in modeling of the radial distortion of the camera [14]. Multiobjective BA is able to reduce the effect by introducing the additional non-vision-related constraints, while the similarity transformation, with its preservation of angles between any three points and subsequent shape preservation, is not adequate for such corrections of the model deformations.
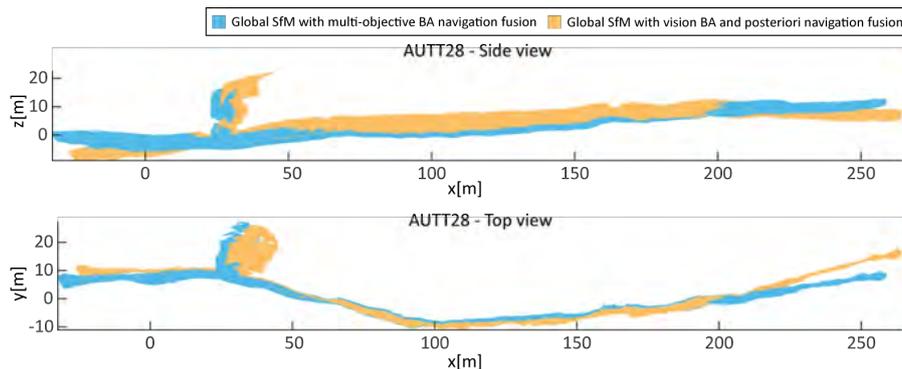

Figure 26. Comparison of navigation fusion strategies used in the reconstruction of 3D models.

6.3.2. Multisurvey Data Fusion

As explained in Section 5.2.2, the multimission data fusion can cause contradictory navigation priors during optimization. We address this by expanding the optimization problem with an additional 3D vector, representing the possible USBL offset between the recorded navigation data of the two dives. To examine the effects of this offset compensation on model construction, an additional model was constructed using raw navigation data (i.e., without offset compensation). Figure 27 depicts errors in the camera pose estimates with respect to their navigation priors, and shows a concentration of errors in areas imaged during both dives (Figure 16), where navigation priors of the two dives are incoherent. The errors dramatically decrease with the introduction of an offset, yielding an improved fitting solution. Alternatively, incoherence can cause model distortions to compensate for contradicting priors, as shown by abrupt changes of scale (area D in Figure 28).
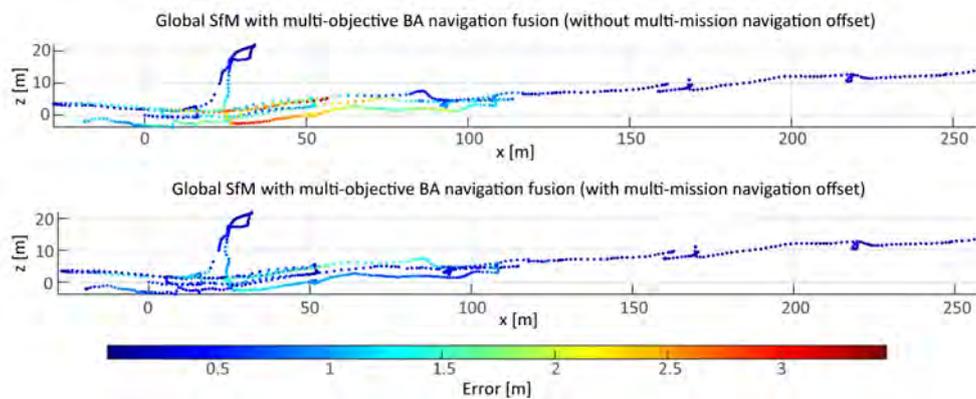
Figure 27. Comparison of the effects of multisurvey data fusion strategies on the estimated camera path.
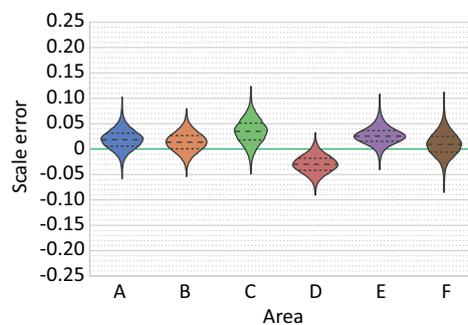

Figure 28. Estimated scale errors per segment for AUTT28 model built with Global SfM with multiobjective BA navigation fusion without an additional offset vector.

6.3.3. Scale Error Estimation Methods

To recover high-resolution and precise information from 3D models (lengths, areas, and volumes) it is important to use the most accurate method. As the nonalignment of lasers with the optical axis of the camera prevents the use of previous image-scaling methods (e.g., Pilgrim et al. [24]; Davis and Tusting [25]), two other methods could be used instead. Minor misalignments of laser scalers may be discarded for simplicity or lack of sufficiently distributed calibration data. In such cases, both our partially-constrained approach and the simplified direct 3D method, that assumes an equivalence of the Euclidean distance between the points of laser intersections and the beams themselves, could be used for the evaluation.

For this comparison the model with least scale drift was selected (Global SfM and multiobjective BA navigation fusion) to emphasize the effects of different methods on the results. Furthermore, as both the simplistic direct 3D and partially-constrained methods assume laser-pair parallelism, the analysis of these two methods was performed on data consisting of only laser pairs that were the closest to being parallel (Figures 29a and 30a), as well as on the complete dataset (Figures 29b and 30b), to show the effect that nonparallelism of laser beams may have on the different methods.
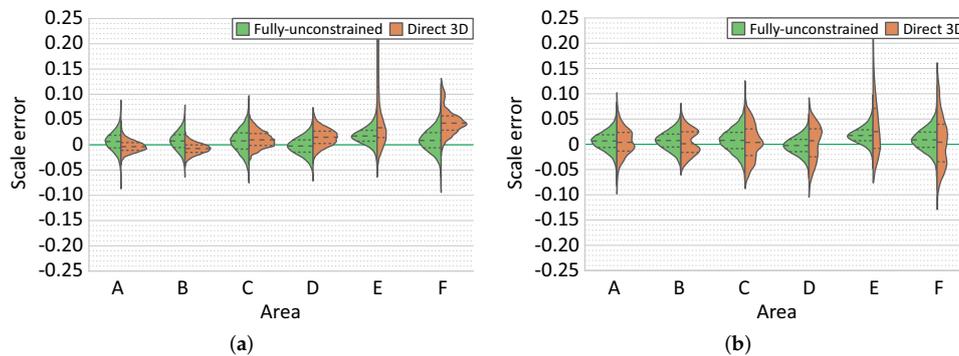
Figure 29. Comparison of estimated scale errors computed with fully-unconstrained and simplistic direct 3D method using (**a**) only nearly-parallel laser pairs and (**b**) all laser pairs.

As expected, in comparison to the simplistic approach (orange) (Figure 29a), our method (green) is significantly less impacted by the discrepancies in camera–scene distances at the two laser intersections caused by the deviation of camera–scene angle from perpendicularity. The spread of the estimated values within each segment for the direct approach is directly correlated with the span of camera–scene distances (most notable in area E in Figure 29). Although varying distances themselves do not play a role, they do however increase the probability of both having deviating camera-surface angles, and of violating the surface flatness requirement, which both result in discrepancies in camera–scene distances between the two laser points.

In contrast, the analysis of the results of the partially-constrained approach (Figure 30a) confirms that this method is unaffected by changes of camera angle and scene roughness. As expected, the results in sections D, E and F are nearly identical, with discrepancies in sections A, B and C. Sections A, B, and C were evaluated using data collected during dive 658, and D, E, and F were evaluated during dive 654; we attribute this discrepancy to the marginally larger error in nonparallelism of the laser configuration used during dive 658 compared to that of dive 654. This is clearly shown when the results are computed on the data from all laser pairs (Figure 30b), as nonparallelism of different laser pairs causes significant variation in the results. Segments acquired at closer ranges (A, B, and C), and therefore less affected by the errors in parallelism, have smaller errors than those of segments D and F, which are evaluated at larger distances. While similar multimodal distributions appear in the results of the simple direct 3D method, the clear multimodal peaks are suppressed by the effects of camera-surface angles and roughness of the surface model.
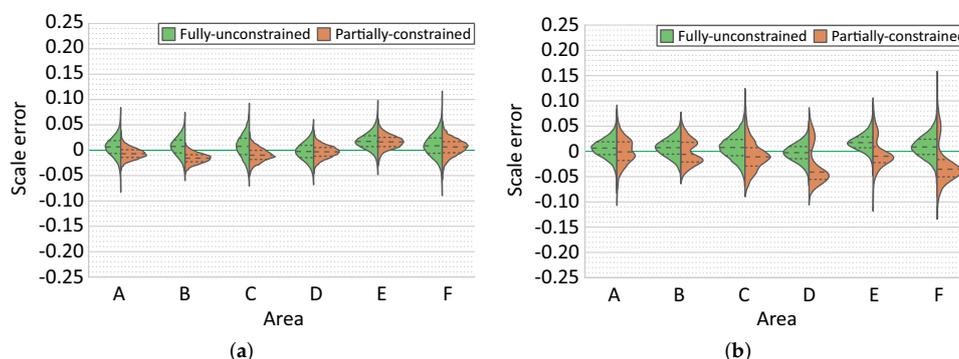

Figure 30. Comparison of estimated scale errors computed with fully- and partially-constrained method using (**a**) only nearly-parallel laser pairs; (**b**) all laser pairs.

7. Discussion

The comparison between the results show that the fully-unconstrained method is more consistent and accurate than the two other approaches, i.e., partially-constrained and simplistic direct 3D method. We note that by limiting the data to parallel laser pairs (dive 654), the partially-constrained method produced similar results. Therefore, the PCM approach can be used when the relation between parallel lasers and the camera is not known, opening up its use to numerous scenarios where strict rigidity between the camera and lasers is not maintained or determined (e.g., legacy data).

The effects of different reconstruction strategies were analyzed using two distinct survey scenarios. The first model (FPA dataset) was acquired with multiple passes over the same areas. Overlap of nonsequential images restricted the potential solutions of the optimization problem to a nearly identical solution regardless of the strategy (SfM or navigation fusion). In a second model (AUTT28 dataset), data were acquired during two separate surveys, and include a long single pass with the camera oriented nearly parallel to a vertical wall. The results demonstrate that surveys where sequential images are weakly connected are prone to produce broad-scale deformation (doming effect) in the final model. Rigorous minimization of the re-projection error, combined with the projective scale ambiguity, distorts the model, and can lead to further drift in the scale estimate. While the navigation fusion strategy did not play a role in the first model (FPA), the results of this second model (AUTT28) demonstrate the advantage of using multiobjective BA navigation fusion to process data with more complex survey patterns. Furthermore, the introduction of additional vectors in the optimization of multisurvey problems successfully accounted for the offset changes present in the underwater USBL-based navigation data, and thus minimize the effect of contradicting navigation priors.

8. Conclusions

This study presented a comprehensive scale error evaluation of four of the most commonly used image-based 3D reconstruction strategies of underwater scenes. This evaluation seeks to determine the advantages and limitations of the different methods, and to provide a quantitative estimate of model scaling, which is required for obtaining precise measurements for quantitative studies (such as distances, areas, volumes and others). The analysis was performed on two data sets acquired during a scientific cruise (SUBSAINTES 2017) with a scientific ROV (VICTOR6000), and therefore under realistic deep sea fieldwork conditions. For models built using multiobjective BA navigation fusion strategy, an L-Curve analysis was performed to determine the optimal weight between competing objectives of the optimization. Furthermore, the potential offset in navigation when using USBL-based positioning from different dives was addressed in a representative experiment.

Building upon our previous work, the lack of readily available measurements of objects of known sizes in large scale models was overcome with the fully-unconstrained method, which exploits laser scaler projections onto the scene. The confidence level for each of the scale error estimates was independently assessed with a propagation of the uncertainties associated with image features and laser spot detections using a Monte Carlo simulation. The number of iterations used in the simulation to satisfactorily represent the complexity of the process was validated through the analysis of the final estimate behavior.

As each scale error estimate characterizes an error at a specific area of the model, independent evaluations across the models enable efficient determination of potential scale drifts. To obtain a sufficient number of accurate laser measurements, an automatic laser spot detector was also developed. By mitigating the effects of scene texture using an auxiliary image, a much larger number of accurate detections was possible, even with greatly attenuated laser beams. The requirement of having laser spots either not present or in at different position in the auxiliary image is easily satisfied in video acquisitions, while an additional image has to be recorded if still images are collected. Furthermore, the recovery of characteristic shapes of laser spots with radially decreasing intensities enabled additional determination of the uncertainty of laser spot detections. In total, the scale errors have been evaluated on a large set of measurements in both models (432/1378) spread across them.

Finally, the comparison of results obtained using different reconstruction strategies were performed using two distinct survey scenarios. In surveys comprising a single dive and with multiple overlapping regions, the choice of reconstruction strategies is not critical, since all strategies perform adequately well. However, in more complex scenarios there is a significant benefit from using optimization including the navigation data. In all cases, the best reconstruction strategies produced models with scale errors inferior to 5%, with errors on the majority of each model area being around 1%. Acquisition of calibration data (points collected over a large range of distances) is indeed critical. Depending on laser setup, a modification of laser geometry is possible (e.g., during the process of diving due to pressure changes). As minor discrepancies in parallelism can cause significant offsets at the evaluating distance, performing a calibration in the field is desirable (e.g., approach of the scene illuminated with laser beams). Furthermore, our results also indicate and justify the importance of collecting a multitude of evaluation data at different locations and moments during the survey.

Author Contributions: Conceptualization, K.I., N.G., J.E., and R.G.; Data curation, K.I. and A.A.; Funding acquisition, J.E. and R.G.; Investigation, K.I.; Methodology, K.I., N.G., and R.G.; Software, K.I. and A.A.; Supervision, N.G., J.E., and R.G.; Writing—original draft, K.I.; Writing—review & editing, N.G., J.E., and R.G.

References

1. Lucieer, A.; de Jong, S.M.; Turner, D. Mapping landslide displacements using Structure from Motion (SfM) and image correlation of multi-temporal UAV photography. *Prog. Phys. Geogr. Earth Environ.* **2014**, *38*, 97–116. [CrossRef]

2. Javernick, L.; Brasington, J.; Caruso, B. Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry. *Geomorphology* **2014**, *213*, 166–182. [CrossRef]

3. Marteau, B.; Vericat, D.; Gibbins, C.; Batalla, R.J.; Green, D.R. Application of Structure-from-Motion photogrammetry to river restoration. *Earth Surf. Process. Landf.* **2017**, *42*, 503–515. [CrossRef]

4. Mathews, A.J.; Jensen, J.L.R. Visualizing and Quantifying Vineyard Canopy LAI Using an Unmanned Aerial Vehicle (UAV) Collected High Density Structure from Motion Point Cloud. *Remote Sens.* **2013**, *5*, 2164–2183. [CrossRef]

5. Campos, R.; Garcia, R.; Alliez, P.; Yvinec, M. A surface reconstruction method for in-detail underwater 3D optical mapping. *Int. J. Robot. Res.* **2014**, 64–89. [CrossRef]

6. Pizarro, O.; Friedman, A.; Bryson, M.; Williams, S.B.; Madin, J. A simple, fast, and repeatable survey method for underwater visual 3D benthic mapping and monitoring. *Ecol. Evol.* **2017**, *7*, 1770–1782. [CrossRef] [PubMed]

7. Bingham, B.; Foley, B.; Singh, H.; Camilli, R.; Delaporta, K.; Eustice, R.; Mallios, A.; Mindell, D.; Roman, C.; Sakellariou, D. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *J. Field Robot.* **2010**, *27*, 702–717. [CrossRef]

8. Rossi, P.; Castagnetti, C.; Capra, A.; Brooks, A.; Mancini, F. Detecting change in coral reef 3D structure using underwater photogrammetry: Critical issues and performance metrics. *Appl. Geomat.* **2019**, 1–15. [CrossRef]

9. Escartín, J.; Leclerc, F.; Olive, J.A.; Mevel, C.; Cannat, M.; Petersen, S.; Augustin, N.; Feuillet, N.; Deplus, C.; Bezos, A.; et al. First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history. *Earth Planet. Sci. Lett.* **2016**, *450*, 96–107. [CrossRef]

10. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2003. [CrossRef]

11. Soloviev, A.; Venable, D. Integration of GPS and vision measurements for navigation in GPS challenged environments. In Proceedings of the IEEE/ION Position, Location and Navigation Symposium, Indian Wells, CA, USA, 4–6 May 2010; pp. 826–833. [CrossRef]

12. Mian, O.; Lutes, J.; Lipa, G.; Hutton, J.; Gavelle, E.; Borghini, S. Accuracy assessment of direct georeferencing for photogrammetric applications on small unmanned aerial platforms. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *40*, 77. [CrossRef]

13. Forlani, G.; Dall'Asta, E.; Diotri, F.; Cella, U.M.d.; Roncella, R.; Santise, M. Quality Assessment of DSMs Produced from UAV Flights Georeferenced with On-Board RTK Positioning. *Remote Sens.* **2018**, *10*, 311. [CrossRef]

14. James, M.R.; Robson, S. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surf. Process. Landf.* **2014**, *39*, 1413–1420. [CrossRef]

15. Eltner, A.; Schneider, D. Analysis of Different Methods for 3D Reconstruction of Natural Surfaces from Parallel-Axes UAV Images. *Photogramm. Rec.* **2015**, *30*, 279–299. [CrossRef]

16. Mertes, J.; Zant, C.; Gulley, J.; Thomsen, T. Rapid, quantitative assessment of submerged cultural resource degradation using repeat video surveys and Structure from Motion. *J. Marit. Archaeol.* **2017**, *12*, 91–107. [CrossRef]

17. Sedlazeck, A.; Koser, K.; Koch, R. 3D reconstruction based on underwater video from ROV Kiel 6000 considering underwater imaging conditions. In Proceedings of the OCEANS 2009-EUROPE, Bremen, Germany, 11–14 May 2009; pp. 1–10. [CrossRef]

18. Pizarro, O.; Eustice, R.M.; Singh, H. Large Area 3-D Reconstructions From Underwater Optical Surveys. *IEEE J. Ocean. Eng.* **2009**, *34*, 150–169. [CrossRef]

19. Campos, R.; Gracias, N.; Ridao, P. Underwater Multi-Vehicle Trajectory Alignment and Mapping Using Acoustic and Optical Constraints. *Sensors* **2016**, *16*, 387. [CrossRef] [PubMed]

20. Garcia, R.; Campos, R.; Escartín, J. High-resolution 3D reconstruction of the seafloor for environmental monitoring and modeling. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems: (IROS 2011), San Francisco, CA, USA, 25–30 September 2011.

21. Cocito, S.; Sgorbini, S.; Peirano, A.; Valle, M. 3-D reconstruction of biological objects using underwater video technique and image processing. *J. Exp. Mar. Biol. Ecol.* **2003**, *297*, 57–70. [CrossRef]

22. Kalacska, M.; Lucanus, O.; Sousa, L.; Vieira, T.; Arroyo-Mora, J. Freshwater fish habitat complexity mapping using above and underwater structure-from-motion photogrammetry. *Remote Sens.* **2018**, *10*, 1912. [CrossRef]

23. Neyer, F.; Nocerino, E.; Gruen, A. Monitoring coral growth - The dichotomy between underwater photogrammetry and geodetic control network. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 2. [CrossRef]

24. Pilgrim, D.A.; Parry, D.M.; Jones, M.B.; Kendall, M.A. ROV Image Scaling with Laser Spot Patterns. *Underw. Technol.* **2000**, *24*, 93–103. [CrossRef]

25. Davis, D.; Tusting, R. *Quantitative Benthic Photography Using Laser Calibrations*; Undersea World: San Diego, CA, USA, 1991.

26. Istenič, K.; Gracias, N.; Arnaubec, A.; Escartin, J.; Garcia, R. Automatic Scale Estimation of Structure from Motion based 3D Models using Laser Scalers. *arXiv* **2019**, arXiv:1906.08019.

27. Nornes, S.M.; Ludvigsen, M.; Ødegard, Ø.; SØrensen, A.J. Underwater Photogrammetric Mapping of an Intact Standing Steel Wreck with ROV. *IFAC-PapersOnLine* **2015**, *48*, 206–211. [CrossRef]

28. Warren, M.; Corke, P.; Pizarro, O.; Williams, S.; Upcroft, B. Visual sea-floor mapping from low overlap imagery using bi-objective bundle adjustment and constrained motion. In Proceedings of the Australasian Conference on Robotics and Automation, Wellington, New Zealand, 3–5 December 2012.

29. Strasdat, H.; Montiel, J.; Davison, A.J. Scale drift-aware large scale monocular SLAM. *Robot. Sci. Syst. VI* **2010**, *2*, 7. [CrossRef]

30. Fonstad, M.A.; Dietrich, J.T.; Courville, B.C.; Jensen, J.L.; Carbonneau, P.E. Topographic structure from motion: a new development in photogrammetric measurement. *Earth Surf. Process. Landf.* **2013**, *38*, 421–430. [CrossRef]

31. Thoeni, K.; Giacomini, A.; Murtagh, R.; Kniest, E. A comparison of multi-view 3D reconstruction of a rock wall using several cameras and a laser scanner. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 573. [CrossRef]

32. Kalacska, M.; Chmura, G.; Lucanus, O.; Berube, D.; Arroyo-Mora, J. Structure from motion will revolutionize analyses of tidal wetland landscapes. *Remote Sens. Environ.* **2017**, *199*, 14–24. [CrossRef]

33. Bythell, J.; Pan, P.; Lee, J. Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques. *Coral Reef.* **2001**, *20*, 193–199. [CrossRef]

34. Courtney, L.A.; Fisher, W.S.; Raimondo, S.; Oliver, L.M.; Davis, W.P. Estimating 3-dimensional colony surface area of field corals. *J. Exp. Mar. Biol. Ecol.* **2007**, *351*, 234–242. [CrossRef]

35. McKinnon, D.; He, H.; Upcroft, B.; Smith, R.N. Towards automated and in-situ, near-real time 3-D reconstruction of coral reef environments. In Proceedings of the OCEANS'11 MTS/IEEE KONA, Waikoloa, HI, USA, 19–22 September 2011; pp. 1–10. [CrossRef]

36. Lavy, A.; Eyal, G.; Neal, B.; Keren, R.; Loya, Y.; Ilan, M. A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modeling. *Methods Ecol. Evol.* **2015**, *6*, 521–531. [CrossRef]

37. Gutiérrez-Heredia, L.; D'Helft, C.; Reynaud, E. Simple methods for interactive 3D modeling, measurements, and digital databases of coral skeletons. *Limnol. Ocean. Methods* **2015**, *13*, 178–193. [CrossRef]

38. Ferrari, R.; McKinnon, D.; He, H.; Smith, R.; Corke, P.; Gonzalez-Rivero, M.; Mumby, P.; Upcroft, B. Quantifying multiscale habitat structural complexity: A cost-effective framework for underwater 3D modeling. *Remote Sens.* **2016**, *8*, 113. [CrossRef]

39. Capra, A.; Dubbini, M.; Bertacchini, E.; Castagnetti, C.; Mancini, F. 3D reconstruction of an underwater archaelogical site: Comparison between low cost cameras. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 67–72. [CrossRef]

40. Guo, T.; Capra, A.; Troyer, M.; Grün, A.; Brooks, A.J.; Hench, J.L.; Schmitt, R.J.; Holbrook, S.J.; Dubbini, M. Accuracy Assessment of Underwater Photogrammetric Three Dimensional Modelling for Coral Reefs. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*. [CrossRef]

41. Raoult, V.; Reid-Anderson, S.; Ferri, A.; Williamson, J. How reliable is Structure from Motion (SfM) over time and between observers? A case study using coral reef bommies. *Remote Sens.* **2017**, *9*, 740. [CrossRef]

42. Gonzalez-Rivero, M.; Bongaerts, P.; Beijbom, O.; Pizarro, O.; Friedman, A.; Rodriguez-Ramirez, A.; Upcroft, B.; Laffoley, D.; Kline, D.; Bailhache, C.; et al. The Catlin Seaview Survey–kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2014**, *24*, 184–198. [CrossRef]

43. Escartín, J.; Le Friant, A.; Feuillet, N. SUBSAINTES Cruise Report, N/O L'Atalante-ROV VICTOR-AUV AsterX. *Fr. Oceanogr. Cruises* **2017**. [CrossRef]

44. Bryson, M.; Johnson-Roberson, M.; Pizarro, O.; Williams, S.B. True Color Correction of Autonomous Underwater Vehicle Imagery. *J. Field Robot.* **2015**. [CrossRef]

45. Hernández, J.D.; Istenič, K.; Gracias, N.; Palomeras, N.; Campos, R.; Vidal, E.; Garcia, R.; Carreras, M. Autonomous underwater navigation and optical mapping in unknown natural environments. *Sensors* **2016**, *16*, 1174. [CrossRef]

46. Bouguet, J.Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Int. Corp.* **2001**, *5*, 4.

47. Prados, R.; Garcia, R.; Gracias, N.; Escartin, J.; Neumann, L. A novel blending technique for underwater gigamosaicing. *IEEE J. Ocean. Eng.* **2012**, *37*, 626–644. [CrossRef]

48. Andono, P.N.; Purnama, I.; Hariadi, M. Underwater image enhancement using adaptive filtering for enhanced sift-based image matching. *J. Theor. Appl. Inf. Technol.* **2013**, *52*. [CrossRef]

49. Bianco, G.; Neumann, L. A fast enhancing method for non-uniformly illuminated underwater images. In Proceedings of the OCEANS-Anchorage, Anchorage, AK, USA, 18–21 September 2017; pp. 1–6.

50. Alcantarilla, P.F.; Solutions, T. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1281–1298. [CrossRef]

51. Agrawal, M.; Konolige, K.; Blas, M.R. Censure: Center surround extremas for realtime feature detection and matching. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2008; pp. 102–115. [CrossRef]

52. Moisan, L.; Moulon, P.; Monasse, P. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Process. On Line* **2012**, *2*, 56–73. [CrossRef]

53. Moulon, P.; Monasse, P.; Marlet, R. OpenMVG. An Open Multiple View Geometry Library. Available online: https://github.com/openMVG/openMVG (accessed on 11 January 2018).

54. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In *Vision Algorithms: Theory and Practice*; Springer: Berlin, Germany, 1999; pp. 298–372. [CrossRef]

55. Moulon, P.; Monasse, P.; Marlet, R. Adaptive structure from motion with a contrario model estimation. In *Asian Conference on Computer Vision*; Springer: Berlin, Germany, 2012; pp. 257–270. [CrossRef]

56. Moulon, P.; Monasse, P.; Marlet, R. Global fusion of relative motions for robust, accurate and scalable structure from motion. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 3248–3255. [CrossRef]

57. Michot, J.; Bartoli, A.; Gaspard, F. Bi-objective bundle adjustment with application to multisensor SLAM. *3DPVT'10* **2010**, *3025*.

58. Pareto, V. *Cours D'Economie Politique*; F. Rouge: Lausanne, Switzerland, 1896.

59. Aitken, A.C. On least squares and linear combination of observations. *Proc. R. Soc. Edinb.* **1936**, *55*, 42–48. [CrossRef]

60. Svärm, L.; Oskarsson, M. Structure from motion estimation with positional cues. In *Scandinavian Conference on Image Analysis*; Springer: Berlin, Germany, 2013; pp. 522–532. [CrossRef]

61. Shen, S. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. *IEEE Trans. Image Process.* **2013**, *22*, 1901–1914. [CrossRef] [PubMed]

62. Jancosek, M.; Pajdla, T. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. *Int. Sch. Res. Not.* **2014**, *2014*. [CrossRef]

63. Waechter, M.; Moehrle, N.; Goesele, M. Let there be color! Large-scale texturing of 3D reconstructions. In *Computer Vision–ECCV*; Springer: Berlin, Germany, 2014; pp. 836–850. [CrossRef]

64. Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. *ACM Trans. Gr. (TOG)* **2003**, *22*, 313–318. [CrossRef]

65. Ke, T.; Roumeliotis, S.I. An efficient algebraic solution to the perspective-three-point problem. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7225–7233.

66. Caimi, F.M.; Kocak, D.M.; Dalgleish, F.; Watson, J. Underwater imaging and optics: Recent advances. In Proceedings of the OCEANS 2008, Quebec City, QC, Canada, 15–18 September 2008; pp. 1–9. [CrossRef]

67. Massot-Campos, M.; Oliver-Codina, G. Underwater Laser-based Structured Light System for one-shot 3D reconstruction. In Proceedings of the SENSORS, Valencia, Spain, 2–5 November 2014; pp. 1138–1141. [CrossRef]

68. Rzhanov, Y.; Mamaenko, A.; Yoklavich, M. UVSD: Software for detection of color underwater features. In Proceedings of the OCEANS 2005 MTS/IEEE, Washington, DC, USA, 19–23 September 2005; Volume 3, pp. 2189–2192. [CrossRef]

69. Schoening, T.; Kuhn, T.; Bergmann, M.; Nattkemper, T.W. DELPHI—Fast and adaptive computational laser point detection and visual footprint quantification for arbitrary underwater image collections. *Front. Mar. Sci.* **2015**, *2*, 20. [CrossRef]

70. Padfield, D. Masked FFT registration. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2918–2925. [CrossRef]

71. Evangelidis, G.D.; Psarakis, E.Z. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1858–1865. [CrossRef]

72. Wan, E.A.; Van Der Merwe, R. The unscented Kalman filter for nonlinear estimation. In Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), Lake Louise, AB, Canada, 4 October 2000; pp. 153–158. [CrossRef]

73. Michel, J.L.; Klages, M.; Barriga, F.J.; Fouquet, Y.; Sibuet, M.; Sarradin, P.M.; Siméoni, P.; Drogou, J.F. Victor 6000: Design, utilization and first improvements. In Proceedings of the Thirteenth International Offshore and Polar Engineering Conference, Honolulu, HI, USA, 25–30 May 2003.

74. Bouguet, J.Y. Camera Calibration Toolbox for Matlab (2008). Available online: http://www.vision.caltech.edu/bouguetj/calib_doc (accessed on 22 January 2018).

5

Mission-time 3D Reconstruction with Quality Estimation

In this chapter we present the results of the collaboration between University of Girona (UdG) and the ARC Centre of Excellence for Robotic Vision at the Australian National University (ANU), through the research stay of the main author of this thesis. In this publication we propose a novel SfM-based system capable of producing a globally consistent 3D reconstruction together with an estimate of its uncertainty in mission-time or shortly after. The system integrates an incremental NLS solver for the BA problem, as well as a new method for fast covariance recovery proposed in another collaborative work, and an outlier rejection scheme based on a dual-map approach. The published work received the Norman Miller Prize for Best Student Poster Award at the MTS/IEEE OCEANS 2017 Conference.

Abstract

Accurate and detailed 3-dimensional (3D) models of the underwater environment are becoming increasingly important in modern marine surveys, since they convey immense information that can be easily interpreted. Techniques such as bundle adjustment (BA) and structure from motion (SfM), which jointly estimate sparse 3D points of the scene and camera poses, have gained popularity in underwater mapping applications. However, for large-area surveys these methods are computationally expensive and not intended for online application. This paper proposes an SfM pipeline based on solving the BA problem in an incremental and efficient way. Furthermore, the new system can provide not only the solution of the optimization (camera trajectory along time and the 3D points of the environment), but also the estimate of the uncertainty associated with the 3D reconstruction. This system is able to produce results in mission-time, i.e. while the robot is in the water or very shortly afterwards. Such quick availability is of great importance during survey operations as it allows data quality assessment in-situ, and eventual replanning of missions in case of need.

Keywords

Cameras; Estimation; Optimization; Feature extraction; Pipelines

6

RESULTS AND DISCUSSION

IN this chapter the main results of the thesis are discussed. Section 6.1 provides a summary of the completed work. Next sections discuss the results obtained along the road map of the thesis: Sections 6.2 and 6.3 discuss and present additional details of the pre-processing steps we often utilized in our processing pipeline. The main contributions of our work are commented in section 6.4 regarding the proposed scale estimation framework, in section 6.5 concerning the proposed laser spot detection method together with additional results showing the applicability of the approach on the legacy data sets, and in section 6.6 detailing the calibration process undertook in order to use the relevant data in the reconstruction strategy analysis discussed in section 6.7. The multi-survey particularities observed in the analysis are further discussed in section 6.8 with the presentation of additional results from a different scientific research cruise. Section 6.9 finally comments the results regarding the mission-time 3D reconstruction system developed during the research stay at the ANU in Canberra, Australia. Lastly, section 6.10 presents a brief overview of a selected subset of 3D models that have been built throughout the duration of this thesis. In order to avoid excessive repetition with the result sections of the publications of this compendium, the reader is referred to the corresponding chapter where the detailed results can be found, and we focus on providing additional insights and results where applicable.

6.1 Overview of completed work

The motivation for the development of this thesis was to improve the resources for the creation of accurate image-based 3D reconstructions of hard-to-access underwater (GPS-denied) environments that are subsequently to be used for a wide range of scientific purposes. Driven by the objective of establishing a framework that is independent of a specific robot platform, camera system and mission trajectory, numerous particularities as well as the effects of the water medium on the quality and availability of data had to be considered.

The basic reconstruction steps, i.e. matching, sparse, dense, surface and texture reconstruction, were supplemented by additional pre-processing steps of keyframe selection and color correction to counteract the downgrading effects of the water medium on the images (e.g. blurriness, low contrast), as described in the chapters 2 (section 3) and 4 (section 2). The benefits of our proposed scheme for efficient subset selection are discussed in more detail in section 6.2 with previously unpublished results, while the effects of image enhancement and color correction on the quality and final perception of the resulting models are presented in section 6.3. By comparing several intermediate results of the reconstruction steps using a real data set, we show the many positive effects that color correction has on the reconstruction process.

Depending on the area of application and the depth of the environment, the data to be processed with our framework is usually collected either by a diver or with a remotely operated / autonomous vehicle. Since optical sensing capabilities in science-focused explorations are still predominantly limited to a single optical camera or at best asynchronous camera systems (e.g. several GoPro cameras), the resulting models, derived from images alone, can only be estimated up to an unknown scale [79, 104]. This makes them useless for quantitative studies (e.g. precise measurement of distances, volumes, orientations, etc.). In the case of robotic vehicles, where the navigation data is significantly noisier and more prone to large spurious errors than in conventional air and terrestrial applications, the fusion strategy in SfM becomes of paramount importance. While there is a growing body of literature supporting the notion that small spatial scale image-based 3D underwater reconstructions (on the order of meters) are accurate, there is an evident lack of such scale accuracy evaluation of models produced on a larger scale (in the range of tens to hundreds of meters, and often in deep-sea scenarios).

In our work we have addressed this issue in several steps. First, we proposed two novel methods for the automatic scaling of SfM-based 3D reconstructions using commonly available laser scalers, namely the fully unconstrained method (FUM) and the partially constrained method (PCM). In the publication presented in chapter 3, we show that the exploitation of the image information beyond the mere localization of the laser spots allows these methods to overcome the limitations imposed by earlier laser photogrammetry methods and thus to accurately estimate the scale of the model. Both methods have been validated by a series of generated data sets based on a real 3D model of a hydrodynamic underwater vent and their applicability in real scenarios has been proven by a data set from the *SUBSAINTES* scientific cruise off Les Saintes Islands (in the French Antilles).

In order to use the above-mentioned scale estimation methods for scaling real models used in quantitative studies as part of accurate metrology applications in such areas as geology, biology, engineering, archaeology, etc. we have further extended our approach with uncertainty characterization. Described in detail in chapter 4 (section 3), we proposed a framework in which the confidence levels for each of the scale estimates are independently

assessed by propagating the uncertainties of the inputs (image features and laser spot detection) using a Monte Carlo (MC) simulation.

We took advantage of our participation in a scientific deep-sea cruise *SUBSAINTES* (doi: 10.17600/17001000) in 2017, where we equipped a ROV both with the navigation sensors (i.e. USBL, inertial measurement unit) and with a laser scaler, to obtain data for an analysis of image-derived underwater 3D models. Since we were able to determine the scale of the model by two approaches independently (i.e. scaling with laser scalers and by navigation fusion), it was possible to perform a scale error analysis on large-scale models of deep-sea underwater environments that had been built using navigation data as the only source of scale information.

Considering that the evaluation framework we propose depends integrally on the extraction of knowledge of the location and the associated uncertainty of the laser spots, we developed an automatic method that can be robustly used on a large number of images. This significantly reduced the very labor-intensive and error-prone nature of the process when it is performed manually. Presented in chapter 4 (section 4), our method facilitates the detection of the centers of the laser beams by taking into account the texture of the scene. To the best of our knowledge, this is the first publication solving the problem of determining the uncertainties of these detections based on the recovered characteristic shapes of the laser spots. While the results of its use on two data sets collected during the *SUBSAINTES* cruise are shown in the publication, section 6.5 contains, in addition to a brief summary, new results of the processing of an unrelated legacy data set. This is to further demonstrate the applicability of our method and its generalization capabilities to other data sets.

The two methods of scale estimation (fully unconstrained method (FUM) and partially constrained method (PCM)) presented here require the knowledge of certain aspects of the geometry between the lasers and the camera, which may be logistically difficult to obtain in a real application scenario. Motivated by this, we also addressed the alternative possibilities of the laser scaler-camera calibration as a way of obtaining that required information from the image data itself. In the case of the *SUBSAINTES* data, we exploited the fact that the laser origins were known and showed that the orientation of the lasers could be determined using data collected during the mission itself and with a 3D model affected by an unknown scale. To account for the uncertainties in the detection of the laser spot positions, we propagated them through the calibration process in order to obtain the associated uncertainties of the final estimated laser directions. Chapter 4 (section 5.3) explains this procedure in detail and presents the results that show the aforementioned independence of the obtained laser directions from the scale of the model from which they were calculated.

Finally, with the use of numerous measurements (detected laser spots) and the estimated laser-camera calibration, we were able to perform a comprehensive analysis of the most commonly used image-based 3D reconstruction strategies in the underwater community. The selected strategies consisted of the combinations of incremental/global SfM and the *a priori / a posteriori* use of navigation data. The advantages and limitations of the different approaches were determined by analysing two distinct underwater survey scenarios acquired during the *SUBSAINTES* cruise with a scientific ROV (VICTOR 6000). They encapsulate realistic deep-sea field working conditions such as different survey strategies, multiple dives and extensive outcrop lengths and sizes.

As part of the analysis, we also investigated some other important aspects of the

image-based reconstruction process in underwater scenarios. For *a priori* (multiobjective BA) navigation fusion strategy we reiterated and analysed the process of determining the optimal weight, to model the ratio of the uncertainties of image localization and navigation data. To avoid its potentially ambiguous determination process in our analysis, we used L-Curve analysis to estimate the optimal weight between competing optimization goals. In line with results in the literature, we showed that while the optimization process is robust to small deviations from the optimal weight, this is no longer true for the case of large deviations, which can cause errors of several percent for our data sets, thus rendering the models unsuitable for scientific applications. The effects of combining data collected in several surveys were also evaluated. We have shown that the introduction of additional vectors in the optimization of such problems successfully takes into account the potential offset in navigation when using USBL-based positioning from different dives. We further confirm these claims by presenting the results of processing an unrelated data set acquired during the *MIDAS* cruise in section 6.8. In this section, a similar navigation offset had to be considered to facilitate the detection of changes between models of the same area.

While the work presented so far has addressed various steps to potentially mitigate the effects of unfavourable water characteristics, the outcome of the reconstruction process still depends, to a large extent, on the conditions and strategies applied in the acquisition process itself. Since the computational requirements of the reconstruction phase dictate its decoupling from the acquisition, especially in large-scale scenarios, it prevents active feedback on the quality of the subsequent reconstruction directly at the time of data acquisition. The immediate consequence is a high demand for careful mission planning and redundant data acquisition, which often still does not prevent the reconstructions from being of poor quality and possibly requiring costly additional missions in the same area.

During my research stay at ANU in 2017 we addressed this issue by proposing a new SfM system capable of estimating the trajectory and a sparse 3D representation of the scene together with the associated uncertainties in mission time, i.e. while the robot is in the water or shortly after. The proposed approach, presented in chapter 5, is able to process large-scale 3D reconstructions in a robust and globally consistent way, exploiting the incremental fashion of solving the BA problem and covariance recovery. We performed and reported on an analysis of the proposed system as a whole and of the individual methods for incremental processing and covariance recovery using a large-scale underwater data set. The related publications [4, 6] received Best Paper Honorable Mention Award at the International Conference on 3D Vision - 3DV 2017 (Qingdao, China) and Norman Miller Best Student Poster Award at the IEEE OCEANS'17 (Aberdeen, UK).

With the general objective of this thesis being the creation of tools and strategies for accurate image-based 3D reconstruction in underwater scenarios for scientific purposes, we present in section 6.10 an overview of the most significant models built during the whole duration of this thesis. These models have been created as a result of collaboration with a large number of other scientists and researchers in many fields. As described, some models have been used for accurate measurements in scientific fields, while others have been used for various purposes ranging from the generation of superior omnidirectional video, augmented reality applications, or simply as evidence that algorithms such as mission planning have successfully ensured satisfactory data acquisition. In this way, we want to show that the work presented in this thesis has contributed to the development of studies in a variety of fields such as robotics, underwater photogrammetry and the natural sciences.

6.2   Keyframe Selection Scheme

Keyframe selection is an important pre-processing step, used for removing unnecessary redundancies in the image set (i.e., images taken from very similar poses), while simultaneously ensuring the selection of the best possible subset of images. Although still photography allows for higher quality image acquisition, video remains a preferred option for situations in which the conditions are unpredictable. By collecting a significantly larger number of images (frames) with a large overlap, video enables a greater flexibility in post-acquisition image selection. Significant changes of camera-scene distance and/or survey speed within a mission, often occurring in dynamic scenarios, can cause considerable variations in frame overlap. If the most commonly used constant-frame selection strategy (i.e., selecting a frame every n-th second) is used, it can cause an over- or under-selection of images. Over selection can not only significantly degrade the efficiency of the reconstruction process due to exaggerated observation redundancy, but also potentially lead to unreliable estimates of the depth information due to small displacements of the camera [79]. Conversely, the under selection of frames will lead to segments of missing observations and thus preventing the reconstruction of a complete model.

As described in chapters 2 (section 3.1) and 4 (section 2.1), our keyframe selection scheme uses a simple dual approach. First, a measurement of feature displacement on an image plane and the ratio of successfully tracked features between keyframes is used to discard frames whose overlap and content is excessively similar to the already selected frames. Once the threshold is reached, a small number of adjacent and thus similar frames are identified and used to find the sharpest among them. This prevents a random selection of a blurry frame when a sharper version is available. The blurriness score is calculated as a variance of the Laplacian over an image [143], since its value increases with rapid changes in intensity in the image, potentially indicating sharper edges of a more in-focus image.

The performance of the scheme is illustrated by the presentation of the results of processing two different video sequences (in both cases 6 min 40 s sequences were used to increase the clarity of the figures). The videos were collected with different cameras in two significantly different mission types. First, we present the data used in our joint work Vidal *et al.* [11] where the robot (Sparus II AUV) explored a natural rock formation using an autonomous mission planning algorithm designed to ensure sufficient data acquisition. While the ideal path would maintain constant both the robot-scene distance and the speed of the robot, the difficult conditions (near-surface depth of 2.5 m and lack of prior information about the environment) led to various changes in the robot-scene distance and rapid changes in speed. The estimated shift and ratio of successfully tracked features between frames is shown in Figure 6.1. The dots represent frames where our system has decided to select a new keyframe, either because the content has changed sufficiently (red dots) or because the number of tracked features from the last keyframe has fallen below the allowable threshold (blue dot). In our tests, the maximum change in overlap was set at 15% and the minimum acceptable tracking rate was set at 70%. It can be seen that, when the robot is near the scene (marked orange), the system extracted frames more frequently because the change in image content was more abrupt. This can be seen particularly well in Figure 6.2, where the determined dynamic frame extraction frequency is shown in blue, with straight red and magenta lines representing the constant extraction rates of 1fps and 0.5fps respectively. The determined extraction frequency varies considerably and increases

in the marked segments where the robot is near the scene. The fact that there is not a single horizontal line that could satisfactorily match the calculated frequency clearly shows the advantage of content-adapted frame extraction. If the constant extraction frequency was used and if it was set high (the line in the graph would be lower), the frames would be over-sampled in many areas, while at lower frequency (the line would be higher) the reconstruction process would result in several unconnected sub-models. The total number of frames extracted for the presented video with our scheme was 158, while at a constant rate of 1/0.5 fps 401/201 frames would have been selected.
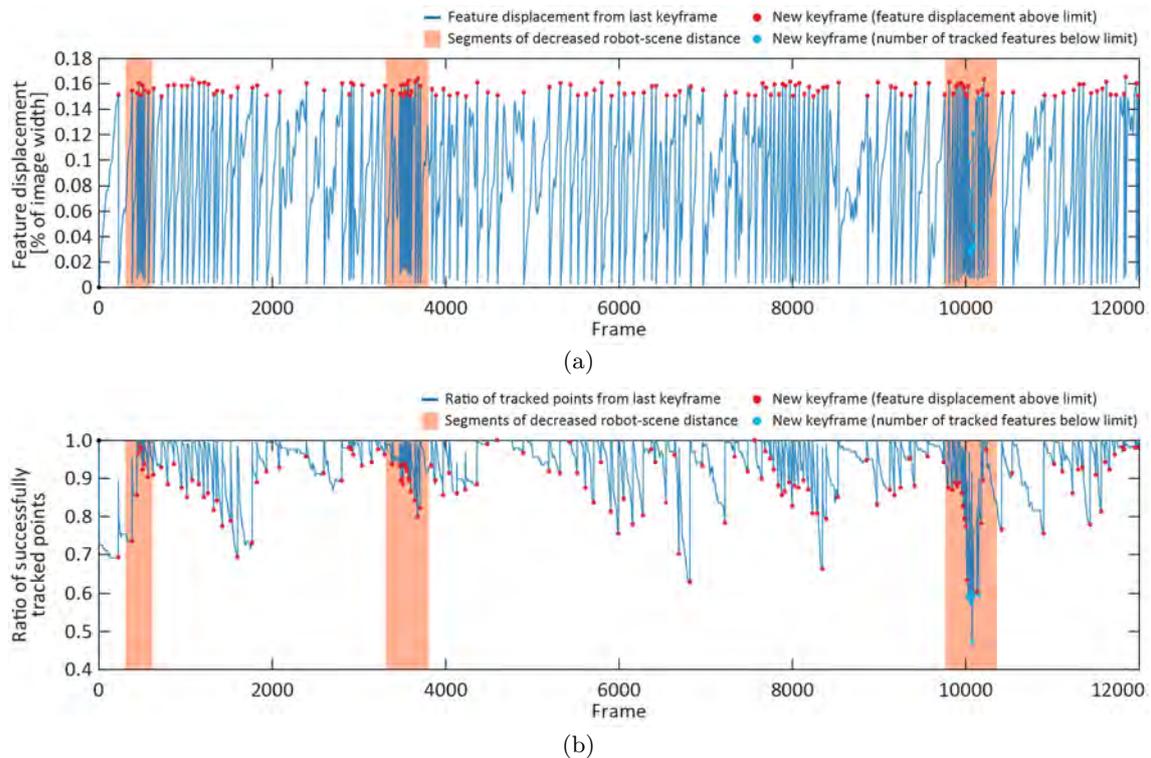

(a)

(b)

Figure 6.1: Feature displacement and ratio of tracked features between each frame and the last selected keyframe for a sequence in the *Punta del Molar* data set [11]. Red dots represent extraction of a new keyframe as a result of sufficient feature shift, cyan dots represent selection as a result of an excessively reduced number of tracked features, and orange areas mark the segments with reduced distances between the robot and the scene.

The advantages of the proposed selection of the best neighbor based on the blurriness score is demonstrated with an example, which occurred at frame 6787. At that moment, the system successfully detected a significant change in content and started collecting neighboring frames (i.e. frames that are very similar). In this particular case, it identified 54 consecutive frames and calculated their individual blurriness scores. The relative differences between them are shown in Figure 6.3a, with the best/worst frame highlighted in green/red and shown in Figures 6.3c and d. While the pair is visually similar, which is to be expected due to the way the frames are selected, the difference in sharpness can be determined by comparing their histograms of the magnitudes of the Laplacians shown in Figure 6.3b. It can be seen that the higher number of elements with larger magnitudes in the high quality image (green) indicates a sharper image.
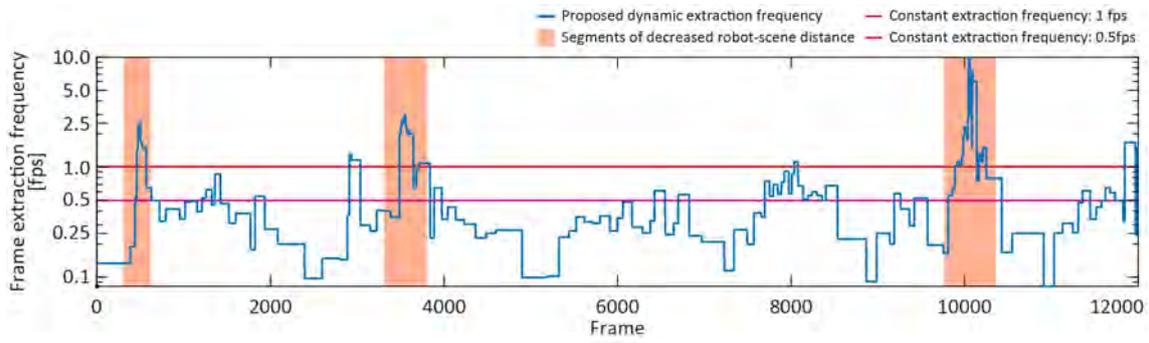
Figure 6.2: Proposed dynamic frame extraction frequency in the *Punta del Molar* data set with constant frame rates of 1/0.5 fps, highlighted in red and magenta.
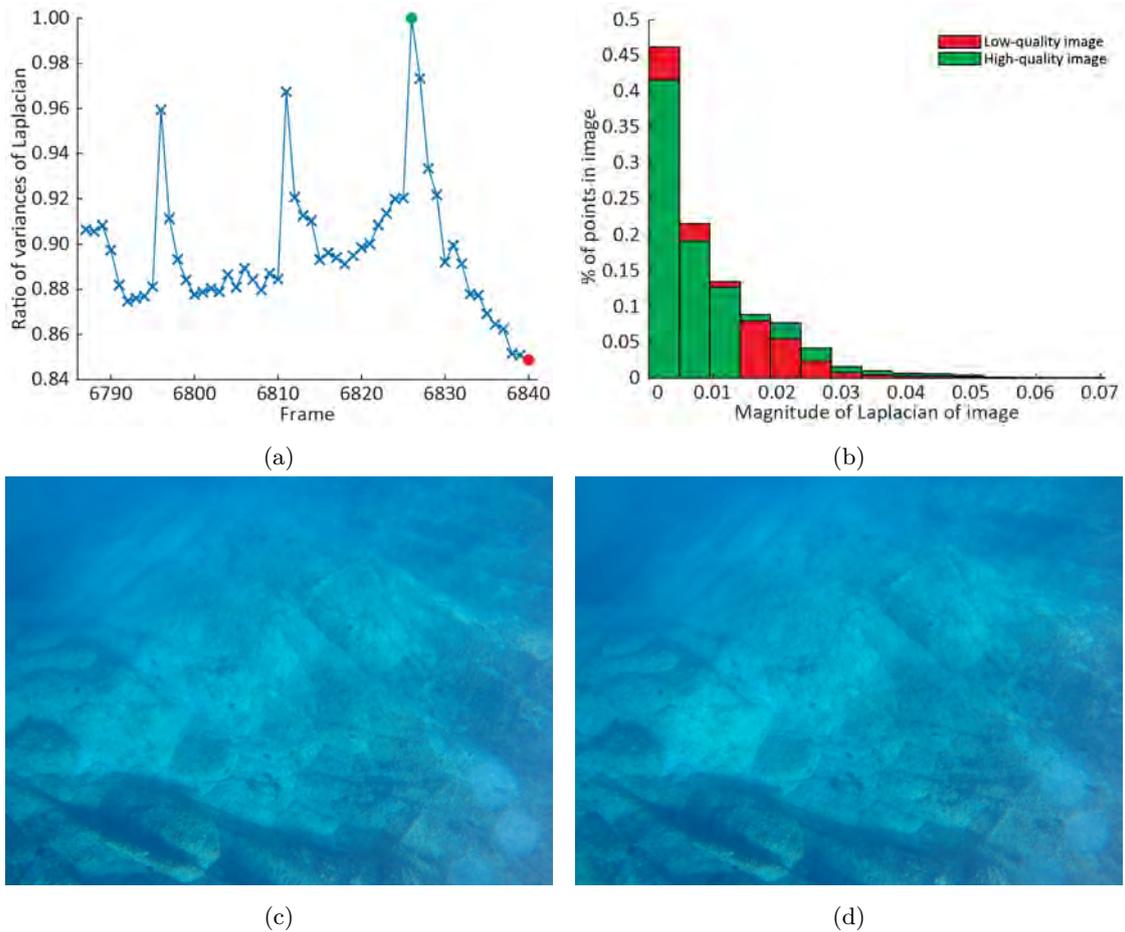

Figure 6.3: Selection of the best (sharpest) keyframe - frame 6787 in the *Punta del Molar* data set. (**a**) Individual blurriness scores for adjacent frames with marked best/worst score (green/red dot). (**b**) Histogram of magnitudes of Laplacians for marked best/worst frames. (**c,d**) The best/worst frame identified by the selection.

In the previous scenario, the dynamic determination of the frame extraction frequency proved to be extremely important for the selection of a high-quality image set. In the second selected video, collected during the *MIDAS* 2016 cruise, the robot (Seaeye Cougar XT ROV) was tele-operated and moved at a nearly constant speed and altitude, with a slightly tilted camera. While the resulting nearly constant frame extraction frequency (Figure 6.4) could easily have been replaced by a simpler method with constant-rate frame extraction, the conditions in the water (numerous suspended particles) caused the camera to constantly refocus, resulting in a multitude of blurred images. The use of such simple strategies could therefore possibly lead to the selection of many blurred frames despite the fact that sharper alternatives were available. Such an example, which occurred in frame 2757 of the *MIDAS D17* data set, is shown in Figure 6.5 with the same graphics as in the previous case. The change of the focus (i.e. increased blur) between frames is reflected in the computed blurriness scores (Figure 6.5a), the frames themselves (Figures 6.5c and 6.5d), and on the associated histograms (Figure 6.5b). Similar to the previous scenario, the selected sequence consisted of 6 min 40 s and the system identifies 229 sharpest keyframes, while 401/201 (random) frames would have been selected if a constant rate of 1/0.5 fps had been used.
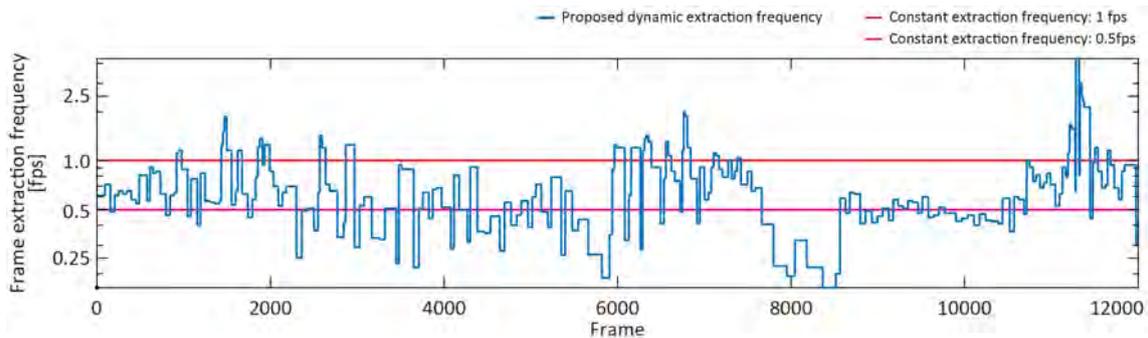

Figure 6.4: Proposed dynamic frame extraction rate in the *MIDAS D17* data set with constant frame rates of 1/0.5 fps, highlighted in red and magenta.

The results presented here (and in our publications) show that the proposed keyframe selection scheme enables an automatic determination of a dynamic frame extraction rate which directly depends on the movement of the camera and the survey speed. We do acknowledge a few shortcomings of the approach, namely the need for empirical selection of threshold values and the sensitivity of our blurriness score computation method to various factors (e.g. small illumination changes, new objects in the scene such as fish). However, even in those rare (worst) cases, in which the scheme might lead to a selection of an image with a slightly poorer quality, the scheme will intrinsically ensure that the selected images have sufficient overlap needed for the successful reconstruction, which is never ensured with the constant-time strategy. While a more complex strategy could easily be introduced to address those issues, the simplicity of the approach enables the rapid processing of a large number of frames.

6.3   Color Correction

The formulation of underwater images is a complex process that is influenced by a variety of effects that are typically negligible in air (e.g. uneven spatial illumination, color-
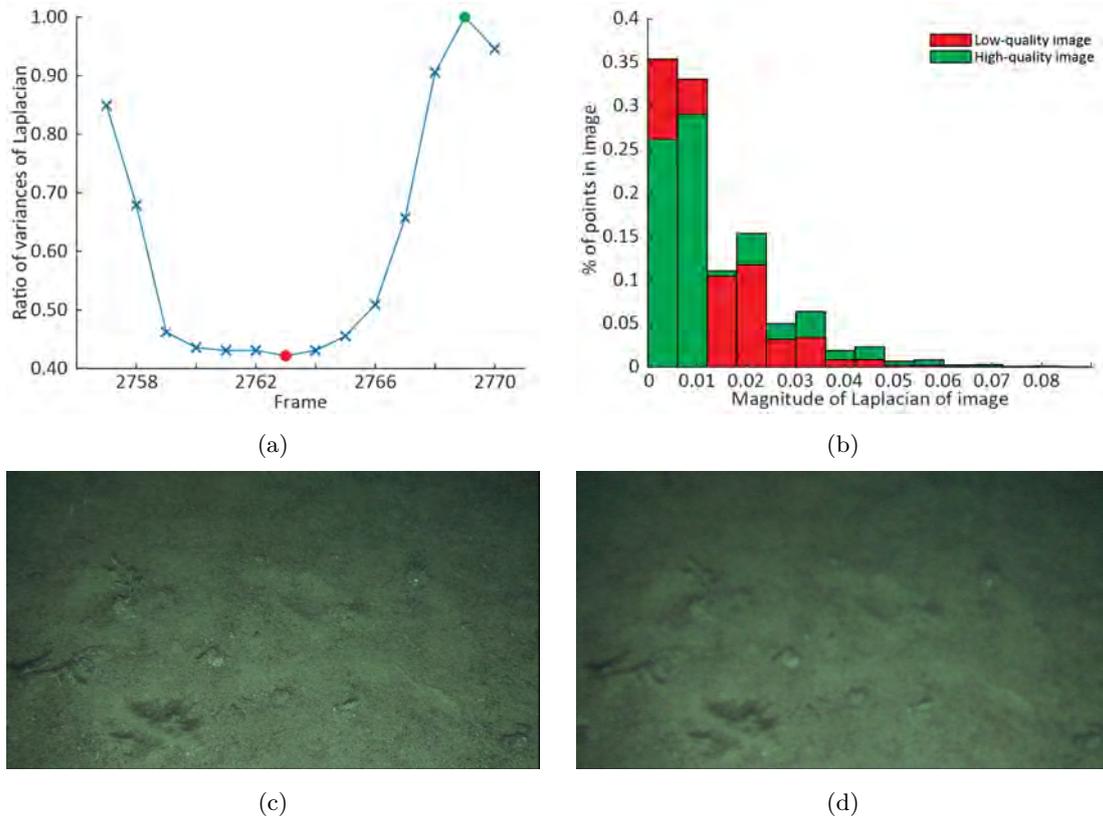
(a)   (b)

(c)   (d)

Figure 6.5: Selection of the best (sharpest) keyframe - frame 2757 in the *MIDAS D17* data set. (**a**) Individual blurriness scores for adjacent frames with marked best/worst (green/red dot). (**b**) Histogram of magnitudes of Laplacians for marked best/worst frames. (**c,d**) The best/worst frame identified by the selection.

dependent attenuation, backscatter, etc.). It often results in significant variance in color and brightness of the same area within the image set, due to images taken from different poses and distances between the camera and the scene. Therefore, the application of color correction not only benefits human perception and interpretation of image content, but also improves the quality and quantity of successful feature matches between image pairs used in the reconstruction process [144].

Dating back to the beginning of the Christian era, the ship incited the interest of archaeologists because it has different naval architecture from those known from the same time when it was discovered. In the process of documenting the structure of the vessel, the images were taken from various distances and views. This caused the attenuation of light to differ considerably both within a single image and in the entire image set, resulting in a highly variable set of images.

To illustrate the beneficial effects of using color correction as a pre-processing step for the reconstruction process, we present various intermediate results of individual reconstruction steps as well as the final textured 3D model obtained with data with and without the applied color correction. We performed the analysis using a challenging *Cap del Vol* data set, which was recorded by a diver while observing a Roman shipwreck located in the bay of Port de la Selva, Spain [145]. The ship, dating back to the beginning

of the Christian era, incited the interest of archaeologists because of its particular naval architecture. In the process of documenting the structure of the vessel, the images were taken from various distances and views. This caused the attenuation of light to differ considerably both within a single image and in the entire image set, resulting in a highly variable set of images. Some examples of original images observing a common area are shown in Figure 6.6(a-c), with color-corrected versions shown in (d-f). Although there are several sophisticated approaches that can robustly and accurately restore measurements of true color (reflectance), such methods without an exception require complex acquisition arrangements setups (e.g. [146]) or careful data acquisition (e.g. [147]). In our example, we have used the method of Ancuti *et al.* [148] because it requires only the selection of a reference image that encapsulates the distribution of expected resulting hues. Despite the use of this particular method, the observed effects of color correction on the reconstruction process can be assumed for almost any correction method, including novel methods based on advanced machine learning approaches (e.g. WaterGAN [149]), as long as they improve the visibility and perceptual uniformity of regions observed in multiple images.
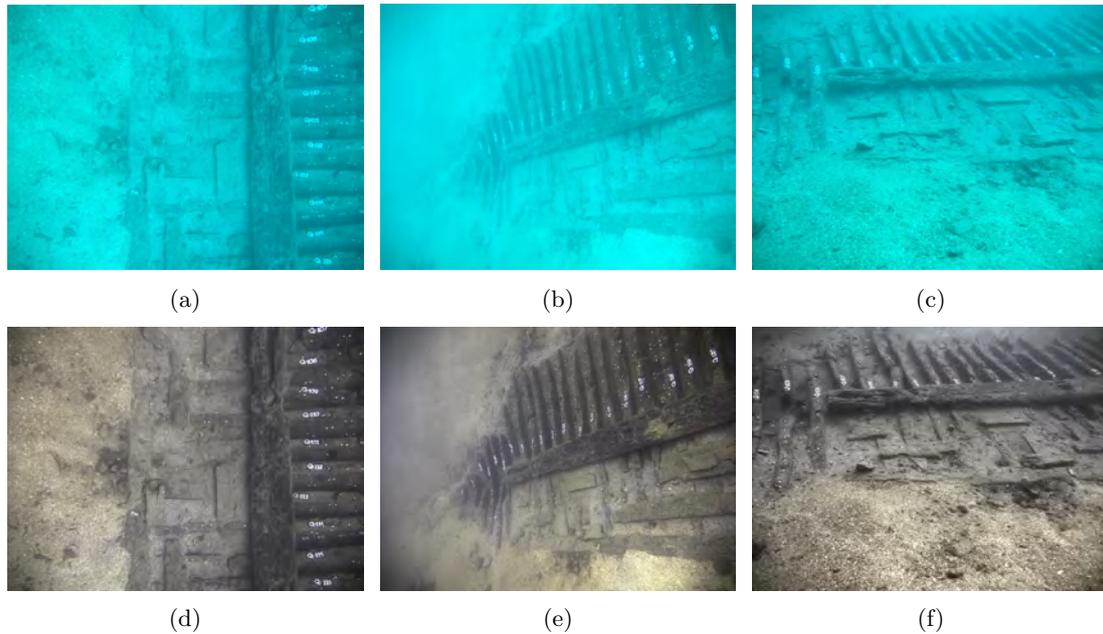

| (a) | (b) | (c) |
| (d) | (e) | (f) |

Figure 6.6: Sample images from the *Cap del Vol* data set: (**a-c**) original color, (**d-f**) corrected with the color correction method of Ancuti *et al.* [148].

The reconstruction process used in this analysis is described in detail in chapter 4 (section 2). To obtain comparable results between the two sets of images, we extracted the best 2000 features from each of the 549 images after reducing their size to 1160×868px. The overview and some detailed views of the two final textured models are shown in Figure 6.7. The reconstructions were scaled to metric units using two physical spatial measurements extracted from a drawing made by divers during the acquisition. The results show a significant benefit of the color correction procedure for the human interpretation of the models. While the original images suffer from a distinct blue color cast, the corrected images produce a model in which the texture of the wood is easier to see and where, among other things, the numbering plates and measurement chips are of correct black,
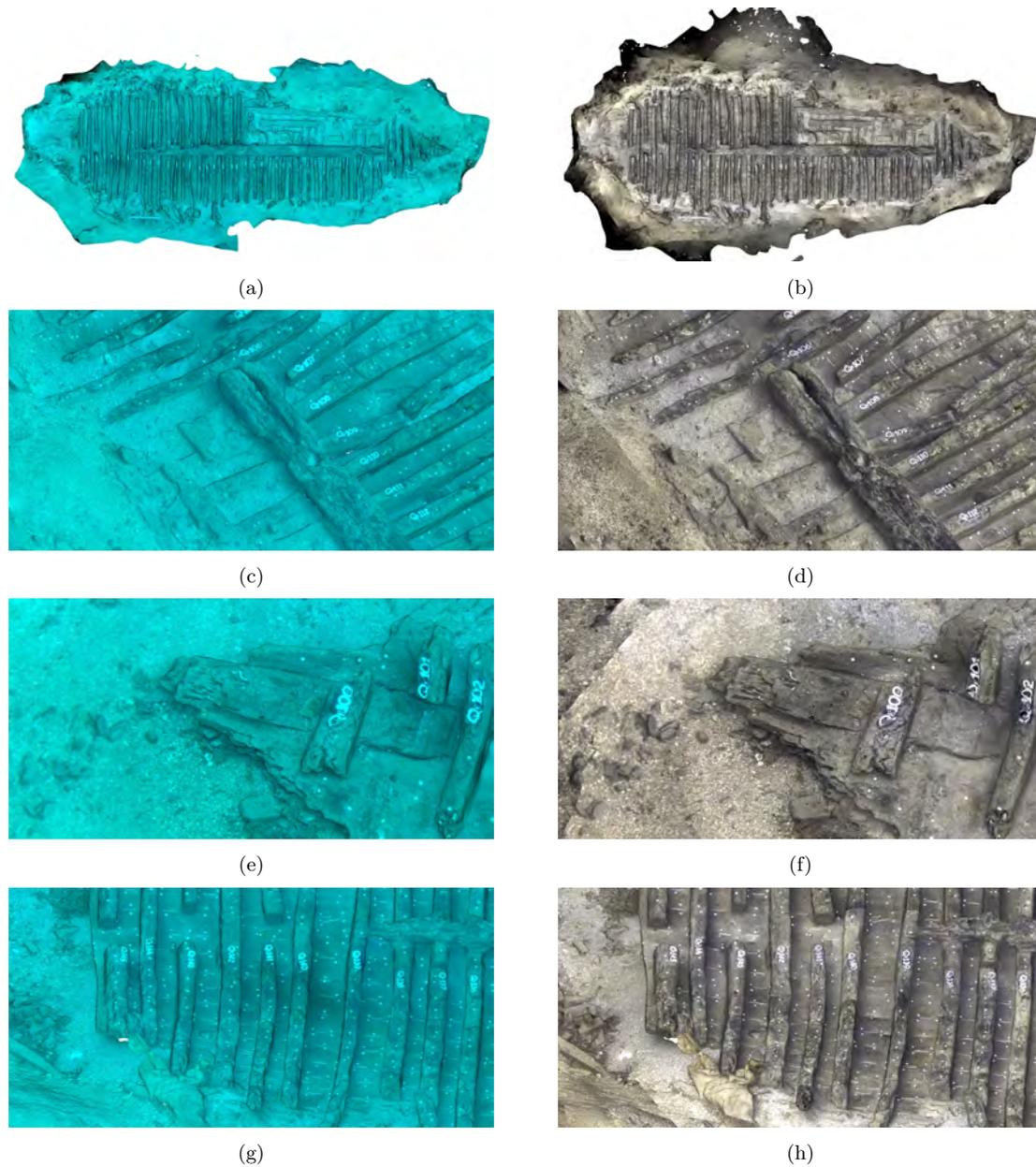
white and yellow color.

Figure 6.7: Textured 3D models of the *Cap del Vol* shipwreck created with our reconstruction pipeline: (**a-b**) top-down view with original images/corrected images, (**c,e,g**) detailed views of a model with original images, (**d,f,h**) detailed views of a model with color corrected images.

The statistics of the various intermediate results are summarized in Table 6.1. The performance of pairwise feature matching is characterized by the ratio between the number of matches obtained for each image pair using original images and color-corrected images. Values greater than 1 represent better performance on the original data set, while values closer to 0 (and less than 1) represent superiority of results with processed images. Since the comparison of individual image pairs is too content-dependent to draw definitive conclusions about the effects of color correction, the data was aggregated over the entire data

set.

To evaluate the possible impact of the color correction process on the reconstruction process, we compared not only the number of successful feature matches validated by the geometric filtering, but also the number of actual feature matches used in the reconstruction process, thus passing the numerous filters based on the re-projection errors. Figure 6.8 shows the distribution of the calculated ratios between the number of successful feature matches after geometric filtering in original and color-corrected image pairs (red columns) and the distribution of the ratios of those matches that passed all the re-projection error tests and were used in the final optimization of the 3D reconstruction.

The comparison suggests that although the matching within the original set results in a slightly higher number of matches (the average ratio is 1.11), such matches are likely to be less accurate and therefore more often rejected in the reconstruction process (the average ratio of successfully used matches in the reconstruction process is 0.86). The somewhat surprisingly larger number of geometric matches in the original set can be explained by the fact that these images are of lower contrast and therefore lead to less pronounced feature descriptors that more often match with imprecisely located partners.
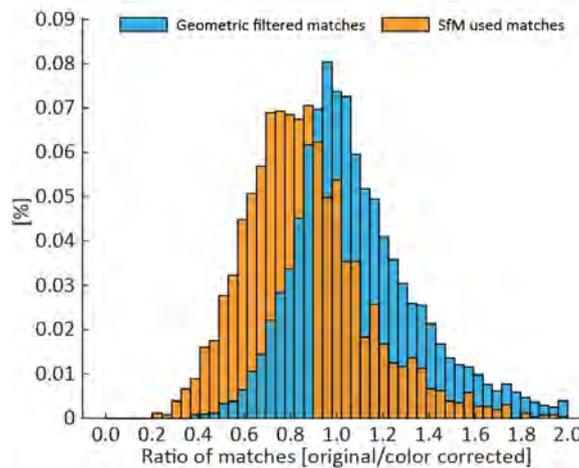

Figure 6.8: [Distribution of feature matches ratios after geometric filtering (blue) and matches used in the 3D reconstruction process (orange) of the *Cap del Vol* model.

Subsequently, and predictably, the larger number of quality feature pairs also directly affects the number of reconstructed 3D points in the sparse point cloud ($64, 683$ versus $73, 646$). The impact of the difference in the number of image-based constraints in the BA optimization was further evaluated by comparing the discrepancies between the locations of the estimated cameras in the two results. As shown in Figure 6.9, the location of the majority (93%) of the cameras is nearly identical (with the Euclidean distance smaller than $1\,\mathrm{cm}$), with only a few cameras having slight discrepancies reaching up to $5\,\mathrm{cm}$. When examining the locations of these cameras, it can be seen that all problematic cameras have nearby cameras farther away. The discrepancies can thus be attributed to the fact that it is more difficult to match features in such images because of changes in appearance. Since this effect is significantly reduced by the color correction process, it subsequently allows more matches and a better estimation of the poses of these images. While we acknowledge that the discrepancies found cannot be unambiguously attributed as errors of the model created with the original images, the significant difference in the number of 3D points

reconstructed and the lower root mean square error (RMSE) of the re-projection (0.32px versus 0.28px) suggests that the color corrected model can be considered as more accurate.
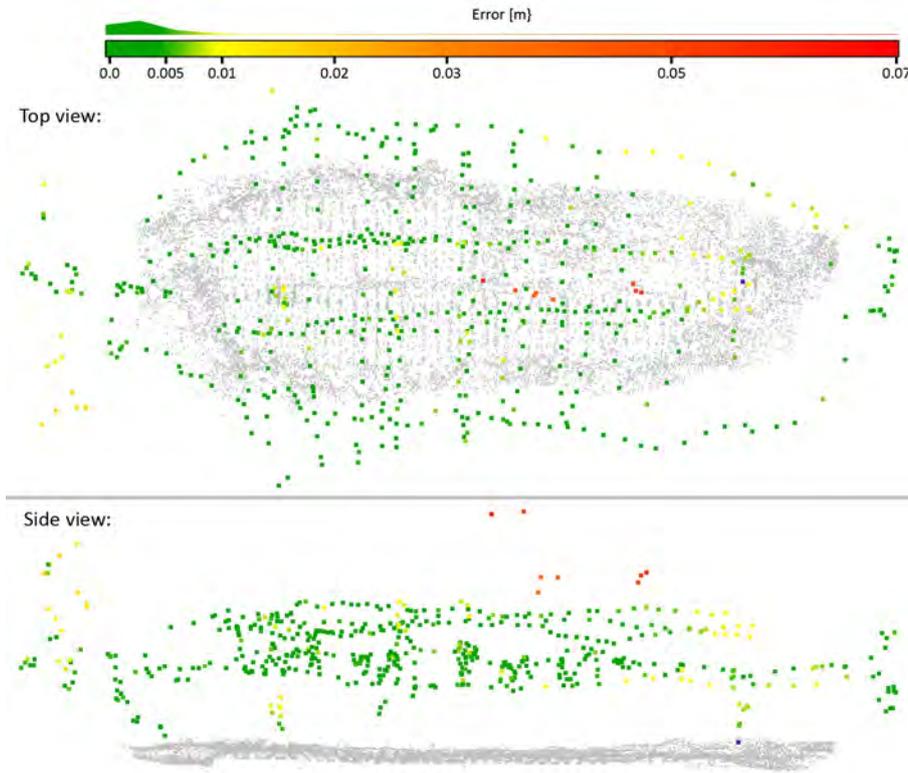

Figure 6.9: Discrepancies between estimated camera locations in the *Cap del Vol* data set: (**a**) top-down view and (**b**) side view.

The advantages of having a color corrected image set are also evident in the results of the densification process. While already starting with a larger number of initial seed points and better estimates of camera poses, the improved visibility of the texture and its uniform appearance in different images allows the densification process to recover a significantly higher number of points ($75,963,734$ versus $68,170,202$). This is also in part due to a larger covered area ($115.49\,\mathrm{m}^2$ versus $100.31\,\mathrm{m}^2$). The surfaces reconstructed from these dense point clouds are further compared and displayed in Figure 6.10 to reveal any geometric differences between the final 3D models. While the majority (88%) of the two models are within a $5\,\mathrm{mm}$ difference, there are two distinct areas where the discrepancies increase up to $1\,\mathrm{cm}$. These areas correspond to the regions observed by the cameras identified as displaced in Figure 6.9.

Although the performance improvement of applying a color correction procedure to an image set prior to the 3D reconstruction will vary depending on the data set [150], the presented analysis shows a variety of benefits. Any method that increases the ability of feature matching to obtain a greater number of accurate feature pairs, will subsequently lead to a better constrained optimization process, resulting in more accurate estimates of camera poses and a larger set of sparse 3D points describing the scene. Together with the increased visibility and uniformity of regions in the images, this not only allows the recovery of more detail and potentially larger areas, but also improves the result of the
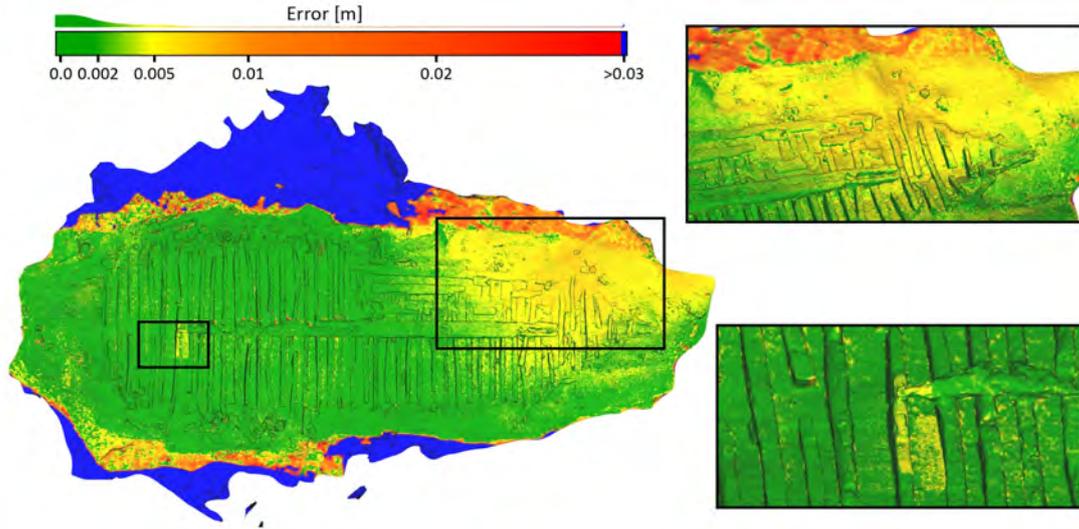
Figure 6.10: Discrepancies between 3D models created with the original and color corrected images from the *Cap del Vol* data set. Areas missing from the original model are highlighted in blue.

Table 6.1: Comparison of different intermediate results of reconstruction steps using original and color corrected image sets of the *Cap del Vol* data set.

	Original images	Color corrected images
Mean ratio of # geo. filtered feature matches		1.03
Mean ratio of # feature matches used in 3D		0.86
Re-projection error of SfM [px]	0.32	0.28
# of 3D points in SfM	64, 683	73, 646
# of 3D points after densification	68, 170, 202	75, 963, 734
Area covered [m^2]	100.31	115.49

texture mapping process.

6.4   SfM-based Scale Estimation Methods

To extend the use of laser scalers from simply providing an absolute size reference on photographs to an automatic scaling of SfM-based 3D reconstructions, we proposed two novel methods. The fully unconstrained method (FUM) and partially constrained method (PCM) were designed to cover a variety of frequently used laser configurations in real underwater scenarios. The FUM can be applied to any laser configuration with known geometric relationships between the camera and the lasers, while PCM addresses the scenarios where the lasers are equidistant from the camera center and parallel to each other, but not necessarily aligned with the camera axis. This relaxation of the required rigidity between the camera and the lasers significantly increases its potential applicability in real scenarios where strict rigidity is often unattainable.

Both methods are based on image localization and ray-casting techniques and exploit

the information obtained with an optical image beyond the sole position of the intersection of the laser beams with the scene (laser spots). As described in detail in chapter 3 (section 2), the camera pose (with respect to the 3D model) is obtained by a feature-based localization process which uses BA to minimize the re-projection error between known (and fixed) 3D points of the model and their 2D image observations. Once estimated, the pose is used together with the location of the laser spots on the images in a ray casting process to obtain the actual location of the 3D points onto which the lasers were projected. Finally, by compensating for the known geometry of the laser scalers, either the perpendicular distance between the estimated laser beams (PCM) or the displacement of the laser origins is used to infer the scale of the model.

In order to use the newly presented methods for scaling real models in quantitative studies (precise measurement of distances, volumes, orientations, etc.), the effects of the uncertainties of the input values on the final estimate of scale are further determined using a MC simulation, as presented in chapter 4 (section 3.2). The propagation is modeled by repeated computations of the same quantities, with the input values being statistically sampled on the basis of their probability distributions. The uncertainty of the imaging process and feature detection is characterized by the re-projection error obtained from the localization process, while the uncertainty of the laser calibration and laser spot detection is obtained from their respective processes, which are reviewed in the following sections and presented in detail in chapter 4 (sections 4 and 5.3).

In the results section of the chapter 3 we show that the proposed methods have several advantages over the existing methods. In order to validate the performance of the two methods and to compare them with the image scaling methods used so far, a synthetic data set based on a real 3D model of a hydrothermal vent was created. The laser and image information was generated as it would have been obtained from a number of scenarios. Three different laser configurations (i.e. aligned with the optical axis of the camera; parallel but not aligned with the optical axis; and freely oriented), a range of camera viewing angles (up to 40° both in pitch and roll relative to the perpendicular view of the camera on the surface) and at several camera-scene distances were considered.

The result of the tests proved that the methods are capable of accurately estimating the model scale independent of camera viewing angle, camera-scene distance or terrain roughness (chapter 3, section 3.2). The FUM provided accurate results under all tested circumstances, while the PCM was subject to only a slight error (2.9% in the most extreme case) due to the approximation used to determine the laser direction. By further evaluating the effect of this approximation on 10,000 randomly selected points (chapter 3, section 3.3), we have shown that the scaling error in PCM depends on the difference in camera-scene distances between the two points of the laser beam-scene intersection, and that this error decreases as the camera scene distance increases. Furthermore, we also show that due to the specificity of the algorithms used (robust pose estimation), noise and potential outliers in the feature detection and matching processes do not have a significant impact on the results. However, the noise induced in the position of laser spots had a direct influence on the estimates. Increases in the distance between the camera and the scene led to higher errors in scale estimation because the displacements are magnified with distance. Additionally, we compared the results of a single laser measurement with the average of multiple measurements within an image, and showed that such fusion further reduces the effects of noise.

The proposed methods were also evaluated using real data from the *SUBSAINTES*

cruise. While the non-alignment of the lasers with the optical axis of the camera prevented a comparison of the methods with previous image scaling approaches, the FUM and PCM were compared to the only other possible method - a simplistic direct 3D approach (assuming the equivalence of the Euclidean distance between the points of the laser intersections and the beams themselves). The comparison of the outcomes reported in chapter 4 (section 6.3.3), clearly shows the advantages of using FUM and potentially PCM over the simplistic 3D method.

To obtain accurate results, the laser spot locations first had to be corrected (undistorted) for the refraction effects caused by the density difference in the water-glass-air interface (between the camera sensor and the scene). Assuming that the optical axis of the camera is approximately perpendicular to the dome port, the primary refractive effects can be assumed to be radially symmetric around the principal point and thus corrected by the radial lens distortion component of the calibration parameters [151, 152]. In the reported experiments, the camera and light refraction at the air-port and port-water interfaces were modeled using a pinhole camera model with the 3rd degree radial distortion [153]. As described in chapter 4 (section 5), the intrinsic parameters of the camera were calculated using a standard calibration procedure [154] with the data collected at the beginning of the cruise. The optics and camera parameters were identical to those used later throughout the entire acquisition process. Although small asymmetric effects caused by misalignment between the optical axis and the housing port, and/or non-uniformities in the thickness or material of the housing are not incorporated into the model, the magnitude of the systematic error is generally negligible as it is below the noise level when measurements are taken in the same ranges as during the calibration process [155, 156]. Similarly, small changes may occur due to vibration, temperature variations, etc. However, these changes are considered too small to significantly affect the final result.

Our FUM approach was significantly less affected both by the deviation of the camera-scene angle from perpendicularity and by the violation of the requirement for surface flatness, which caused a much greater dispersion of the estimated values within each segment for the direct approach. The ability of the PCM to overcome different camera angles and scene roughness was also pointed out, although the results also clearly showed the negative effects of using laser pairs that violate the requirement for laser parallelism. The use of these measurements led to significantly larger deviations in the estimated scales and were directly affected by the distance between camera and scene. However, by limiting the data to parallel laser pairs, PCM produced comparable results indicating its possible use in numerous scenarios where the strict rigidity between camera and lasers was not maintained or not accurately determined (e.g. legacy data).

In addition to validating the ability and accuracy of our SfM-based scale estimation methods, we have showed in chapter 3 (section 3.5) that the methods can be used directly to estimate and then scale an arbitrarily scaled 3D model. Since our methods (FUM or PCM) estimate a scale with associated uncertainty for each individual image containing laser spots, a general scale estimate for the model must then be determined. In the results section in chapter 3, we show that such a general scale estimate can be obtained by averaging the individual scale estimates, taking into account their individual uncertainties. Since calculating and applying the scale correction in an *a posteriori* step forces a single scale correction for the entire model, any scale drifts within the model cannot be corrected in this way. Furthermore, if the model contains significant scale drifts, simple averaging would yield inaccurate results. In such cases, a more scenario-dependent selection of the

individual scale estimates used in a global scale estimate would need to be performed to obtain satisfactory results, or the calculation of the scale estimates themselves would need to be included in SfM process, rather than as an *a posteriori* step. We consider this as a possible future work and briefly comment on the implications of such an implementation in section 7.2

6.4.1   Usage of legacy data

Since the use of laser scales to provide scale references has been in use since the late 1980s [157, 158, 159], there is a significant amount of data (i.e. legacy data) that can potentially be used with our SfM-based scale estimation methods. Although until recently the acquired data was mostly used for measurements within single images, it is likely that the images with identifiable laser projections can be used to provide the scale information if the recorded sequence allows the reconstruction of a 3D model using an image-based approach. However, this limits its application to moving objects (e.g. fish, benthonic species, soft corals, etc.) or in highly dynamic environments. Apart from this, our methods are considered universal as they can be applied to standard image acquisitions and are not tied to data acquired with specific sensors or hardware (e.g. stereo cameras, underwater vehicle, color laser).

The ability to use legacy data with our scaling methods was one of the main reasons for developing PCM. While FUM requires full knowledge of both the origins and orientations of the lasers (with respect to the camera), the required knowledge is significantly reduced in PCM. Based on the fact that many (old and commercially available) laser scaler setups consist of at least two parallel lasers, PCM is designed to exploit this knowledge. The method is thus capable of determining scale, with the sole additional requirement that the laser origins be equidistant from the camera center (without the need for precise knowledge of their position). If these conditions are met, additional calibration of the laser-camera system is not required.

Furthermore, unlike previous image scaling methods, the lasers do not need to be aligned with the optical axis of the camera. This relaxes the constraints considerably and increases the applicability of our method to an even wider range of legacy data sets, as well as to many non-professional applications where laser-camera axis alignment is extremely rarely performed.

As demonstrated with the data collected during the SUBSAINTES cruise, even legacy data where the lasers are not exactly parallel can potentially be used. However, in such situations it is necessary that the position of the laser origins is known. Such scenarios often occur when the lasers are fixed individually (the laser directions may change independently) and the calibration process is not performed before each acquisition, or the accurate knowledge of this detailed information is lost over time. As shown in chapter 4 (section 5.3) and summarized in section 6.6, the missing information about the direction of the lasers can be recovered from the data itself as long as there is a sufficient number of laser intersections along a range of distances.

As with all applications of our approach, the location of the laser spots detected on the images must be corrected to mitigate the effects of refraction at the water-glass-air interface. Since legacy data often do not contain accurate camera intrinsic and distortion parameters, these parameters can be included in the BA and thus estimated together with the extrinsic camera parameters and the geometry of the scene. Although an accurate and deliberate calibration process is preferable, such joint estimation usually yields

incomparably better results than using inaccurate intrinsic information.

6.5   Laser Spot Detection Method

The scale estimation framework presented depends heavily on the ability of the system to reliably detect laser spots on an image and subsequently characterize the uncertainties of these detections. The apparent lack of such tools in the literature has been addressed in the work presented in chapter 4 (section 4). The proposed method differs from other approaches in that the point with the highest image intensity in a given image area is not directly considered as the center of the laser beam. Instead, it additionally considers the underlying scene texture and its influence on the resulting intensities. This is done by using an auxiliary image that is assumed to have been taken from a similar distance and with laser spots that either do not exist or are in a different position. The auxiliary image provides pixel-wise texture information at the positions of the laser spots through the process of alignment. This becomes especially important in scenarios where the brightness of the laser beam does not sufficiently overlay the texture of the scene due to the significant attenuation that occurs along its path. In such cases, the final intensity of the beam reaching the camera can be overly influenced by the texture at the point of impact, resulting in imprecise localization using methods based solely on image intensity.

The proposed approach for laser spot detection has a number of important novel aspects with respect to previously published approaches. The foremost is the ability to detect even *strongly attenuated* laser spots. Additionally it estimates the uncertainties of individual laser detections on an image by means of a MC simulation that accounts for the inherent camera noise in terms of image intensities. Each iteration of the simulation thus yields a set of independent detections whose uncertainties are described by a Gaussian distribution based on the recovered characteristic shapes of laser spots with radially decreasing intensities. If the same laser spot has been consistently detected over the majority of MC iterations, the merged distribution is determined using the Unscented Transform [160], or otherwise discarded if it is considered unstable.

The application of the methods to two data sets collected during *SUBSAINTES* cruise is reported in chapter 4 (sections 4 and 6). Due to the strong attenuation of the laser beams, our approach to mitigate the effects of scene texture contributed significantly to the fully automatic detection of over 1800 laser spots on approximately 580 pre-selected images. Although the method was developed for the *SUBSAINTES* data, the approach itself is not limited to a specific data set. However, since the method uses a helper image to improve the visibility of the laser spots, it is necessary for any data set to have such additional images. This requirement is easily met in video recordings, since even slight changes in camera pose will sufficiently alter the positions of the lasers, whereas in still image acquisition additional effort must be made to collect either images from slightly altered poses or in the absence of laser projections.

To demonstrate the applicability of our approach to other data sets, we present examples of laser detection on image data collected during *CALDERA* 2012 cruise [161]. Since the data was acquired long before our method was developed, it was collected without regard to the requirements of our laser detection method and is therefore a good representation of legacy data. The acquisition was done in video format, with the robot (Girona 500 AUV) in constant motion, implicitly fulfilling the requirement of availability of auxiliary images. The laser scaler consisted of two red lasers, which were clearly visible

throughout the mission. For the demonstration we chose a 5 min segment in which a robot makes several passes over an area at different distances. Various example frames are shown in Figure 6.11.
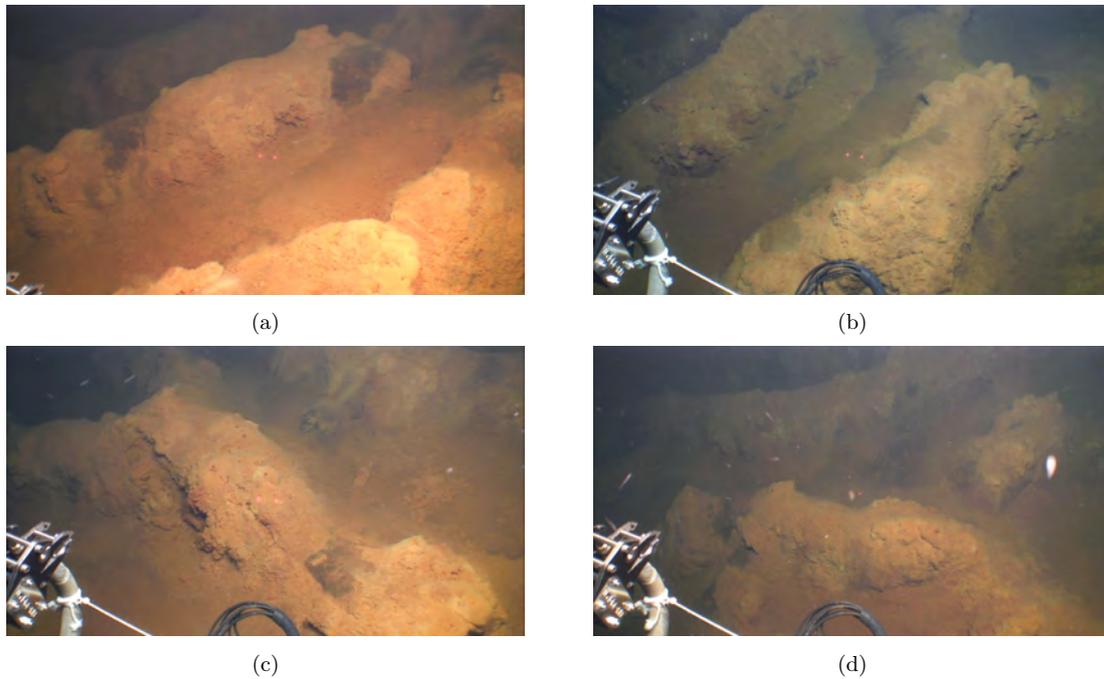

(a)   (b)

(c)   (d)

Figure 6.11: Sample frames from the *Caldera* data set: (**a-c**) frames successfully used in the detection process, (**d**) frame discarded due to incorrect detection (identified due to a significant distance from the expected ray line) caused by the presence of a shrimp or other fauna.

A total of 181 frames were selected (one image every two seconds) in which the two lasers were constantly present. An example of the successful detection of the two laser spots and the meaningful characterization of their uncertainties is shown in Figure 6.12. The positions of all detected laser spots in the sequence are further shown in Figure 6.13a together with the two fitted laser beams. Since the lasers did not move during the entire mission, all laser spots should lie on a laser ray line and thus any spot that deviates by more than 2px is considered an outlier (marked with X). Overall, in 153 frames (84.5%) successful and accurate detection of both laser spots was achieved, while in another 19 (10.5%) at least one laser spot was detected correctly. Only in 9 frames (5.0%) either no laser spot was detected or it was detected in a wrong place.

The missing and the falsely detected laser spots can be explained by the presence of shrimp or other suspended particles in the vicinity of the laser spots (Figure 6.11d). As a result, the method either detected a laser spot incorrectly or rejected a correct laser spot due to increased unreliability. However, the fact that only 11 of 336 (3.3%) laser spots were identified as outliers shows the robustness of our method. Furthermore, Figure 6.13b depicts the cumulative distribution of errors in the laser spot, measured as the perpendicular distance between the laser spot and the projected laser beam. The results show that there is no significant difference between the two lasers and that at least 90% of the detections are within a 1px error.
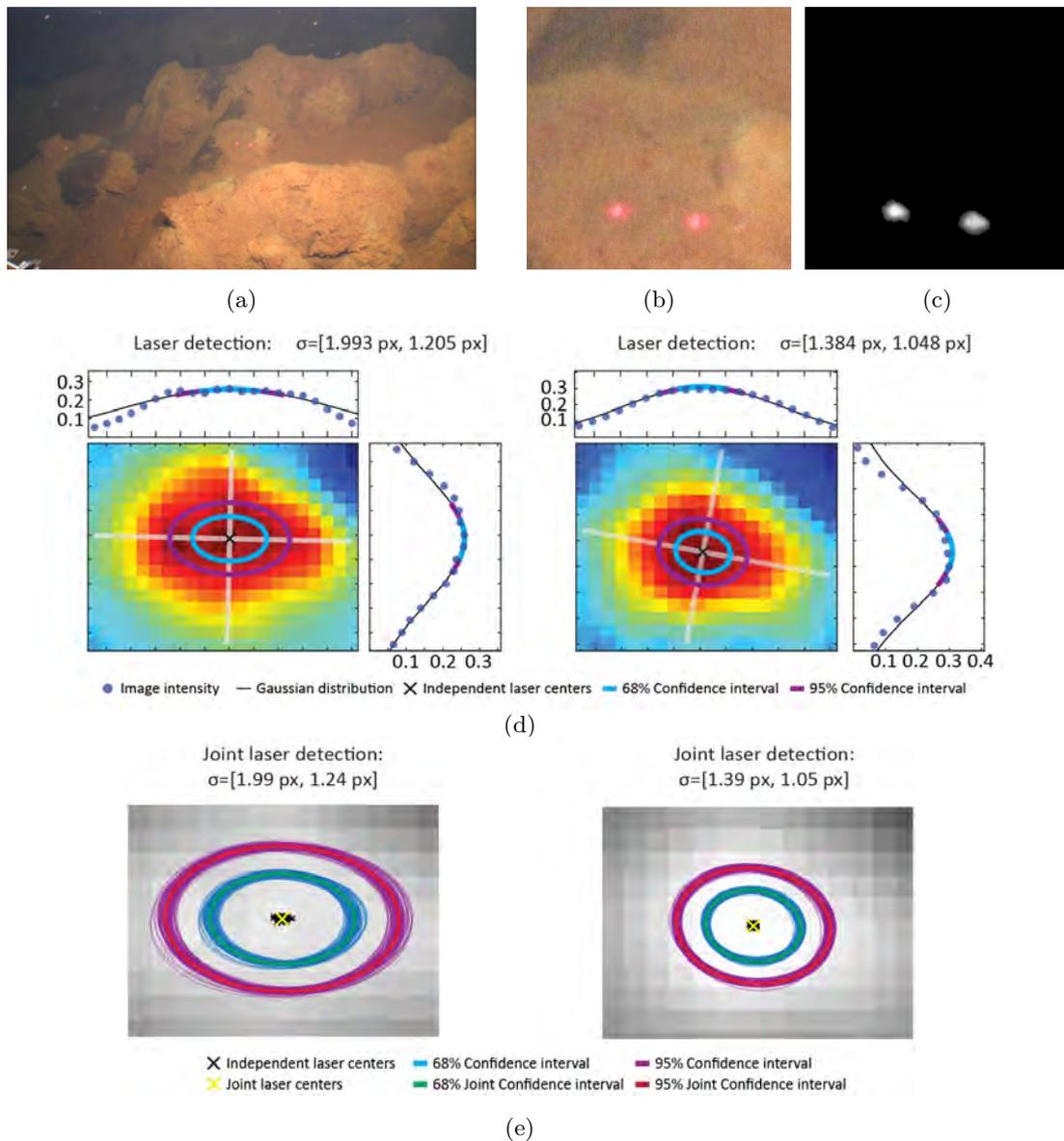
(a)　　　　　　　　　　　　(b)　　　　　　　　　　　(c)

(d)

(e)

Figure 6.12: Example of successful detection of laser spots in the *Caldera* sequence: (**a**) original frame, (**b**) predicted region of interest, (**c**) estimated laser beam brightness after texture removal, (**d**) detected laser spots in an individual MC iteration, (**e**) final detection result with uncertainty estimated by MC simulation. Individual detections are marked with black crosses and blue/violet ellipses for their uncertainties (68%/95% confidence level), and the final uncertainty estimate is shown in green and red, respectively.

6.6　Laser-Camera Calibration Process

Laser-camera calibration is a process of determining the relative geometry between the two devices that form our measurement system, i.e. the optical camera and the laser scaler. Depending on the method used in our scale estimation framework, either a complete geometry of the laser configuration must be known (i.e. origin and directions of each laser with respect to the camera center) or, in the case of the partially calibrated method, only
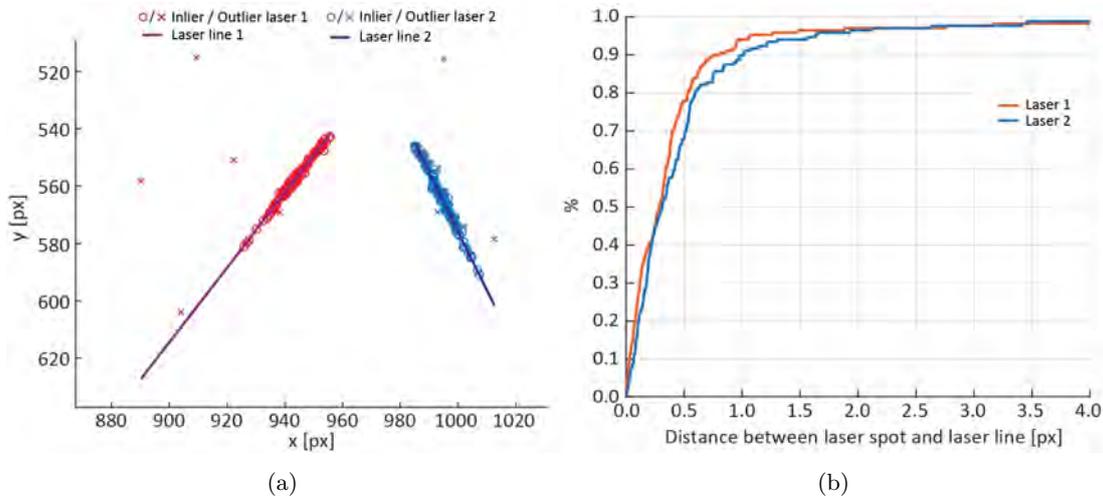
Figure 6.13: Detected laser spots in the *Caldera* sequence: (*a*) projected onto an image plane along with the estimated laser lines, and (*b*) cumulative distribution of detection errors, measured as the perpendicular distance between the laser spot and the projected laser beam for each of the two lasers.

the distance between the parallel lasers must be known (details in chapter 3 section 2). Normally, the calibration process consists of an initial acquisition of images containing clearly visible laser spots at a range of known or easily determined distances (e.g. using a checkerboard pattern), followed by a line-fitting process based on the 3D positions of the laser-surface intersections extracted from the aforementioned images. Ideally, it should be performed after each operation that could affect the geometry of the camera and the lasers, such as the servicing or replacement of equipment in the vehicle frame, but this is rarely possible in real-world scenarios.

In the case of the *SUBSAINTES* cruise, whose data we used for our scale error analysis, the evaluation images were collected during several dives separated by days, during which both the camera and the lasers were moved several times, and operational constraints on the ship prevented repeated acquisitions of calibration images. In this particular case, the laser-scaler mounting brackets ensured that the laser origins remained constant and only the laser directions with respect to the camera changed slightly each time the camera was installed. Given this is not an uncommon situation, we presented in chapter 4 (section 5.3) the approach we followed to obtain the missing and necessary information. Since the lasers were slightly misaligned, and not parallel, the fully unconstrained method had to be applied and thus its prerequisite need to know the complete geometry of the measurement system had to be fulfilled. Taking advantage of the facts that only the laser directions were changed between the dives and that the directions themselves do not encapsulate scale information, we were able to use images (with visible laser spots) collected during the missions themselves to estimate the required information. A set of 3D points lying on the laser beams were recovered by ray-casting using a reconstructed model of the scene and the locations of the laser spots detected on the images. As detailed in section 6.4, the laser spots first had to be corrected for refraction effects within the water-glass-air interface, which in our case was done using the 3rd degree radial distortion model [153]. The 3D points were then used to determine the laser directions by independent line fitting

methods. To maximize the conditioning of the line fitting, it was important to select and use the reconstructed model with the largest distance range of such intersections.

Although the evaluation data was collected over two distinct areas (AUTT28 and FPA), the requirement of having a constant scale limited the use of points in calibration to only those collected from the same area. Finally, the calibration was only performed with selected images from the AUTT28 area, as this provided the largest distance range of laser intersection points. The model used for ray casting was created using Global SfM and multiobjective BA, since the global nature of SfM and the internal fusion of navigation data reduces the possibility of scale drifts within the model most significantly.

Furthermore, due to the use of noisy laser spot positions, we have taken into account the inherent uncertainty of the resulting calibration by propagating the uncertainties of the laser spot through the calibration process by means of a MC simulation. The method thus provided us with the laser directions and the associated uncertainties, which, together with the *a priori* known origins of the lasers gave us all the information we needed to perform the scale estimation/error evaluation on the models.

In chapter 4 (section 5.3), we show that the calibration of the laser directions is independent of the 3D model used as well as of the associated model scale. We present the results depicting adequate fit between the projected laser line and the FPA related points (RMSE <0.8px) which were not used in the calibration and were affected by a completely different scaling factor. Since the lasers were moved between dives 654 and 658, the result shows two distinct sets of calibrations, and our assumption is thus justified with data from 2 different dives with 4 independent lasers. In addition, a similar fit to the vast majority of AUTT28 points (RMSE <0.6px) shows that the model used in the calibration did not exhibit any significant scale drift.

6.7   Scale Error Analysis

The lack of a systematic study of the performance of image-based 3D reconstruction and navigation fusion strategies on large scale models in GPS-denied environments (e.g. in deep-sea scenarios) was addressed by a comprehensive analysis presented in chapter 4. In this, we evaluated the four most commonly used approaches, which result from the combination of incremental/global SfM and the *a priori* (multiobjective BA) and *a posteriori* (similarity transformation) use of navigation data. The scale accuracies of the resulting reconstructions and the associated uncertainties were automatically estimated by our model evaluation framework and using numerous measurements obtained with our laser spot detection method previously mentioned.

The benefits and limitations of the evaluated strategies were analyzed on the basis of two distinct survey scenarios, which were carried out during the *SUBSAINTES 2017* cruise (doi: 10.17600/17001000). In the first survey (FPA data set) the data was collected in a single dive by several passes over the same area, resulting in a set of images with a significant overlap between non-consecutive images. In contrast, the second set (AUTT28 data set) consisted of two separate dives over the same area. The first of the two dives was a long, single pass dive, with the camera aligned almost parallel to a vertical wall. The additional dive subsequently recorded the data of a new horizontal segment and re-observed the areas already seen on the first dive, albeit at a closer range (see Figure 4.16). The selection of these sets allowed us to capture several real challenges in deep-sea fieldwork, such as combining different surveying strategies, multiple dives, viewing distances

and perspectives, and large outcrop lengths and sizes.

The models for each of the two data sets (FPA and AUTT28) were constructed using all four strategies and then evaluated on several segments distributed over the observed area. The accuracy of the models was analyzed using 148 images (432 lasers) grouped into 7 segments (FPA model) and 432 images with 1378 laser measurements grouped into 6 segments (AUTT28 model). The individual lasers were clustered into segments based on their proximity to ensure that the scale of the model did not vary within each segment. Because the estimates of each segment characterized the error in a specific area of the model, and because the segments were well distributed across the models (see Figures 4.24 and 4.25), we were able to determine not only the average scale error of each model, but also potential scale drifts within the models themselves.

As explained in detail in the results section of the chapter 4, the scale error analysis revealed that in surveys where the observed area is imaged with several overlapping passes (e.g. FPA), the accuracy does not differ significantly between different strategies or within the models themselves. The errors were determined to be within the range of -1% and -5% with uncertainties of about $\pm 3\%$. The presence of only minor differences in error between the segments/models can be attributed to the fact that the greater number of constraints between non-sequential images significantly limited potential deviations of the result regardless of the approach. As expected, the results also showed that the uncertainty of the estimates is closely related to the camera-scene distance, since small uncertainties in the laser direction lead to larger discrepancies at larger distances.

While the results of the FPA data set might indicate that the choice of reconstruction strategy is unimportant, the analysis of a more complex scenario (AUTT28) showed significant advantages of using navigation within the optimization using multiobjective BA - *a priori*. The models built with *a posteriori* navigation fusion suffered from broad-scale deformation (doming effect) due to the minimization of the re-projection error in combination with the inherent scale ambiguity of the two-view image pair geometry. The long single pass of the ROV, as performed in the first dive of the data set, introduced numerous successive two-view image pairs, magnifying the scale drift. As the results of the analysis show, the additional constraints introduced by the navigation data help to reduce ambiguity and ultimately virtually eliminate the scale drift. In addition, we have shown that the best strategy (global SfM with multiobjective BA) produced a total scale error of less than 1%, while the incremental SfM produced models with a scale drift of 2% along its 300 m length.

The analysis not only provided insight into the best possible reconstruction strategies, but also showed that the models created using the *SUBSAINTES* data have scale inaccuracies much smaller than 5%, with the majority being around 1%. The precise quantification of the model uncertainty has allowed the geologists from the Paris Institute of Earth Physics (IPGP) to use these models to study geological features associated with an earthquake by measuring the associated displacement along a fault rupture. In this way, they can extend their preliminary studies, previously carried out with a single model, with the information from more than 30 different sites along a $\sim 20\,\mathrm{km}$ transect. This collection of models generated from the *SUBSAINTES* data (using our pipeline) is, to date, considered one of the largest sets of image-derived underwater 3D models acquired by deep-sea vehicles for geological purposes (e.g. [9, 10, 15, 16, 17]).

6.8   Multi-survey Fusion

In real scenarios, several deployments over the same area are often necessary to obtain sufficient data for adequate 3D mapping. As we analyzed in chapter 4, it is important to consider the consequences of the merging when combining the data. If the scene has not changed significantly between the dives, the optical images can be easily combined, while special attention has to be devoted to the merging of the navigation data. Typically the available USBL data (and possibly pressure sensor data) have been smoothed and have usually expected errors in the acoustic positioning of ~1% or higher of the distance to the ship. This can lead to significant increases in nominal positioning error at greater depths. However, these errors do not represent the relative uncertainty between nearby poses, but rather a general bias of the collected data for a particular dive. Although they are constant within each dive, the errors will vary between dives in the same area and become especially problematic when data from multiple dives are combined. Models built with data from a single dive will therefore only be affected by a small error in georeferencing, while multi-survey optimization may have to deal with conflicting navigation priors; images taken from identical positions would have different acoustic positions, with offsets of the order of several meters or more. We have shown that the problem of merging multiple surveys can be solved by introducing an additional unknown parameter in the form of a 3D displacement vector for each subsequent dive, representing the difference between the USBL-induced offsets.

Here we present another example of such a multi-survey scenario, which was processed in 2016 for the purposes of the MIDAS project. During the industry-led field testing cruise - *MIDAS* 2016, an area of $150\,\mathrm{m}^2$ over the Condor Seamount in the Mid Atlantic Ridge was surveyed twice (in dives 17 and 26). The aim was to collect data for an experiment to detect changes before and after the deployment of steel structures (landers) for an unrelated scientific experiment. Both models, shown in Figures 6.14 and 6.15, were built using our initial reconstruction pipeline.
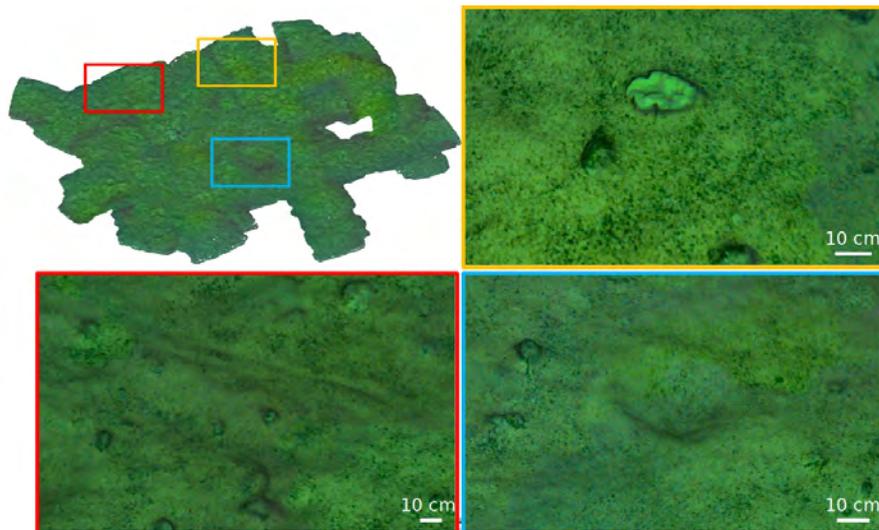
Figure 6.14: Textured 3D model reconstructed from the pre-impact survey (dive 17) during the *MIDAS* cruise.
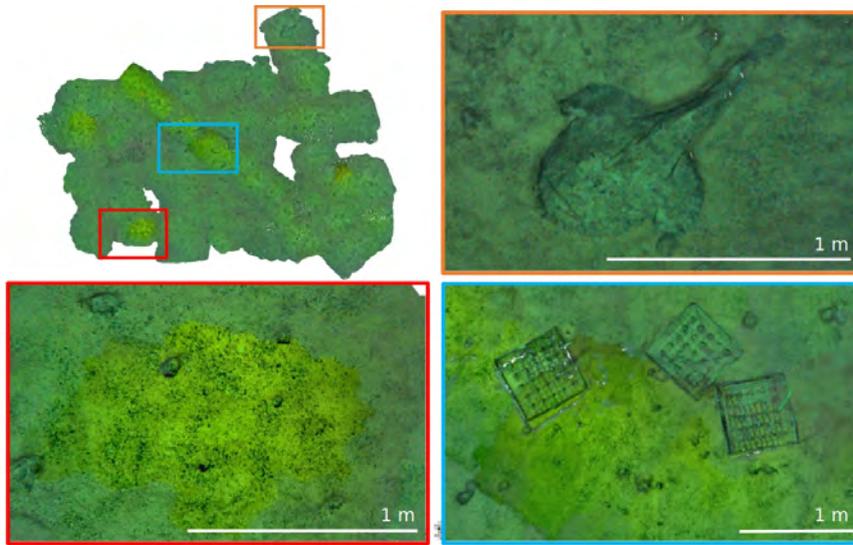
Figure 6.15: Textured 3D model reconstructed from post-impact survey (dive 26) during the *MIDAS* cruise.

Initially the models were built independently, using optical and navigation data for each respective dive. Although they appeared to be aligned as shown in Figure 6.16a, where the sparse point clouds reconstructed for models D17 (red) and D26 (blue) closely overlap, the alignment based on navigation data alone was not sufficiently consistent with the reconstructed scene (optical data). To obtain a better alignment, the model D26 was realigned based on the optical data from dive 17. By localizing images taken in dive 26 within the previously built model D17, the new realigned model D26 was effectively estimated based on the optical data acquired during dive 26 and indirectly with the navigation data from dive 17. Figure 6.16b shows the sparse point cloud of the new realigned reconstruction in green along with the original model D17 in blue. The change between the two models, namely the independently georeferenced model D17 and realigned model D26, is explicitly shown in Figure 6.16c with the same color coding. In the Figures 6.16d-f, the location of the cameras reconstructed in the three cases is also shown.

The constant shift between the locations of the two camera sets (independently georeferenced model D17 and realigned reconstruction D26) clearly indicate the existence of the bias mentioned above, which influences navigation priors. By estimating the displacement between the two sets, the average bias was estimated to be (2.09 m, 1.86 m, 0.015 m) in the x (E-W), y (N-S), and z (depth) directions, with the final misalignment of the cameras being 0.029 m ± 0.022 m. Similar to the *SUBSAINTES* data sets in chapter 4, the z offset is disproportionately smaller because the pressure sensor gives discrepancies between dives that are orders of magnitude smaller than the USBL positions.

The improved alignment subsequently enabled us to detect several geometric changes based on the 2.5D height maps and direct point cloud comparison. The 2.5D height maps were constructed by projecting points onto a grid of cells on a common plane, which were then compared between the models, while the direct point cloud comparison measured the Euclidean distance between the two closest points in the two models. From the results of the 2.5D height map comparison (Figure 6.17) we can see that we were able to detect changes caused by the deployment of the landers (red square in Figure 6.17 and top row in
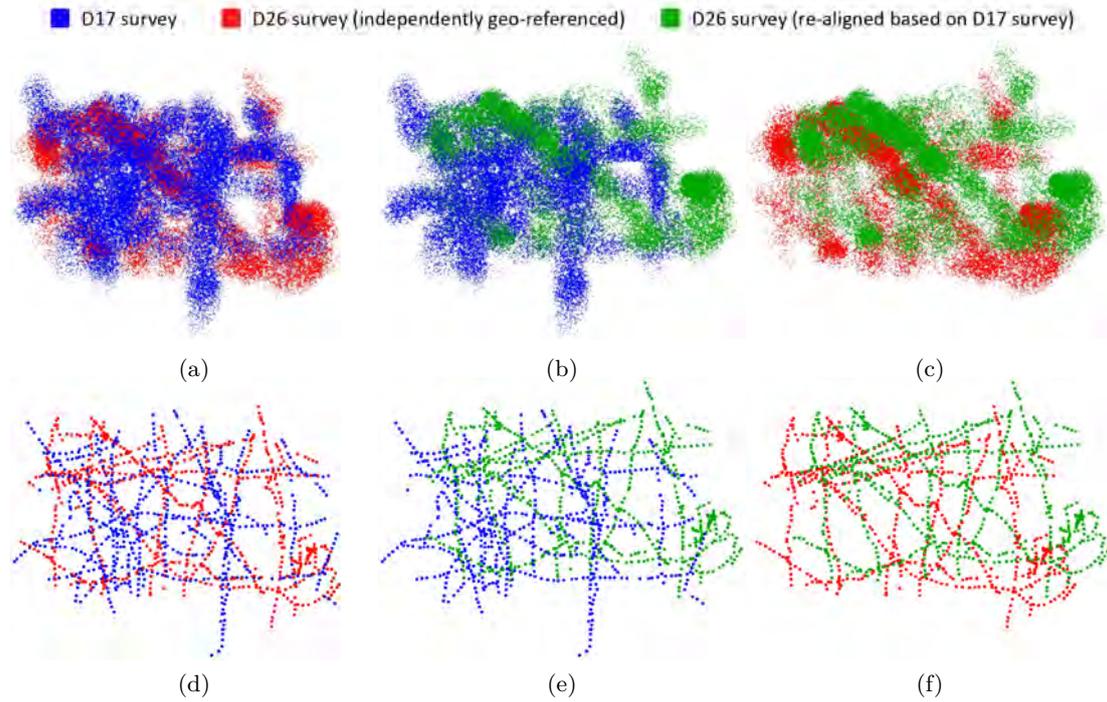
Figure 6.16: Reconstructed sparse point clouds (**a-c**) and camera locations (**d-f**) from: independently georeferenced model D17 (blue), independently georeferenced model D26 (red), and realigned model D26 based on visual data from D17 (green).

Figure 6.18) and the presence of new fish during post-impact acquisition (blue square in Figure 6.17 and bottom row in Figure 6.18). The results also show some problems, in the form of false indicators of change, which are due to the bending of the model (red squares in figure 6.17). This was especially pronounced at the corners due to uncertain navigation and sparsely overlapping image tracks, which did not provide sufficient constraints in the optimization process.

6.9 Mission-time 3D Reconstruction with Quality Estimation

One of the perennial problems of image-based underwater 3D reconstruction processes was partially tackled during my research stay at the Australian National University. The significant dependence of the outcome of the reconstruction process on the conditions and the strategy applied in the acquisition process has long limited its applicability in real-world scenarios.

Described in detail in the conference paper titled *Mission-time 3D reconstruction with quality estimation* given in chapter 5, the problematic decoupling of the offline reconstruction process was addressed by proposing a new reconstruction pipeline capable of robust and globally consistent processing of large-scale 3D reconstructions by using an incremental way of solving the BA problem. An incremental NLS solver implemented in the SLAM++ library was used to efficiently apply incremental updates to the BA solution, thus enabling high performance in sequential image processing, which is the predominant
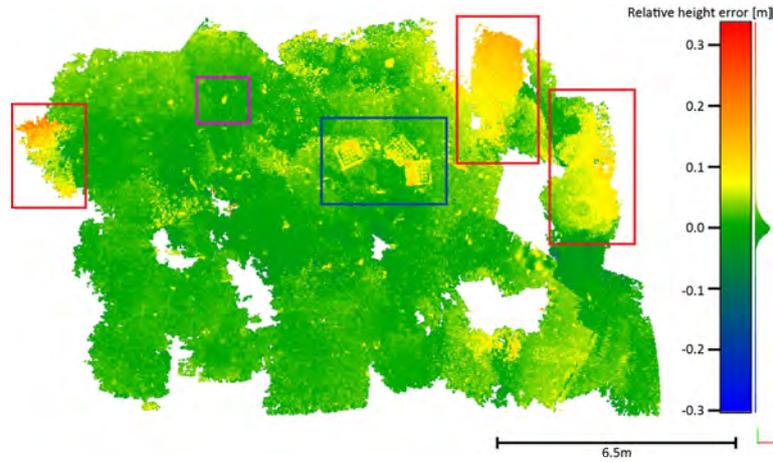
Figure 6.17: Discrepancies between calculated 2.5D height maps from pre-impact (D17) and post-impact (D26) surveys based on realigned 3D models. Examples of true changes are highlighted with magenta and blue rectangles, and false changes due to bending of the models are highlighted in red.
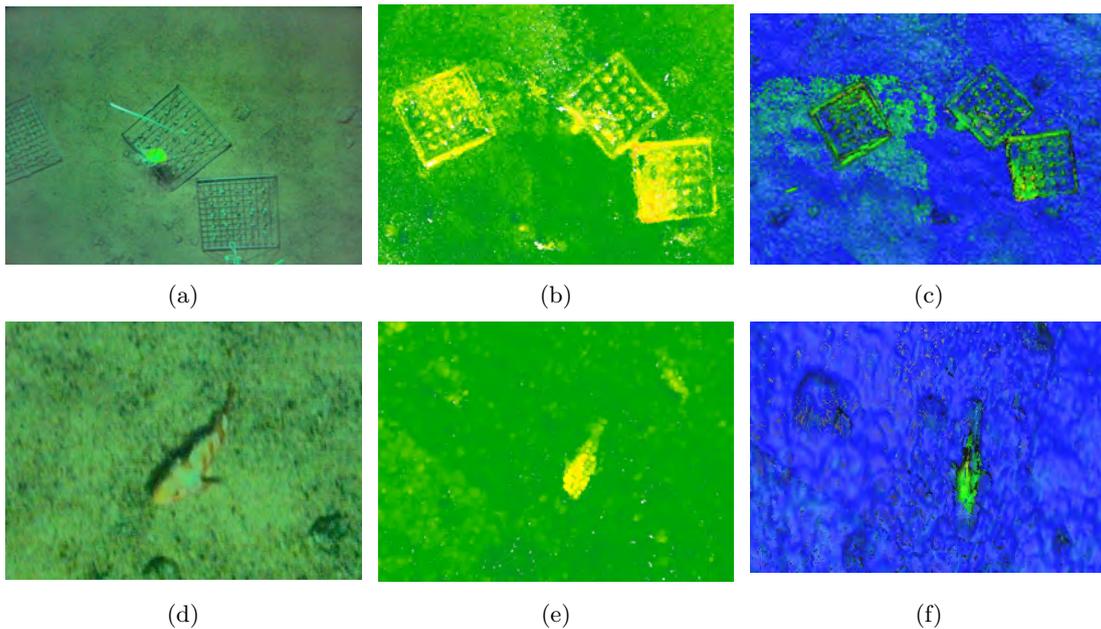

(a)	(b)	(c)
(d)	(e)	(f)

Figure 6.18: Two examples of detected changes between pre-impact (D17) and post-impact (D26) surveys based on the realigned 3D models (top row - blue highlighted area in figure 6.17; bottom row - magenta highlighted area in figure 6.17). (**a,d**) Optical images, (**b,e**) difference in 2.5D height maps, and (**c,f**) difference between point clouds, shown as colorcoded distance measured as Euclidean distance between the nearest points in the two point clouds.

nature of our target problem.

To reduce the generation of noisy and erroneous 3D points resulting from noisy underwater images, the system employs an outlier rejection strategy based on a dual-map approach. Each point is assigned to either a local or a global map, depending on the confi-

dence in the points (calculated from the number of successful observations). Local points can be reassigned and transferred to the global map once their confidence has increased sufficiently, or discarded as outliers upon a certain period of inactivity. By decoupling the two sets and constantly culling the local map, the tracking process is able to use a large set of points at any given time, while still operating within the computational limits of on-board processing. This significantly increases the robustness of the tracking process while ensuring that only well-observed points are included in the global optimization, thus further improving the quality and global consistency of the estimate at each step.

The BA, formulated as a probabilistic framework, provides not only the mean estimate, but also the uncertainty of camera poses and 3D points encoded within the marginal covariances. This normally prohibitively expensive procedure is possible by using a fast recovery covariance method [6]. In it, the key elements of the covariance matrix are calculated by identifying the variables affected by each update and by exploiting the algebraic manipulations of the operations within the computations to significantly reduce the computational time required.

Some methods underlying our system (i.e. incremental NLS solver and fast covariance recovery) were proposed by Dr. Ila and Dr. Polak during my research stay at the ANU and are part of the SLAM ++ library. As co-author of the paper ([6]) describing the methods published at International Conference on 3D Vision - 3DV 2017 (Qingdao, China), my main task was to compare the proposed methods with the implementation available in Ceres and to create and process the large-scale underwater data set used in the analysis. The paper mentioned above was given the Best Paper Honorable Mention Award.

In the paper presented in chapter 5 we have made an analysis of the proposed system as a whole and of the individual methods for incremental processing and covariance recovery. Using a large-scale underwater data set of inspection of a shipwreck, we demonstrated that our proposed SfM system is capable of successfully reconstructing a scene by applying a dual-map filtering approach, retaining only informative points and images in the global representation, while maintaining its ability to perform robust tracking under limited computational resources. This was demonstrated by the ability of the system to maintain an approximately constant size for the local map by removing inactive points throughout the mission while continuously adding newly observed points to the global map. The display of recovered covariance also showed its usefulness in identifying regions of the reconstruction that had been poorly observed. The performance of our incremental NLS solver and our method for fast covariance recovery was also compared to two other widely used approaches (g2o [162] and Ceres [163]). The results proved that they not only outperform the others in terms of computation time with similar accuracy, but that they are the only methods that allow estimating the uncertainties in the targeted mission-time.

6.10   Various results

The fulfillment of one of the main objectives of this thesis, namely the creation of tools and strategies for accurate image-based 3D reconstructions that can be used for scientific purposes in underwater scenarios, was also demonstrated by the numerous 3D models built in different application fields and for various purposes. In this section we provide an overview of the most significant projects and models from various collaborations with researchers in natural sciences, in robotics and in computer vision.

Subsaintes cruise

As previously mentioned, a significant part of the work presented in this thesis was made possible by the close collaboration with Dr. Javier Escartín and his research team at the IPGP, as well as with other research groups participating in the SUBSAINTES 2017 cruise (doi: 10.17600/17001000) and in the subsequent project ANR SERSURF (ANR-17-CE31-0020, France). Our participation in the project focused on the acquisition and subsequent processing of optical data with the ROV VICTOR 6000, with emphasis on 3D reconstruction. As described in detail in the dissertation, the obvious lack of a systematic accuracy assessment of larger-scale 3D models built in GPS-denied environments was addressed. Our work allowed us to precisely quantify the scale accuracy (and the associated uncertainties) of reconstructed 3D models recorded at more than 30 different sites along a $\sim 20\,$km transect at the base of the fault scarp. This now enables Dr. Escartín and other researchers to systematically use the models to measure the detected displacements, fractures and other deformations of the fault as well as study links between erosion/sedimentation and seismicity.

While the two most prominent models were already shown in the chapter 4, we built a total of 36 3D models using images acquired in 11 different dives conducted over 20 days at depths between 500 and 1200 m. Of these, 28 models were based on the data from a single dive, 7 used information from two dives, and one model was created by merging 3 dives that spanned over 7 days. The models vary in length between $\sim 10\,$m and $\sim 300\,$m horizontally, and extend vertically up to $\sim 30\,$m. The geological studies, which were partly made possible by the existence of these 3D models, have now been presented at the world's largest international Earth and space science meeting - the American Geophysical Union Fall Meeting in 2018 ([9, 10]), as well as at the general assembly of the European Geosciences Union in 2019 ([15, 16]).

As part of the ongoing cooperation, we were also involved in the development of a VR application aimed at recreating the experience of classic onshore survey work for the training of marine geologists. As explained in detail in the conference paper titled *Performing submarine field survey without scuba gear using geographic information system (GIS)-like mapping in a Virtual Reality environment* [17], we tried to mitigate the limited physical and visual capabilities of ROVs by combining the visualization of SfM-based 3D models with other digital elevation models in an immersive 3D environment. The VR system not only allows a good estimation of the scale and geometry of the objects under investigation, but also enables the users to perform actions on the underwater outcrops similar to those normally performed during geomorphological or geological fieldwork, as shown in Figure 6.19.

Boreas shipwreck

Throughout the thesis, one of the closest and most rewarding collaborations came from joint work with fellow researchers from the Underwater Vision and Robotics Center, who led the development of the omnidirectional camera. The camera, which was designed and built in-house, was extensively used in numerous projects as an acquisition device for optical images used in 3D reconstruction processes. Apart from these projects, some of which are presented in the following sections, the focus of the joint research was on overcoming the challenges in the creation of omnidirectional video. Since the camera is made up of several individual GoPro Hero4 Black cameras, each of which is strategically
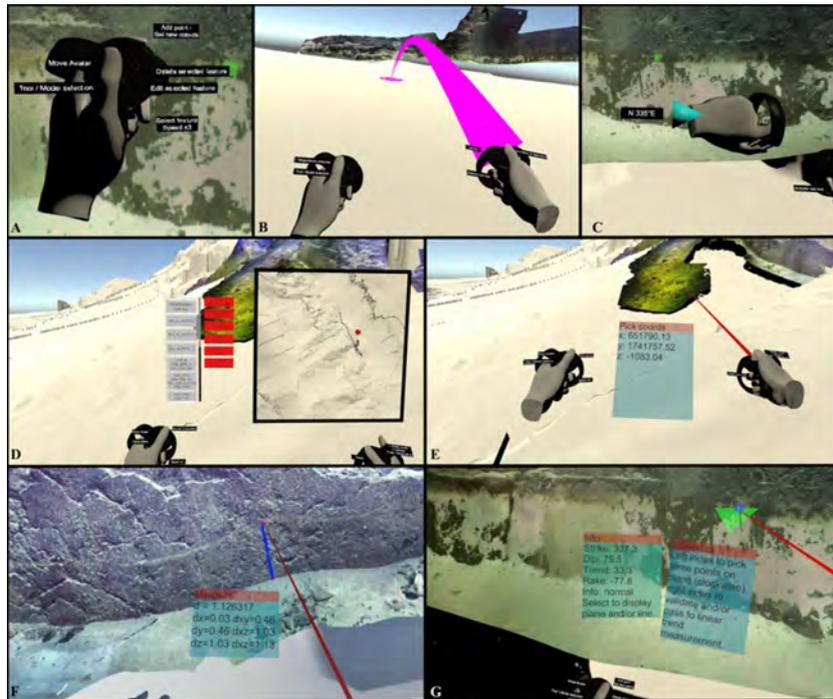
Figure 6.19: Overview of the interactive tools within the developed VR environment: (A) controlling the display, (B) teleportation, (C) compass, (D) localization, (E) coordinate selection, (F) ruler, and (G) plane and line measurement tool. Figure taken from Billant *et al.* [17]

oriented in a different direction, the images must be stitched together in a post-processing step to create a unified panorama as if it was shot from a single location, i.e. without any visible alignment errors.

As described in detail in the publication titled *Omnidirectional Multicamera Video Stitching Using Depth Maps* [12], we developed novel online and offline stitching methods that make use of the available information about both the multi-camera system and the observed environment. The online process uses sparse samples of the environment obtained through the simultaneous localization and mapping process, while the (offline) process of generating high-quality panoramic images requires a prior reconstruction of the scene in the form of a 3D model/depth maps.

The individual cameras in our custom-built system are not electronically synchronized. As such, only an approximate synchronization can be achieved, with an accuracy in the order of the frame duration. This approximate synchronization did not cause any significant problems in the creation of 360°videos. However it prevented the direct use of known rigid body transformations between the cameras in the reconstruction process. In order to create a correctly scaled 3D model, which is needed for the proposed offline stitching method, we proposed a bi-objective optimization where the known relative transformations are softly enforced by introducing an additional cost term. The new term penalizes large deviations in the relative transformations calculated from the newly estimated poses and their *a priori* known values. Since approximate synchronization dictates that the difference in acquisition time between cameras should be less than the elapsed time between two frames, we can reasonably assume that these differences in transformations should be small. Thus, the bi-objective optimization simultaneously compensates for the lack of

synchronization and ensures that the reconstructed model is scaled correctly.

This approach was applied to the data set collected in a survey over a World War II shipwreck that had sunk near the coast of Palamós, Spain, named *Boreas*. The 40 m-long wreck is a well-known attraction for recreational divers and lies between 25 and 30 m below the surface of the sea. The omnidirectional camera was mounted on the Sparus II AUV (Figure 6.20 and was guided by the diver to ensure safety when passing through narrow passages and tight cargo holds like those shown in Figure 6.21.
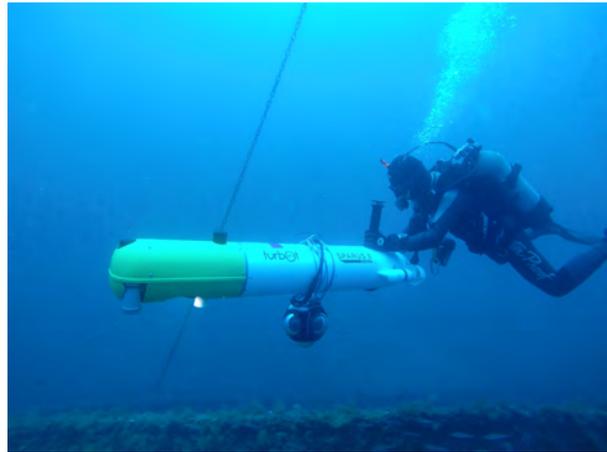
Figure 6.20: Sparus II AUV equipped with the omnidirectional camera while surveying over the *Boreas* shipwreck.
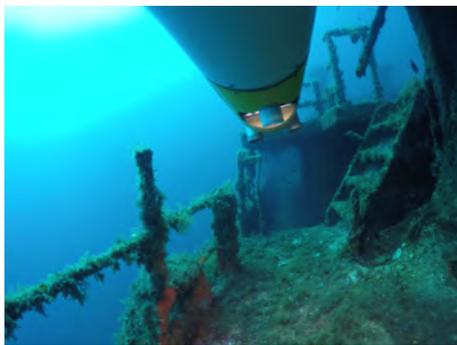
(a)   (b)

Figure 6.21: Example images from the forward-facing camera of the omnidirectional system at: (**a**) narrow passage and (**b**) the entry of the narrow cargo hold.

The advantages of using an omnidirectional camera for 3D reconstruction in such diverse terrain are evident. By selecting a total of $3,595$ individual images from 5 cameras acquired in only about 20 min we were able to produce a textured 3D model shown in Figure 6.22. The ability to capture images in all directions enabled us to obtain sufficient data for reconstruction even with just a single pass over narrower areas that would normally require several passes and would significantly extend the capture time. The large number of images also allowed us to successfully reconstruct some challenging parts of the ship, such as the cargo hold, as shown in Figure 6.23.
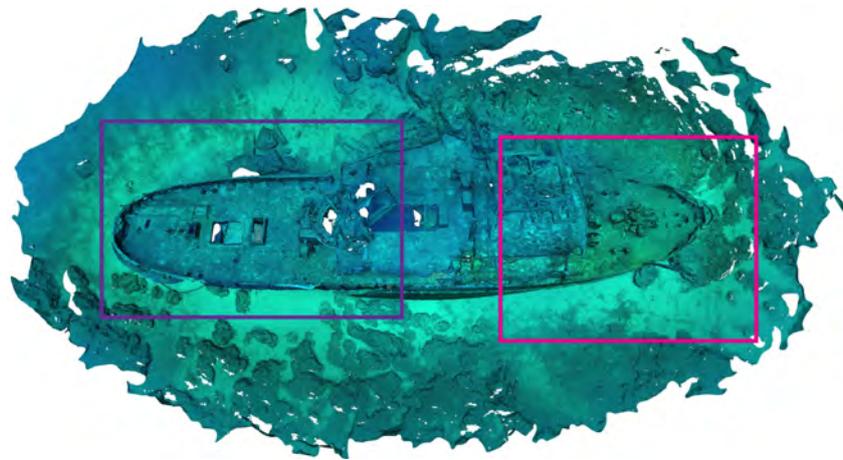
(a) Top-down view
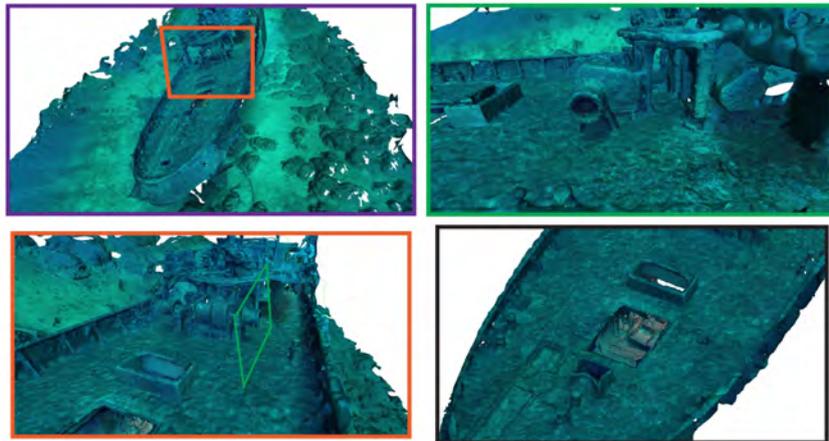
(b) Stern
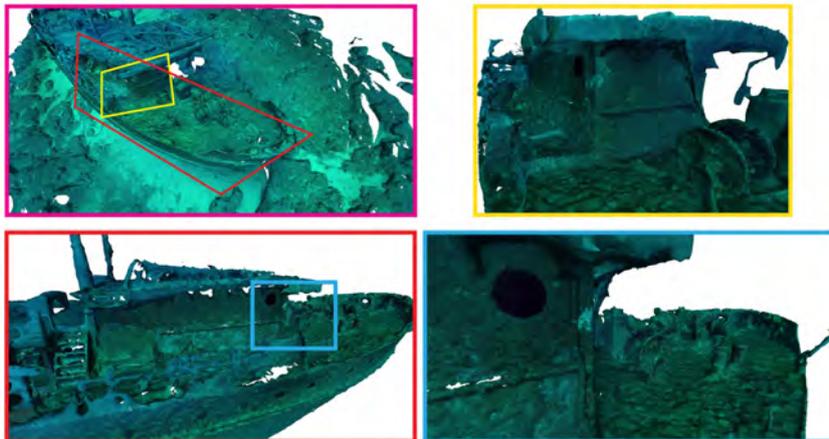
(c) Bow

Figure 6.22: Textured 3D model of the *Boreas* shipwreck created with omnidirectional imagery (top view with a few magnified details).

Figure 6.23: Side view of a fused dense/textured 3D model of the *Boreas* shipwreck with highlighted successfully reconstructed cargo hold area.

In the context of omnidirectional video generation, the reconstructed model enabled us to create equirectangular depth maps (Figure 6.24a), which are required for high-resolution stitching in post-processing. The fact that the direct projections of the images onto a panorama are precise (figure 6.24b) indicates that our depth maps are correct and, our proposed approach with soft enforcement of the transformations worked correctly, since the accuracy of the scale of the model is an important factor for the accuracy of the equirectangular depth maps.
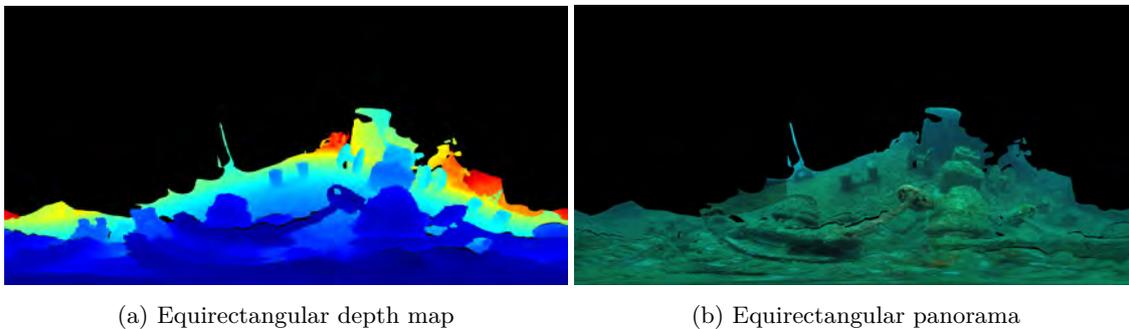
(a) Equirectangular depth map

(b) Equirectangular panorama

Figure 6.24: Equirectangular depth map and panorama generated according to individual depth maps extracted from 3D model of *Boreas* shipwreck. Black regions correspond to parts not reconstructed on the model or regions without observations from the individual cameras.

While the process and details of creating the final panoramas are outside the scope of this thesis, we nonetheless present the advantages of using the proposed (offline) stitching method compared to the usual spherical-world approach. The results are taken from our joint publication [12]. To keep the transition between the cameras visible in the panoramas, the presented images were not compensated for the individual exposure and gain settings of the cameras, except for the final result in Figure 6.26. The panorama created with the spherical-world assumption is shown in Figure 6.25a. In the highlighted areas, the misalignment is clearly visible at the boundaries of the contributions of different cameras, since the incorrectly assumed distances to the objects manifest themselves as errors in

the projection process. By better estimating these distances, e.g. by using a 3D model in our approach, the misalignments become hardly visible, as shown in Figure 6.25b. The finished, high-quality, fully color-blended panorama, on which independent camera contributions are imperceptible, can be seen in figure 6.26.
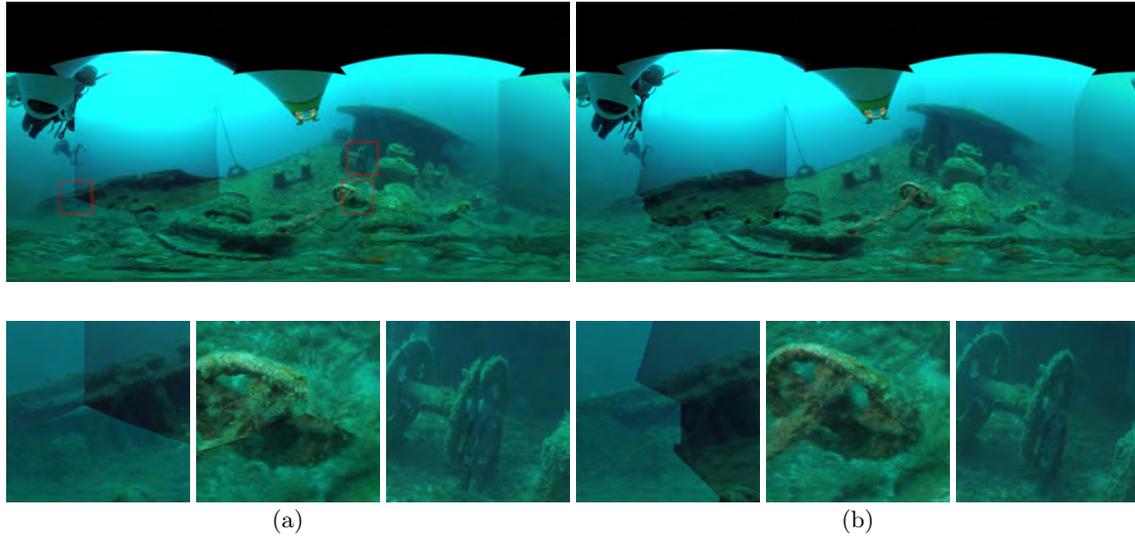

(a)   (b)

Figure 6.25: Equirectangular panorama generated with: (a) spherical-world assumption, and (b) depth maps from 3D reconstruction with a few magnified details. Misalignments in the spherical-world panorama are due to the incorrectly estimated depths.
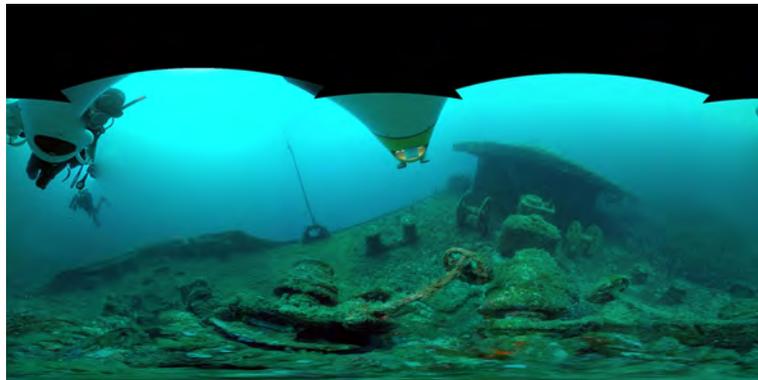
Figure 6.26: Final equirectangular panorama with color blending technique applied. The contributions of each Individual camera are indistinguishable.

Punta Del Molar and Amarrador

Similar to the case of the omnidirectional camera project, another collaboration with research colleagues within the CIRS research center has led to my participation in a series of real-world experiments in which our pipeline was used to demonstrate the effectiveness of the developed robotic exploration algorithm for the exploration of unknown environments while ensuring full coverage of the observed scene with multiple sensors (e.g. profiling sonar and optical camera).

First, an isolated rock - *Punta del Molar*, which is located next to the coastal cliffs of St. Feliu de Guixols (Girona, Spain), was inspected with the Sparus II AUV. The approximately 60 m long rock was chosen because of its irregular shape and complex geometry, as shown in the Figure 6.27. The experiment, which is explained in detail in the journal paper titled *Two-Dimensional Frontier-Based Viewpoint Generation for Exploring and Mapping Underwater Environments* [13] was carried out at a shallow depth of 2.5 m in about 17 min. The final exploration map is shown in Figure 6.28, where the orange cells represent successfully viewed areas of the detected environment. The inspection was performed following a clockwise trajectory, since the cameras were mounted on the right side, and ended at approximately the same position where the robot started.

Figure 6.27: Aerial view of marked *Punta del Molar* rock.
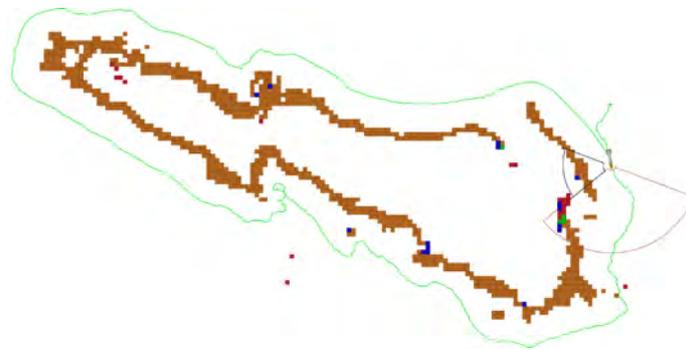
Figure 6.28: Exploration map at the end of the *Punta del Molar* inspection mission. The orange inspection mission. Orange color represents the investigated cells and the green line the trajectory of Sparus II AUV.

The optical images were captured with 3 un-synchronized GoPro Hero 4 Black cameras positioned at the front of the vehicle and oriented right, right-down and forward-to-right-down. The final result of the reconstruction obtained from images of the exploratory mission is shown in the Figure 6.29. As explained in the journal paper, although the algorithm successfully ensured the complete coverage of the scene, the accumulated drift in the last part of the mission resulted in the optical images being effectively captured from a greater distance. Due to the poor visibility conditions, the observed scene was not visible, so our pipeline was not able to reconstruct the complete 3D model. The data from this mission was also used to demonstrate the power of our keyframe selection scheme
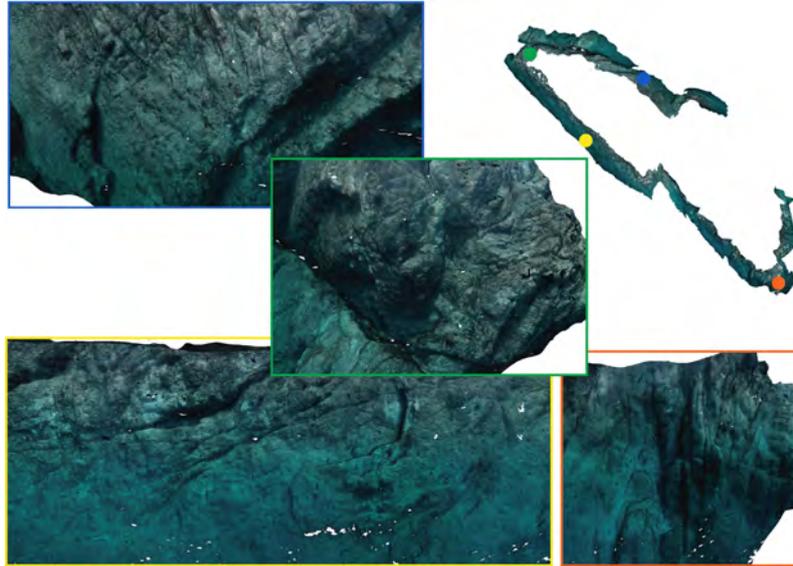
described in section 6.2.

Figure 6.29: Textured 3D model of the explored *Punta del Molar* rock.

In the second scenario, an underwater boulder - *Amarrador* was observed with the Girona 500 AUV, which was equipped with the omnidirectional camera. The boulder, also located near St. Feliu de Guixols (Girona, Spain) at a depth of 40 m, spans an area of 30×15 m and is about 12 m high. The exploration algorithm, which is described in detail in the publication entitled *Multisensor online 3D view planning for autonomous underwater exploration* [18], ensured that the images were taken at a distance of about 4 m from the boulder and the seabed. Using our pipeline, a total of 2,239 images were identified as keyframes (Figure 6.30) and used in the reconstruction process. The final result, in the form of a textured model, is shown in Figure 6.31. As can be seen in the figure, the exploration algorithm ensured that the *Amarrador* boulder was indeed completely observed, with only small patches missing in the surrounding sand areas due to the lack of visible features. Borrowing from our joint publication, we also show the reconstruction obtained by projecting multibeam data onto a single coordinate frame. The multibeam-based reconstruction shown side-by-side with the optical-based model in Figure 6.32 is significantly less detailed due to noise in the multibeam measurements. This demonstrates the advantage of using optical images in scenarios with sufficient visibility.

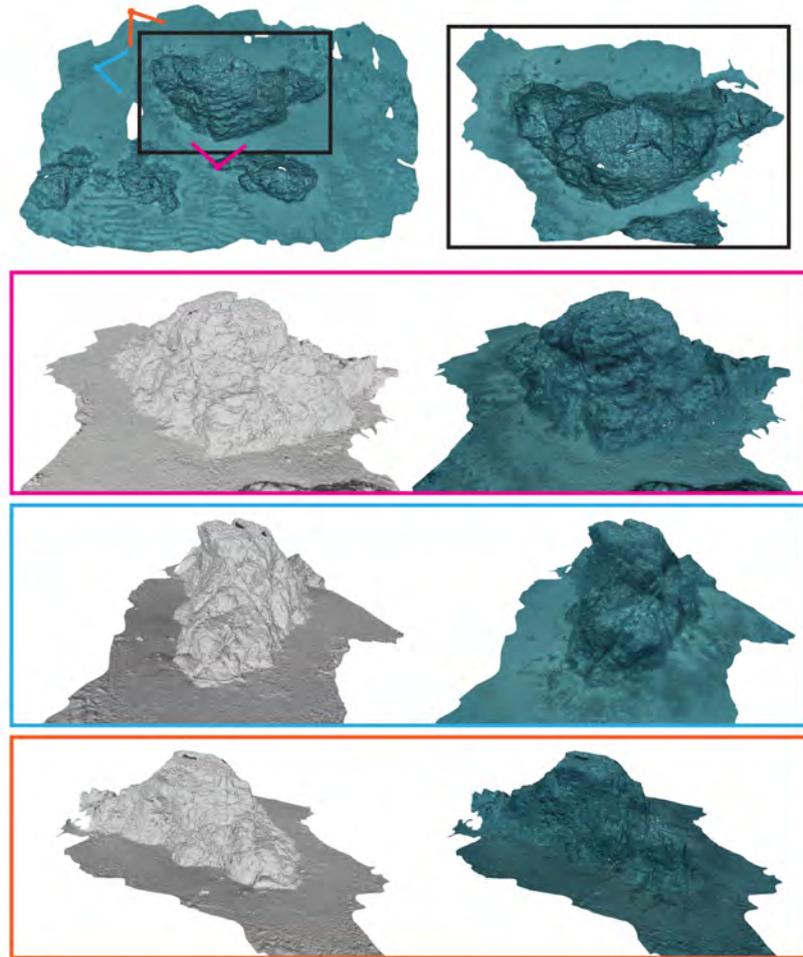**Figure 6.30:** Sample images captured during the exploration of the underwater boulder - *Amarrador*.

Figure 6.31: Surface and textured 3D model of the *Amarrador* boulder reconstructed from images captured by the omnidirectional camera during an autonomous exploration.
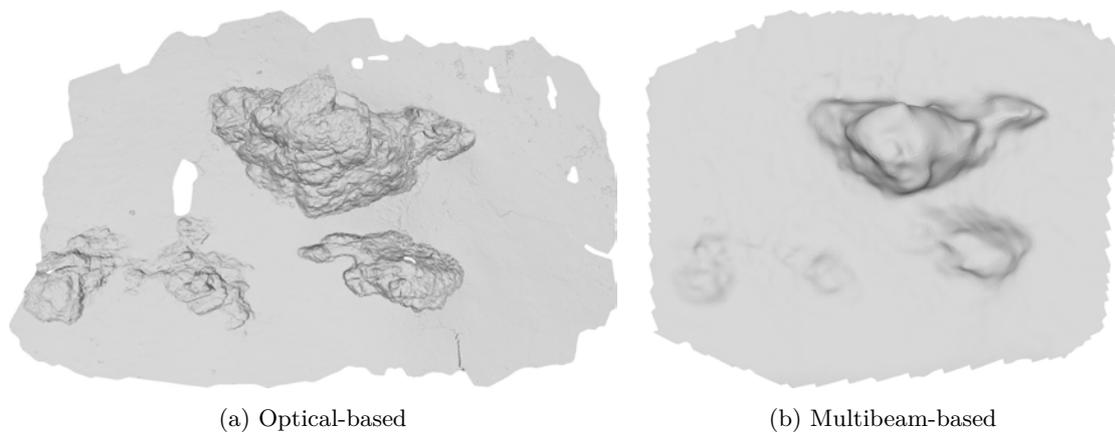

(a) Optical-based

(b) Multibeam-based

Figure 6.32: Comparison of 3D reconstructions created using optical and multibeam echo sounder sensors during the *Amarrador* exploration mission.

Gnalić

The project of surveying a shipwreck near the island of Gnalić was part of a cooperation with archaeologists from the University of Zadar. It demonstrated the methodology for rapid high-resolution mapping of shipwrecks in marine archaeology applications during the "Breaking the Surface 2016" workshop. The Girona 500 AUV was equipped with a high-resolution stereo imaging system (with two Canon EOS 5D Mark II still cameras) and an omnidirectional camera consisting of five individual GoPro Hero4 Black cameras. The robot was programmed to survey a predetermined area of the shipwreck at a depth of $\sim 25\,\text{m}$ for approximately one hour, during which it collected data to create 360° panoramic videos and image-based 3D reconstructions, among other things. Examples of images collected with both camera systems are shown in the Figures 6.34 and 6.36.

Figure 6.33: Girona 500 AUV with omnidirectional and stereo camera system.
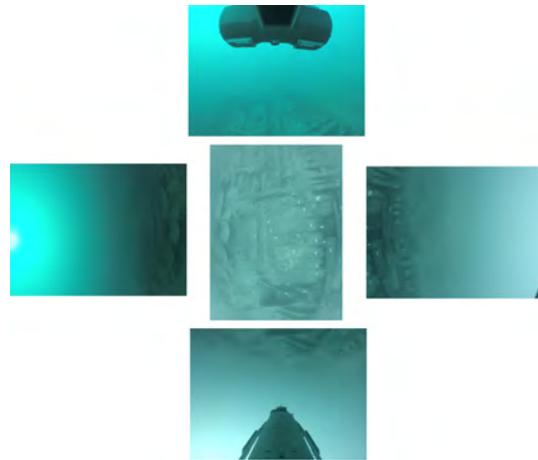
Figure 6.34: Example of original images from each independent camera comprising the omnidirectional camera in *Gnalić* data set.

The results of the reconstruction efforts are presented for both omnidirectional and high-resolution imaging. As stated in the conference paper entitled *Immersive Touring for Marine Archaeology. Application of a New Compact Omnidirectional Camera to Mapping the Gnalić shipwreck with an AUV* [7], the omnidirectional images taken simultaneously from multiple angles on the investigated area not only reduce the time required for surveying, but also enable the introduction of the additional geometric and temporal constraints by matching image pairs between different individual cameras. The complete 3D texture model created using the images from the independent cameras of the omnidirectional system is shown in Figure 6.35.

Due to the relatively high attenuation of light in the water (especially on the red component), the images recorded with GoPro cameras were severely impaired which led to a sub-optimal restoration of color information. However, the more sensitive image sensors in the Canon cameras were able to record sufficient information in the red channel, allowing us to perform better color correction on high-resolution images from the stereo system. Since the camera was positioned almost perpendicular to the seafloor and most of the mission (at least in the part above the shipwreck) was performed at a constant altitude, a single 3×4 color correction matrix was used to remove the color cast caused

Figure 6.35: Textured 3D model of the *Gnalić* shipwreck created using omnidirectional imagery (top view with a few magnified details).

by the attenuation of sunlight. The matrix was calculated from white and black reference points manually determined in the images. An example of the color correction result is shown in Figure 6.36, while the final textured 3D model is shown in Figure 6.37. Although stereo images were available, only images from the left camera were used given that our pipeline did not allow the use of stereo constraints at the time.

(a)   (b)

Figure 6.36: Example of a high-resolution image taken with the Canon EOS 5D Mark II camera of the *Gnalić* shipwreck in: (**a**) original, and (**b**) color- corrected colors.
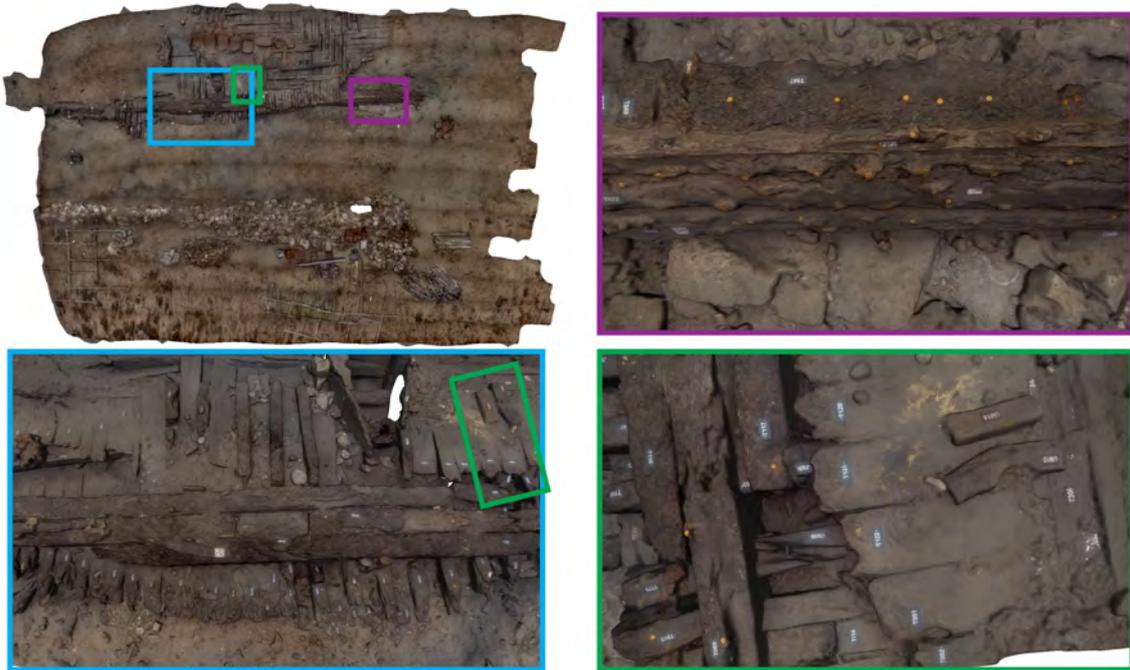
Figure 6.37: Textured 3D model of the *Gnalić* shipwreck created with color corrected images from the left camera of the stereo camera system (top view with a few magnified details).

Gorgonians

In cooperation with the Institute of Marine Sciences (ICM–CSIC) in Barcelona, Spain, the 3D reconstruction pipeline was also successfully used for marine biology purposes. In a pilot study to evaluate the feasibility of recovery and return of gorgonians to their natural environment [14], the marine biologists transplanted the bycatch gorgonians onto three steel structures and deployed them in the marine protected area of Cap de Creus (northwest Mediterranean Sea, Spain). The structures were then monitored over various periods of time with the Girona 500 AUV. In those surveys, optical images were taken with a PointGrey Bumblebee stereo camera by orbiting each of the structures while maintaining an approximately constant distance of 2 m to the structure, as shown with some selected frames in the Figure 6.38.

Figure 6.38: Sample images from a survey monitoring a steel structure with transplanted gorgonians.

In order to provide the researchers with additional material to successfully assess the

survivability of the observed colonies, each of the three structures shown in the Figure 6.39 were reconstructed independently, using images from a single camera of the stereo pair, as the pipeline did not support the introduction of stereo constraints at the time of the project. The unknown scale was unambiguously determined based on measurements taken from the CAD model of the steel structure.

(a) Lander 1
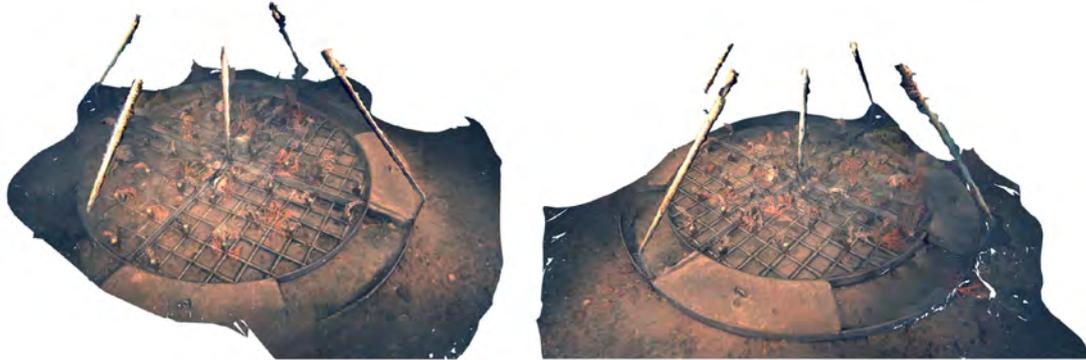
(b) Lander 2
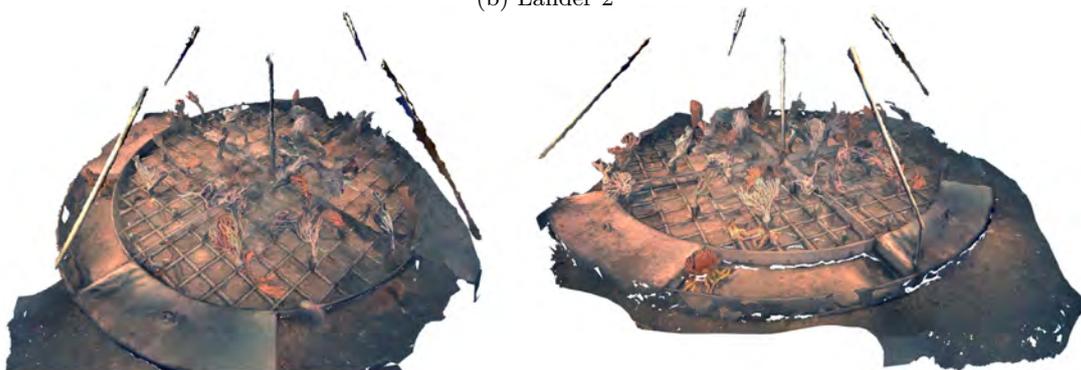
(c) Lander 3

Figure 6.39: Different views of the three textured 3D models of the steel structures with transplanted gorgonians deployed in the marine protected area of Cap de Creus (north-west Mediterranean Sea, Spain).

7

CONCLUSIONS AND FUTURE WORK

THIS chapter closes this thesis by presenting the main conclusions in Section 7.1 and proposing some research lines for future work in Section 7.2.

7.1 Conclusions

This thesis presented a comprehensive exploration and exploitation of image-based 3D reconstruction techniques for practical applications in underwater environments. It has contributed to the advancement of the state of the art in accurate scaling and uncertainty estimation of 3D models built with monocular or un-synchronized camera systems in GPS-denied environments, and to the development of studies in a variety of fields such as robotics, underwater photogrammetry and natural sciences. Through the work presented in the dissertation, the objectives detailed in chapter 1 were successfully completed and led to the following contributions:

Versatile image-based 3D reconstruction framework
We integrated several open-source solutions (openMVG, openMVS, MVS-Texturing) and developed additional functionalities and capabilities (e.g. navigation fusion, soft enforcement of known relative constraints, underwater color correction) in order to develop an end-to-end pipeline capable of reconstructing 3D models for scientific purposes. The framework for creating photo-realistic textured 3D models allows reconstruction based on optical and navigation data and is independent of a specific robot platform, camera or mission.

Extensive use of a 3D reconstruction framework in real-world applications
The developed 3D reconstruction pipeline was extensively used on data collected with six different underwater robotic platforms and numerous different cameras and camera types (monocular, stereo, multi-camera setups). The versatility of the framework led to its use in numerous collaborations with various scientists and researchers. As a result, the models created have been used for accurate measurements in scientific fields as well as for various purposes ranging from the generation of superior omnidirectional video, augmented reality applications or simply as proof that algorithms such as mission planning have successfully ensured satisfactory data acquisition. The pipeline was also used to process data from the *SUBSAINTES* scientific cruise which led to the creation of one of the largest collections of image-derived underwater 3D models acquired by a deep-sea vehicle for geological purposes to date.

Automatic scaling method for SfM-based 3D models with uncertainty estimation
Two novel methods for automatic scaling of SfM-based 3D reconstructions using commonly available laser scalers were proposed. By exploiting image information beyond its use in previous laser photogrammetry methods, the limitations on specific camera poses and/or roughness of the observing scene while estimating the scale were eliminated, and their applicability under a larger variety of circumstances was made possible. Furthermore, we alleviated the need for the manual, time-consuming and error-prone step of image-model point association. Additionally, by endowing the approaches with the process of uncertainty characterization, the scale estimation methods are also suitable for scaling 3D models used in quantitative studies as part of accurate metrology applications. Both methods were validated by a series of generated data sets based on a real 3D model of a deep-sea hydrothermal vent, and their applicability in real scenarios was demonstrated by a data set from the *SUBSAINTES* scientific cruise.

Laser spot detection algorithm with an uncertainty estimation
In order to reduce the very labor-intensive and error-prone process of laser spot detection,

we have developed a method that facilitates the detection of the centers of laser beams, by taking into account the texture of the scene. The uncertainty of these detections is determined based on the recovered characteristic shapes of the laser spots. We have demonstrated the versatility of the approach by processing several data sets with different characteristics.

Laser-Camera Calibration Process
We addressed the possibility of *a posteriori* calibration of the laser-camera orientation based on the image data and a 3D model affected by an unknown scale. Taking advantage of the fact that, in the *SUBSAINTES* data set, the origins of the lasers are stationary and known, we showed that the direction of the lasers can be determined using the data collected during the mission itself. Therefore, this method eliminates the need for a separate acquisition process, enabling its use on numerous legacy data sets in which the location of laser origins in known.

Analysis of image and navigation fusion strategies for underwater 3D reconstruction
Using novel methods for scale estimation and laser spot detection, we performed a comprehensive analysis of the most common image-based 3D reconstruction strategies used by the underwater research community. The advantages and limitations of the different approaches, encapsulating realistic deep-sea field working conditions such as different surveying strategies, multiple dives and extensive outcrop lengths and sizes, were considered. A scale error analysis on large-scale models of deep-sea underwater environments was performed using two different underwater survey scenarios recorded during the *SUBSAINTES* cruise with a scientific ROV (VICTOR 6000). In total, our scaling methods yielded 580 estimates from 1810 laser measurements. The study has shown that surveys with multiple overlaps of non-sequential images result in a nearly identical solution regardless of the strategy (SfM or navigation fusion), whereas surveys with weakly connected sequentially acquired images tend to produce broad-scale deformation (doming effect) if navigation is not included in the optimization. Besides providing insight into the best possible reconstruction strategies, the precise quantification of the uncertainty of the estimated scale for models created with the *SUBSAINTES* data enabled the geologists from IPGP to use these models in their studies. In this way, they were able to extend the preliminary studies of geological features associated with an earthquake, previously conducted with a single model, with the information from more than 30 different sites along a $\sim 20\,\mathrm{km}$ transect.

In addition, we reiterated and analysed the process of determining the optimal weight in *a priori* (multiobjective BA) navigation fusion strategy. We also showed that the introduction of additional vectors in the process of model reconstruction with data collected by multiple surveys successfully mitigates the potential offset in navigation of USBL-based positioning.

Mission time 3D reconstruction and uncertainty estimation
The proposed SfM system, which is capable of estimating the trajectory and a sparse 3D representation of the scene together with the associated uncertainties in mission time, i.e. while the robot is in the water or very shortly after, contributes to the field by demonstrating the feasibility of mission-time 3D reconstruction and uncertainty estimation. By exploiting the incremental way of solving the BA problem and covariance recovery, the undesirable dependence of the outcome of the reconstruction process on the conditions

and strategies of the acquisition process is partially addressed. This opens up the possibility of quality-aware data acquisitions which will consequently not only increase the quality of the final acquired data and the survey efficiency, but at the same time reduce the possibility of conducting unsatisfactory optical surveys.

7.2   Future work

Although this dissertation covered a wide range of topics related to image-based 3D reconstruction, there are still many different possible extensions and future lines of research that would go beyond the scope of this thesis.

The complete image-based end-to-end framework for 3D reconstruction was created based on the conventional approach to multi-view geometry. We used numerous traditional computer vision techniques to perform steps such as feature extraction and matching, pose and structure estimation, meshing and texture mapping. Similarly, we relied on statistical methods for underwater color correction and keyframe extraction. With the emergence and rapid development of powerful machine learning techniques, such as deep learning (DL), which are able to solve problems that were considered unsolvable with superhuman accuracy [164], we believe that promising future work would be to integrate novel DL solutions to these problems independently, and to explore the potential applicability of the techniques in an end-to-end manner, with the aim of benefiting most from both learnable priors and geometric rules. Given that DL excels at solving problems where sufficient data is available and new data closely resembles the training set, the fusion of traditional computer vision and DL techniques should be investigated. This can improve performance and help reduce some of the weaknesses of DL in underwater applications where data is scarce and extremely dependent on environmental conditions.

While the resulting 3D models are already used in numerous applications, e.g. for scientific quantitative studies, the generation of omnidirectional video and in augmented reality scenarios, we further envision their use as additional information in various other research areas. The information about the pose of the cameras and the geometry of the scene could be further leveraged to achieve better results in image dehazing and color recovery processes. Done iteratively, the improvement of the original images will lead to an even better and more accurate 3D model which will hence lead to a better image correction. Similarly, 3D models and auxiliary camera pose information could be further used to detect temporal changes in the observed environment. During our wcork with the *MIDAS* data set, we identified several improvements that could increase the robustness and accuracy of change detection. An additional, non-rigid alignment of 3D models that goes beyond the alignment of common sparse points would significantly reduce the erroneous detection of changes that occur due to differences caused by the densification and meshing processes. Moreover, in environments where the changes are both geometrically and photometrically significant, the association between multi-temporal image sets has proven to be extremely difficult for traditional hand-crafted image features. The use of DL techniques known for their ability to learn different priors could prove extremely beneficial.

As already briefly mentioned in section XX, the purpose of our proposed scale estimation approaches can be further extended. Currently the methods provide estimates of scale for the reconstructed 3D models in an *a posteriori* step. The scale information is computed based on the back-projection of 3D points determined as intersections between the reconstructed model of the scene and the laser rays. As both FUM and PCM

achieve high accuracies by compensating for terrain roughness (among other things), it is important that the scene representation used in the ray casting procedure is as detailed and accurate as possible. This therefore implies the use of a highly detailed description of the scene obtained through multiple steps following SfM (densification and surface reconstruction). Such approach limits its use to scale error evaluations, and as means of determining a general scale correction factor used in rigid transformation of arbitrarily scaled 3D models. While this enables the use of the models for quantitative studies, the *a posteriori* nature of our approach (with regards to SfM) precludes it from being directly included in SfM as well as to be used to correct potential scale drifts within the model.

Since there are clear benefits of incorporating the scale estimates within the optimization (e.g., optimal scaling of the model given the laser measurements, possible prevention/minimization of scale drifts and decrease in broad-scale systematic deformations), this is seen as a natural next step in the development of the approach.

The use of scale estimates within SfM would require several adjustments to our current proposal. The pose estimation step would no longer be a separate image localization step based on known and fixed 3D points representing the environment. Instead, the poses related to the images containing laser spots would be estimated simultaneously with the other camera poses. In addition the 3D intersections between the reconstructed model and the laser rays would no longer be computed based on a highly detailed 3D model to eliminate the *a posteriori* requirement of the scale estimation process. In this case, the locations of 3D points would have to be approximated using the spare set of 3D points jointly estimated within SfM. Finally, the cost function would also need to be augmented to include appropriate cost terms representing the error in the scale estimate. While the benefits of such a complex implementation should not be ignored, it is important to note that determining scale based solely on the 3D geometry of the scene described with sparse 3D points could lead to erroneous scale estimates if the roughness of the scene is not satisfactorily represented, and therefore such an approach would need to be used carefully and with full knowledge of its limitations.

With regard to the precise scale error analysis, the proposed future work addresses the problem of the inability of methods to decouple two different sources of error. As discussed in the publication, the fact that our method estimates the scale factor exclusively from image data containing laser spots leads to the possibility of the resulting error being caused by errors in the camera and camera-laser calibrations and not an actual error in the model. While we believe that small deviations in between the independent scale estimates obtained from each laser scaler within an image indicate that the calibrations are correct, this could not be unambiguously determined. Performing a similar analysis with a data set containing several auxiliary objects of known dimensions placed in the scene would allow us to clearly decouple the two sources of error and check our assumptions that the reported errors are indeed the actual errors in the model. Furthermore, such an analysis would also contribute significantly to the generalization argument of our conclusions (if confirmed).

Since the scale estimation framework depends heavily on the ability to reliably detect laser spots, any improvement in the detection process will lead to a more accurate and/or usable approach. Although we have shown in chapter 6.5 that the method we propose is applicable and accurate for different data sets, the reliability of the detections can be further improved. Currently, the detection method is designed as a stand-alone approach that uses only the information from the main and auxiliary images. By integrating the

method into the estimation framework and using the reconstructed 3D model, the additional geometric knowledge of the environment could be used to further narrow down the search area for laser spots and robustly reject outliers. In cases where the collected data is in a video format, the known temporal relationship of the detections and the known geometry could be further exploited.

Finally, in relation to the proposed mission-time SfM system, a full implementation allowing it to be integrated into our Girona 500 AUV and/or Sparus II AUV is still to be done. Although we have made the first approach to implementing the system, which could be used in real applications, the processing pipeline would greatly benefit from further parallelization of the tracking module and extensive testing is needed. In addition, a fully functional application must have an intuitive visual interface for pilots to enable quality-conscious acquisition missions.

Bibliography

[1] Juan David Hernández, **Klemen Istenič**, Nuno Gracias, Narcis Palomeras, Ricard Campos, Eduard Vidal, Rafael Garcia, and Marc Carreras. "Autonomous underwater navigation and optical mapping in unknown natural environments". In: *Sensors* 16.8 (2016), page 1174. DOI: 10.3390/s16081174 (cited on pages ix, 17, 21).

[2] **Klemen Istenič**, Nuno Gracias, Aurélien Arnaubec, Javier Escartín, and Rafael Garcia. "Scale Accuracy Evaluation of Image-Based 3D Reconstruction Strategies Using Laser Photogrammetry". In: *Remote Sensing* 11.18 (2019), page 2093. DOI: 10.3390/rs11182093 (cited on pages ix, 63).

[3] **Klemen Istenič**, Nuno Gracias, Aurélien Arnaubec, Javier Escartín, and Rafael Garcia. "Automatic scale estimation of structure from motion based 3D models using laser scalers in underwater scenarios". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (2020), pages 13–25. DOI: 10.1016/j.isprsjprs.2019.10.007 (cited on pages x, 49).

[4] **Klemen Istenič**, Viorela Ila, Lukáš Polok, Nuno Gracias, and Rafael García. "Mission-time 3D reconstruction with quality estimation". In: *OCEANS 2017-Aberdeen*. IEEE. 2017, pages 1–9. DOI: 10.1109/OCEANSE.2017.8084708 (cited on pages x, 17, 97, 110).

[5] Juan David Hernández, **Klemen Istenič**, Nuno Gracias, Rafael García, Pere Ridao, and Marc Carreras. "Autonomous seabed inspection for environmental monitoring". In: *Robot 2015: Second Iberian Robotics Conference*. Springer. 2016, pages 27–39. DOI: 10.1007/978-3-319-27149-1_3 (cited on page xi).

[6] Viorela Ila, Lukáš Polok, Marek Solony, and **Klemen Istenič**. "Fast incremental bundle adjustment with covariance recovery". In: *2017 International Conference on 3D Vision (3DV)*. IEEE. 2017, pages 175–184. DOI: 10.1109/3DV.2017.00029 (cited on pages xi, 17, 110, 134).

[7] Josep Bosch, Nuno Gracias, Pere Ridao, David Ribas, **Klemen Istenič**, Rafael Garcia, and Irena Radic Rossi. "Immersive Touring for Marine Archaeology. Application of a New Compact Omnidirectional Camera to Mapping the Gnalić shipwreck with an AUV". In: *Iberian Robotics conference*. Springer. 2017, pages 183–195. DOI: 10.1007/978-3-319-70833-1_16 (cited on pages xi, 17, 144).

[8]   Eduard Vidal, Juan David Hernández, **Klemen Istenič**, and Marc Carreras. "Online view planning for inspecting unexplored underwater structures". In: *IEEE Robotics and Automation Letters* 2.3 (2017), pages 1436–1443. DOI: `10.1109/LRA.2017.2671415` (cited on pages xi, 17).

[9]   Javier Escartin, Frédérique Leclerc, Feuillet Nathalie, Anne Le Friant, Jérémy Billant, Jean-Arthur L Olive, Mathilde Henri, Muriel Andreani, Aurelien Arnaubec, Alexandre Dano, et al. "Mapping the Mw6. 3 2004 Les Saintes earthquake seafloor rupture with deep-sea vehicles: Length, displacement, nature, and links between coseismic deformation and erosion/sedimentation". In: *AGU Fall Meeting Abstracts*. 2018 (cited on pages xi, 17, 129, 135).

[10]   Jérémy Billant, Frédérique Leclerc, Javier Escartin, Nuno Gracias, **Klemen Istenič**, Aurelien Arnaubec, Feuillet Nathalie, and Christine Deplus. "Submarine expression of the earthquake cycle: morphotectonics analysis of the normal Roseau fault (Lesser Antilles, France) from high resolution bathymetry and 3D photogrammetric models to reconstruct its slip history". In: *AGU Fall Meeting Abstracts*. 2018 (cited on pages xi, 17, 129, 135).

[11]   Eduard Vidal, Juan David Hernández, **Klemen Istenič**, and Marc Carreras. "Optimized environment exploration for autonomous underwater vehicles". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2018, pages 6409–6416. DOI: `10.1109/ICRA.2018.8460919` (cited on pages xi, 17, 111, 112).

[12]   Josep Bosch, **Klemen Istenič**, Nuno Gracias, Rafael Garcia, and Pere Ridao. "Omnidirectional Multicamera Video Stitching Using Depth Maps". In: *IEEE Journal of Oceanic Engineering* (2019). DOI: `10.1109/JOE.2019.2924276` (cited on pages xi, 12, 17, 136, 139).

[13]   Eduard Vidal, Narcís Palomeras, **Klemen Istenič**, Juan David Hernández, and Marc Carreras. "Two-Dimensional Frontier-Based Viewpoint Generation for Exploring and Mapping Underwater Environments". In: *Sensors* 19.6 (2019), page 1460. DOI: `10.3390/s19061460` (cited on pages xi, 17, 141).

[14]   Maria Montseny, Cristina Linares, Núria Viladrich, Alejandro Olariaga, Marc Carreras, Narcís Palomeras, Nuno Gracias, **Klemen Istenič**, Rafael Garcia, Stefano Ambroso, et al. "First attempts towards the restoration of gorgonian populations on the Mediterranean continental shelf". In: *Aquatic Conservation: Marine and Freshwater Ecosystems* 29.8 (2019), pages 1278–1284. DOI: `10.1002/aqc.3118` (cited on pages xii, 17, 146).

[15]   Javier Escartin, Jeremy Billant, Frederique Leclerc, Jean-Arthur Olive, **Klemen Istenič**, Nuno Gracias, Aurelien Arnaubec, Rafael Garcia, Christine Deplus, Nathalie Feuillet, et al. "Seafloor rupture of a normal fault: geometry, displacement, and links between erosion/sedimentation and seismicity." In: *Geophysical Research Abstracts*. Volume 21. 2019 (cited on pages xii, 17, 129, 135).

[16]   Jérémy Billant, Frédérique Leclerc, Javier Escartin, Nuno Gracias, **Klemen Istenič**, Rafael Garcia, Aurélien Arnaubec, Alexandre Dano, and Claire Marchand. "Development of a Unity package allowing GIS-like mapping in a Virtual Reality environment." In: *Geophysical Research Abstracts*. Volume 21. 2019 (cited on pages xii, 17, 129, 135).

[17] Jérémy Billant, Júlia Bozzinio, Frédérique Leclerc, Javier Escartin, Nuno Gracias, **Klemen Istenič**, Aurelien Arnaubec, and Rafael Garcia. "Performing submarine field survey without scuba gear using GIS-like mapping in a Virtual Reality environment". In: *OCEANS 2019-Marseille*. IEEE. 2019, pages 1–6. DOI: `10.1109/OCEANSE.2019.8867408` (cited on pages xii, 17, 129, 135, 136).

[18] Eduard Vidal, Narcís Palomeras, **Klemen Istenič**, Nuno Gracias, and Marc Carreras. "Multisensor online 3D view planning for autonomous underwater exploration". In: *Journal of Field Robotics* (2020). DOI: `10.1002/rob.21951` (cited on pages xii, 17, 142).

[19] Josep Bosch, Nuno Gracias, Pere Ridao, **Klemen Istenič**, and David Ribas. "Close-range tracking of underwater vehicles using light beacons". In: *Sensors* 16.4 (2016), page 429. DOI: `10.3390/s16040429` (cited on page xii).

[20] Eduardo Ochoa, Nuno Gracias, **Klemen Istenič**, Rafael García, Josep Bosch, and Patryk Cieslak. "Allowing untrained scientists to safely pilot ROV's: Early collision detection and avoidance using omnidirectional vision". In: *OCEANS Global 2020*. IEEE. 2020 (cited on page xii).

[21] Hawaii Pacific University Oceanic Institute. *Aqua Facts*. `https://reefdivers.io/x-adventurer-m6000-video-lights/7457`. [Online; accessed 11-January-2020] (cited on page 8).

[22] National Oceanic and Atmospheric Administration. *How much of the ocean have we explored?* `https://oceanservice.noaa.gov/facts/exploration.html`. [Online; accessed 03-October-2019]. 2018 (cited on page 8).

[23] Robert D Ballard. "The MEDEA/JASON remotely operated vehicle system". In: *Deep Sea Research Part I: Oceanographic Research Papers* 40.8 (1993), pages 1673–1687. DOI: `10.1016/0967-0637(93)90021-T` (cited on page 8).

[24] Jean-Louis Michel, Michaël Klages, Fernando JAS Barriga, Yves Fouquet, Myriam Sibuet, Pierre-Marie Sarradin, Patrick Siméoni, Jean-François Drogou, et al. "Victor 6000: design, utilization and first improvements". In: *The Thirteenth International Offshore and Polar Engineering Conference* (Honolulu, HI, USA). International Society of Offshore and Polar Engineers. May 2003 (cited on pages 8, 16).

[25] Carl L Kaiser, Dana R Yoerger, James C Kinsey, Sean Kelley, Andrew Billings, Justin Fujii, Stefano Suman, Michael Jakuba, Zachary Berkowitz, Christopher R German, et al. "The design and 200 day per year operation of the Autonomous Underwater Vehicle Sentry". In: *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*. IEEE. 2016, pages 251–260. DOI: `10.1109/AUV.2016.7778680` (cited on page 8).

[26] Woods Hole Oceanographic Institution. *REMUS 6000 - Oceanographic Systems Lab*. `https://www2.whoi.edu/site/osl/vehicles/remus-6000/?pid=38144`. [Online; accessed 14-January-2020] (cited on page 8).

[27] David Ribas, Narcís Palomeras, Pere Ridao, Marc Carreras, and Angelos Mallios. "Girona 500 AUV: From Survey to Intervention". In: *IEEE/ASME Transactions on Mechatronics* 17.1 (2012), pages 46–53. DOI: `10.1109/TMECH.2011.2174065` (cited on pages 8, 16, 17).

[28]   Hanumant Singh, Ali Can, Ryan Eustice, Steve Lerner, Neil McPhee, and Chris Roman. "Seabed AUV offers new platform for high-resolution imaging". In: *Eos, Transactions American Geophysical Union* 85.31 (2004), pages 289–296. DOI: `10.1029/2004EO310002` (cited on page 8).

[29]   Saab Seaeye. *Cougar-XT*. `https://www.saabseaeye.com/solutions/underwater-vehicles/cougar-xt`. [Online; accessed 14-January-2020] (cited on pages 8, 16).

[30]   Marc Carreras, Juan David Hernández, Eduard Vidal, Narcís Palomeras, David Ribas, and Pere Ridao. "Sparus II AUV-A Hovering Vehicle for Seabed Inspection". In: *IEEE Journal of Oceanic Engineering (JOE)* 43.2 (2018), pages 344–355. DOI: `10.1109/JOE.2018.2792278` (cited on pages 8, 16, 17).

[31]   Sofar. *Trident Underwater Drone*. `https://www.sofarocean.com/products/trident`. [Online; accessed 14-January-2020] (cited on page 8).

[32]   SeaDrone. *Inspector 3*. `https://seadronepro.com/inspector-3`. [Online; accessed 14-January-2020] (cited on page 8).

[33]   Blue Robotics. *BlueROV2 - Affordable and Capable Underwater ROV*. `https://bluerobotics.com/store/rov/bluerov2/`. [Online; accessed 14-January-2020] (cited on page 8).

[34]   Daniel J Fornari and WHOI TowCam Group. "A new deep-sea towed digital camera and multi-rock coring system". In: *Eos, Transactions American Geophysical Union* 84.8 (2003), pages 69–73. DOI: `10.1029/2003EO080001` (cited on page 8).

[35]   Peter Hill. "Designing a Deep-Towed Camera Vehicle Using Single Conductor Cable". In: *Sea Technology* 50 (Dec. 2009), pages 49–51 (cited on page 8).

[36]   Daniel OB Jones, Brian J Bett, Russell B Wynn, and Douglas G Masson. "The use of towed camera platforms in deep-water science". In: *Underwater Technology* 28.2 (2009), pages 41–50. DOI: `10.3723/ut.28.041` (cited on page 8).

[37]   Autun Purser, Yann Marcon, Simon Dreutter, Ulrich Hoge, Burkhard Sablotny, Laura Hehemann, Johannes Lemburg, Boris Dorschel, Harald Biebow, and Antje Boetius. "Ocean Floor Observation and Bathymetry System (OFOBS): a new towed camera/sonar system for deep-sea habitat surveys". In: *IEEE Journal of Oceanic Engineering* 44.1 (2018), pages 87–99. DOI: `10.1109/JOE.2018.2794095` (cited on page 8).

[38]   Matthew Dunbabin and Lino Marques. "Robots for environmental monitoring: Significant advancements and applications". In: *IEEE Robotics & Automation Magazine* 19.1 (2012), pages 24–39. DOI: `10.1109/MRA.2011.2181683` (cited on page 8).

[39]   National Deep Submergence Facility. *HOV Alvin*. `https://ndsf.whoi.edu/alvin/`. [Online; accessed 27-December-2019] (cited on page 8).

[40]   Colleen M Cavanaugh, Stephen L Gardiner, Meredith L Jones, Holger W Jannasch, and John B Waterbury. "Prokaryotic cells in the hydrothermal vent tube worm Riftia pachyptila Jones: possible chemoautotrophic symbionts". In: *Science* (1981), pages 340–342. DOI: `10.1126/science.213.4505.340` (cited on page 8).

[41]   C IHO. *Manual Of Hydrography*. 13 (cited on page 9).

[42]   Justin Marshall. "Vision and lack of vision in the ocean". In: *Current Biology* 27.11 (2017), R494–R502. DOI: `10.1016/j.cub.2017.03.012` (cited on pages 9, 10).

[43] Curtis D Mobley. "Radiative transfer in the ocean". In: (2001). DOI: 10.1006/rwos.2001.0469 (cited on page 9).

[44] Chongyi Li, Jichang Guo, and Chunle Guo. "Emerging from water: Underwater image color correction based on weakly supervised color transfer". In: *IEEE Signal processing letters* 25.3 (2018), pages 323–327. DOI: 10.1109/LSP.2018.2792050 (cited on page 9).

[45] Nils Gunnar Jerlov. "Irradiance optical classification". In: *Optical Oceanography* (1968), pages 118–120. DOI: 10.1016/S0422-9894(08)70929-2 (cited on page 9).

[46] Michael G Solonenko and Curtis D Mobley. "Inherent optical properties of Jerlov water types". In: *Applied optics* 54.17 (2015), pages 5392–5401. DOI: 10.1364/AO.54.005392 (cited on page 9).

[47] Seibert Q Duntley. "Light in the sea". In: *JOSA* 53.2 (1963), pages 214–233. DOI: 10.1364/JOSA.53.000214 (cited on page 9).

[48] Derya Akkaynak, Tali Treibitz, Tom Shlesinger, Yossi Loya, Raz Tamir, and David Iluz. "What is the space of attenuation coefficients in underwater computer vision?" In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2017, pages 4931–4940. DOI: 10.1109/CVPR.2017.68 (cited on page 10).

[49] Rafael Garcia, Tudor Nicosevici, and Xevi Cufí. "On the way to solve lighting problems in underwater imaging". In: *OCEANS'02 MTS/IEEE.* Volume 2. IEEE. 2002, pages 1018–1024. DOI: 10.1109/OCEANS.2002.1192107 (cited on page 10).

[50] Nuno Gracias, Shahriar Negahdaripour, Laszlo Neumann, Ricard Prados, and Rafael Garcia. "A motion compensated filtering approach to remove sunlight flicker in shallow water images". In: *OCEANS 2008.* IEEE. 2008, pages 1–7. DOI: 10.1109/OCEANS.2008.5152111 (cited on pages 10, 11).

[51] Panagiotis Agrafiotis, Dimitrios Skarlatos, Timothy Forbes, Charalambos Poullis, Margarita Skamantzari, and Andreas Georgopoulos. "Underwater photogrammetry in very shallow waters: Main challenges and caustics effect removal". In: International Society for Photogrammetry and Remote Sensing. 2018. DOI: 10.5194/isprs-archives-XLII-2-15-2018 (cited on page 10).

[52] Nicole Helgason. *X-Adventurer M6000 Video Lights Are Making Our Divers A Heck Of A Lot Brighter!* https://reefdivers.io/x-adventurer-m6000-video-lights/7457. [Online; accessed 11-January-2020] (cited on page 11).

[53] Miquel Massot-Campos and Gabriel Oliver-Codina. "Optical sensors and methods for underwater 3D reconstruction". In: *Sensors* 15.12 (2015), pages 31525–31557. DOI: 10.3390/s151229864 (cited on pages 11, 12).

[54] Enric Galceran, Ricard Campos, Narcís Palomeras, David Ribas, Marc Carreras, and Pere Ridao. "Coverage path planning with real-time replanning and surface reconstruction for inspection of three-dimensional underwater structures using autonomous underwater vehicles". In: *Journal of Field Robotics* 32.7 (2015), pages 952–983. DOI: 10.1002/rob.21554 (cited on page 11).

[55] Bruno Jouvencel, Xianbo Xiang, Lei Zhang, and Zheng Fang. "3d reconstruction of seabed surface through sonar data of auvs". In: *Indian Journal of Geo-Marine Sciences* 41.6 (2012), pages 509–515 (cited on page 11).

[56] Georgios Papadopoulos, Hanna Kurniawati, Ahmed Shafeeq Bin Mohd Shariff, Liang Jie Wong, and Nicholas M Patrikalakis. "3D-surface reconstruction for partially submerged marine structures using an autonomous surface vehicle". In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2011, pages 3551–3557. DOI: 10.1109/IROS.2011.6094966 (cited on page 11).

[57] Enrique Coiras and Johannes Groen. "3D target shape from SAS images based on a deformable mesh". In: *Proceedings of the 3rd International Conference on Underwater Acoustic Measurements (UAM)*. 2009, pages 382–390 (cited on page 11).

[58] Krzysztof Bikonis, Andrzej Stepnowski, and Marek Moszynski. "Computer vision techniques applied for reconstruction of seafloor 3d images from side scan and synthetic aperture sonars data". In: *Journal of the Acoustical Society of America* 123.5 (2008), page 3748. DOI: 10.1121/1.2935303 (cited on page 11).

[59] Thomas Guerneve, Kartic Subr, and Yvan Petillot. "Three-dimensional reconstruction of underwater objects using wide-aperture imaging SONAR". In: *Journal of Field Robotics* 35.6 (2018), pages 890–905. DOI: 10.1002/rob.21783 (cited on pages 11, 12).

[60] Kio Kim, Nicola Neretti, and Nathan Intrator. "Mosaicing of acoustic camera images". In: *IEE Proceedings-Radar, Sonar and Navigation* 152.4 (2005), pages 263–270. DOI: 10.1049/ip-rsn:20045015 (cited on page 11).

[61] Shahriar Negahdaripour, Pezhman Firoozfam, and Payam Sabzmeydani. "On processing and registration of forward-scan acoustic video imagery". In: *The 2nd Canadian Conference on Computer and Robot Vision (CRV'05)*. IEEE. 2005, pages 452–459. DOI: 10.1109/CRV.2005.57 (cited on page 11).

[62] Murat D Aykin and Shahriar Negahdaripour. "On feature matching and image registration for two-dimensional forward-scan sonar imaging". In: *Journal of Field Robotics* 30.4 (2013), pages 602–623. DOI: 10.1002/rob.21461 (cited on page 11).

[63] Natàlia Hurtós, David Ribas, Xavier Cufí, Yvan Petillot, and Joaquim Salvi. "Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments". In: *Journal of Field Robotics* 32.1 (2015), pages 123–151. DOI: 10.1002/rob.21516 (cited on page 11).

[64] Paul Ozog, Giancarlo Troni, Michael Kaess, Ryan M Eustice, and Matthew Johnson-Roberson. "Building 3d mosaics from an autonomous underwater vehicle, doppler velocity log, and 2d imaging sonar". In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2015, pages 1137–1143. DOI: 10.1109/ICRA.2015.7139334 (cited on page 12).

[65] Tiffany A Huang and Michael Kaess. "Towards acoustic structure from motion for imaging sonar". In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2015, pages 758–765. DOI: 10.1109/IROS.2015.7353457 (cited on page 12).

[66] Shahriar Negahdaripour. "On 3-D reconstruction from stereo FS sonar imaging". In: *OCEANS 2010 MTS/IEEE SEATTLE*. IEEE. 2010, pages 1–6. DOI: 10.1109/OCEANS.2010.5664569 (cited on page 12).

[67] Ricard Campos, Rafael Garcia, Pierre Alliez, and Mariette Yvinec. "A surface reconstruction method for in-detail underwater 3D optical mapping". In: *The International Journal of Robotics Research* (2014), pages 64–89. DOI: 10.1177/0278364914544531 (cited on page 12).

[68] F Bruno, Gianfranco Bianco, Maurizio Muzzupappa, Sandro Barone, and Armando Viviano Razionale. "Experimentation of structured light and stereo vision for underwater 3D reconstruction". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.4 (2011), pages 508–518. DOI: 10.1016/j.isprsjprs.2011.02.009 (cited on page 12).

[69] Albert Palomer, Pere Ridao, Dina Youakim, David Ribas, Josep Forest, and Yvan Petillot. "3D laser scanner for underwater manipulation". In: *Sensors* 18.4 (2018), page 1086. DOI: 10.3390/s18041086 (cited on page 12).

[70] Miquel Massot-Campos, Gabriel Oliver-Codina, Hashim Kemal, Yvan Petillot, and Francisco Bonin-Font. "Structured light and stereo vision for underwater 3D reconstruction". In: *OCEANS 2015-Genova*. IEEE. 2015, pages 1–6. DOI: 10.1109/OCEANS-Genova.2015.7271433 (cited on page 12).

[71] Pedro J Sanz, A Penalver, Jorge Sales, David Fornas, J Javier Fernández, Javier Pérez, and J Bernabé. "GRASPER: A multisensory based manipulation system for underwater operations". In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE. 2013, pages 4036–4041. DOI: 10.1109/SMC.2013.689 (cited on page 12).

[72] Adrian Bodenmann, Blair Thornton, and Tamaki Ura. "Development of long range color imaging for wide area 3D reconstructions of the seafloor". In: *2013 IEEE International Underwater Technology Symposium (UT)*. IEEE. 2013, pages 1–5. DOI: 10.1109/UT.2013.6519824 (cited on page 12).

[73] Hayato Kondo, TOSHIHIRO Maki, TAMAKI Ura, Yoshiaki Nose, T Sakamaki, and M Inaishi. "Structure tracing with a ranging system using a sheet laser beam". In: *Proceedings of the 2004 International Symposium on Underwater Technology (IEEE Cat. No. 04EX869)*. IEEE. 2004, pages 83–88. DOI: 10.1109/UT.2004.1405483 (cited on page 12).

[74] Jules S Jaffe and Chris Dunn. "A model-based comparison of underwater imaging systems". In: *Ocean Optics IX*. Volume 925. International Society for Optics and Photonics. 1988, pages 344–350. DOI: 10.1117/12.945742 (cited on page 12).

[75] Gabrielle Inglis, Clara Smart, Ian Vaughn, and Chris Roman. "A pipeline for structured light bathymetric mapping". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, pages 4425–4432. DOI: 10.1109/IROS.2012.6386038 (cited on page 12).

[76] Adrian Bodenmann, Blair Thornton, Takeshi Nakataniy, and Tamaki Ura. "3D colour reconstruction of a hydrothermally active area using an underwater robot". In: *OCEANS'11 MTS/IEEE KONA*. IEEE. 2011, pages 1–6. DOI: 10.23919/OCEANS.2011.6107059 (cited on page 12).

[77] Gianfranco Bianco, Alessandro Gallo, Fabio Bruno, and Maurizio Muzzupappa. "A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects". In: *Sensors* 13.8 (2013), pages 11007–11031. DOI: 10.3390/s130811007 (cited on page 12).

[78]   Noah Snavely, Steven M Seitz, and Richard Szeliski. "Photo tourism: exploring photo collections in 3D". In: *ACM transactions on graphics (TOG)*. Volume 25. 3. ACM. 2006, pages 835–846. DOI: 10.1145/1141911.1141964 (cited on page 13).

[79]   Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. 2nd edition. New York, NY, USA: Cambridge University Press, 2003. DOI: 10.1017/CBO9780511811685.001 (cited on pages 13, 108, 111).

[80]   Noah Snavely, Steven M. Seitz, and Richard Szeliski. "Modeling the World from Internet Photo Collections". In: *International Journal of Computer Vision* 80.2 (Nov. 2008), pages 189–210. DOI: 10.1007/s11263-007-0107-3 (cited on page 13).

[81]   F. Remondino, S. F. El-Hakim, A. Gruen, and L. Zhang. "Turning images into 3-D models". In: *IEEE Signal Processing Magazine* 25.4 (July 2008), pages 55–65. DOI: 10.1109/MSP.2008.923093 (cited on page 13).

[82]   S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. "Building Rome in a day". In: *2009 IEEE 12th International Conference on Computer Vision*. Sept. 2009, pages 72–79. DOI: 10.1109/ICCV.2009.5459148 (cited on page 13).

[83]   Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. "Bundle adjustment – a modern synthesis". In: *Vision algorithms: theory and practice*. Berlin, Germany: Springer, 1999, pages 298–372. DOI: 10.1007/3-540-44480-7_21 (cited on page 13).

[84]   Oscar Pizarro, Ariell Friedman, Mitch Bryson, Stefan B Williams, and Joshua Madin. "A simple, fast, and repeatable survey method for underwater visual 3D benthic mapping and monitoring". In: *Ecology and Evolution* 7.6 (2017), pages 1770–1782. DOI: 10.1002/ece3.2701 (cited on page 13).

[85]   Curt D Storlazzi, Peter Dartnell, Gerald A Hatcher, and Ann E Gibbs. "End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology". In: *Coral Reefs* 35.3 (2016), pages 889–894. DOI: 10.1007/s00338-016-1462-8 (cited on page 13).

[86]   P Rossi, C Castagnetti, A Capra, AJ Brooks, and F Mancini. "Detecting change in coral reef 3D structure using underwater photogrammetry: critical issues and performance metrics". In: *Applied Geomatics* (2019), pages 1–15. DOI: 10.1007/s12518-019-00263-w (cited on page 13).

[87]   Brian Bingham, Brendan Foley, Hanumant Singh, Richard Camilli, Katerina Delaporta, Ryan Eustice, Angelos Mallios, David Mindell, Christopher Roman, and Dimitris Sakellariou. "Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle". In: *Journal of Field Robotics* 27.6 (2010), pages 702–717. DOI: 10.1002/rob.20350 (cited on page 13).

[88]   Javier Escartín, Frédérique Leclerc, Jean-Arthur Olive, Catherine Mevel, Mathilde Cannat, Sven Petersen, Nico Augustin, Nathalie Feuillet, Christine Deplus, Antoine Bezos, et al. "First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history". In: *Earth and Planetary Science Letters* 450 (2016), pages 96–107. DOI: 10.1016/j.epsl.2016.06.024 (cited on page 13).

[89] Adrian Bodenmann, Blair Thornton, and Tamaki Ura. "Generation of High-resolution Three-dimensional Reconstructions of the Seafloor in Color using a Single Camera and Structured Light". In: *Journal of Field Robotics* 34.5 (2017), pages 833–851. DOI: 10.1002/rob.21682 (cited on page 13).

[90] A. Sedlazeck, K. Koser, and R. Koch. "3D reconstruction based on underwater video from ROV Kiel 6000 considering underwater imaging conditions". In: *OCEANS 2009-EUROPE* (Bremen, Germany). May 2009, pages 1–10. DOI: 10.1109/OCEANSE.2009.5278305 (cited on page 13).

[91] O. Pizarro, R. M. Eustice, and H. Singh. "Large Area 3-D Reconstructions From Underwater Optical Surveys". In: *IEEE Journal of Oceanic Engineering* 34.2 (Apr. 2009), pages 150–169. DOI: 10.1109/JOE.2009.2016071 (cited on page 13).

[92] R Garcia, R Campos, and J Escartín. "High-resolution 3D reconstruction of the seafloor for environmental monitoring and modelling". In: *Proc. Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on* (San Francisco, CA, USA). Sept. 2011 (cited on page 13).

[93] David McKinnon, Hu He, Ben Upcroft, and Ryan N Smith. "Towards automated and in-situ, near-real time 3-D reconstruction of coral reef environments". In: *OCEANS'11 MTS/IEEE KONA*. IEEE. 2011, pages 1–10. DOI: 10.23919/OCEANS.2011.6106982 (cited on page 13).

[94] Katleen Robert, Veerle AI Huvenne, Aggeliki Georgiopoulou, Daniel OB Jones, Leigh Marsh, Gareth DO Carter, and Leo Chaumillon. "New approaches to high-resolution mapping of marine vertical structures". In: *Scientific reports* 7.1 (2017), page 9005. DOI: 10.1038/s41598-017-09382-z (cited on pages 13, 14).

[95] Fabio Bruno, A Lagudi, Alessandro Gallo, M Muzzupappa, B Davidde Petriaggi, and S Passaro. "3D documentation of archeological remains in the underwater park of Baiae". In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40.5 (2015), page 41. DOI: 10.5194/isprsarchives-XL-5-W5-41-2015 (cited on page 13).

[96] S Cocito, S Sgorbini, A Peirano, and M Valle. "3-D reconstruction of biological objects using underwater video technique and image processing". In: *Journal of Experimental Marine Biology and Ecology* 297.1 (2003), pages 57–70. DOI: 10.1016/S0022-0981(03)00369-1 (cited on page 13).

[97] Margaret Kalacska, Oliver Lucanus, Leandro Sousa, Thiago Vieira, and Juan Arroyo-Mora. "Freshwater fish habitat complexity mapping using above and underwater structure-from-motion photogrammetry". In: *Remote Sensing* 10.12 (2018), page 1912. DOI: 10.3390/rs10121912 (cited on page 13).

[98] F. Neyer, E. Nocerino, and A. Gruen. "Monitoring coral growth - The dichotomy between underwater photogrammetry and geodetic control network". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2018), page 2. DOI: 10.5194/isprs-archives-XLII-2-759-2018 (cited on page 13).

[99] Miran Erič, Gregor Berginc, Rok Kovačič, Mitja Pugelj, and Franc Solina. "Successful use of temporary underwater 3D documenting methodology: Early Roman barge from Ljubljanica river, Slovenia". In: (2016) (cited on page 13).

[100]   John McCarthy and Jonathan Benjamin. "Multi-image photogrammetry for underwater archaeological site recording: an accessible, diver-based approach". In: *Journal of maritime archaeology* 9.1 (2014), pages 95–114. DOI: `10.1007/s11457-014-9127-7` (cited on page 13).

[101]   Stella Demesticha, Dimitrios Skarlatos, and Andonis Neophytou. "The 4th-century BC shipwreck at Mazotos, Cyprus: new techniques and methodologies in the 3D mapping of shipwreck excavations". In: *Journal of Field Archaeology* 39.2 (2014), pages 134–150. DOI: `10.1179/0093469014Z.00000000077` (cited on page 13).

[102]   Chris Beall, Brian J Lawrence, Viorela Ila, and Frank Dellaert. "3D reconstruction of underwater structures". In: *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on.* IEEE. 2010, pages 4418–4423. DOI: `10.1109/IROS.2010.5649213` (cited on pages 13, 15).

[103]   Natalia Hurtós, Sharad Nagappa, Narcis Palomeras, and Joaquim Salvi. "Real-time mosaicing with two-dimensional forward-looking sonar". In: *2014 IEEE International Conference on Robotics and Automation (ICRA).* IEEE. 2014, pages 601–606. DOI: `10.1109/ICRA.2014.6906916` (cited on pages 13, 14).

[104]   Manolis Lourakis and Xenophon Zabulis. "Accurate scale factor estimation in 3D reconstruction". In: *International Conference on Computer Analysis of Images and Patterns.* Springer. 2013, pages 498–506. DOI: `10.1007/978-3-642-40261-6_60` (cited on pages 13, 108).

[105]   Gunnar Taraldsen, Tor Arne Reinen, and Tone Berg. "The underwater GPS problem". In: *OCEANS 2011 IEEE-Spain.* IEEE. 2011, pages 1–8. DOI: `10.1109/Oceans-Spain.2011.6003649` (cited on page 13).

[106]   Dean M Steinke and Bradley J Buckham. "A Kalman filter for the navigation of remotely operated vehicles". In: *Proceedings of OCEANS 2005 MTS/IEEE.* IEEE. 2005, pages 581–588. DOI: `10.1109/OCEANS.2005.1639817` (cited on page 13).

[107]   Ricard Campos, Nuno Gracias, and Pere Ridao. "Underwater Multi-Vehicle Trajectory Alignment and Mapping Using Acoustic and Optical Constraints". In: *Sensors* 16.3 (2016), page 387. DOI: `10.3390/s16030387` (cited on page 13).

[108]   Dimitrios Skarlatos, Fabio Menna, Erica Nocerino, and Panagiotis Agrafiotis. "Precision potential of underwater networks for archaeological excavation through trilateration and photogrammetry". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42.2/W10 (2019), pages 175–180. DOI: `10.5194/isprs-archives-XLII-2-W10-175-2019` (cited on page 13).

[109]   Melanie Bergmann, Nina Langwald, Jorg Ontrup, Thomas Soltwedel, Ingo Schewe, Michael Klages, and Tim W. Nattkemper. "Megafaunal assemblages from two shelf stations west of Svalbard". In: *Marine Biology Research* 7.6 (2011), pages 525–539. DOI: `10.1080/17451000.2010.535834` (cited on page 14).

[110]   D. M. Kocak, T. H. Jagielo, F. Wallace, and J. Kloske. "Remote sensing using laser projection photogrammetry for underwater surveys". In: *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium.* Volume 2. Sept. 2004, 1451–1454 vol.2. DOI: `10.1109/IGARSS.2004.1368693` (cited on page 14).

[111]   D. A. Pilgrim, D. M. Parry, M. B. Jones, and M. A. Kendall. "ROV Image Scaling with Laser Spot Patterns". In: *Underwater Technology* 24.3 (2000), pages 93–103. DOI: `10.3723/175605400783259684` (cited on page 14).

[112]   Robert F Tusting and DL Davis. "Improved methods for visual and photographic benthic surveys". In: *Diving for Science* (1993), pages 157–172 (cited on page 14).

[113]   Francisco Bonin-Font, Aleksandar Cosic, Pep Lluis Negre, Markus Solbach, and Gabriel Oliver. "Stereo SLAM for robust dense 3D reconstruction of underwater environments". In: *OCEANS 2015-Genova*. IEEE. 2015, pages 1–6. DOI: `10.1109/OCEANS-Genova.2015.7271333` (cited on pages 14, 15).

[114]   James Servos, Michael Smart, and Steven Lake Waslander. "Underwater stereo SLAM with refraction correction". In: *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE. 2013, pages 3350–3355. DOI: `10.1109/IROS.2013.6696833` (cited on page 15).

[115]   Val E Schmidt and Yuri Rzhanov. "Measurement of micro-bathymetry with a GO-PRO underwater stereo camera pair". In: *Oceans, 2012*. IEEE. 2012, pages 1–6. DOI: `10.1109/OCEANS.2012.6404786` (cited on page 15).

[116]   Stein M Nornes, Martin Ludvigsen, Øyvind Ødegard, and Asgeir J SØrensen. "Underwater Photogrammetric Mapping of an Intact Standing Steel Wreck with ROV". In: *IFAC-PapersOnLine* 48.2 (2015), pages 206–211. DOI: `10.1016/j.ifacol.2015.06.034` (cited on page 15).

[117]   Stefan B Williams, Oscar R Pizarro, Michael V Jakuba, Craig R Johnson, Neville S Barrett, Russell C Babcock, Gary A Kendrick, Peter D Steinberg, Andrew J Heyward, Peter J Doherty, et al. "Monitoring of benthic reference sites: using an autonomous underwater vehicle". In: *Robotics & Automation Magazine, IEEE* 19.1 (2012), pages 73–84. DOI: `10.1109/MRA.2011.2181772` (cited on page 15).

[118]   Andrew Hogue, Andrew German, and Michael Jenkin. "Underwater environment reconstruction using stereo and inertial data". In: *2007 IEEE International Conference on Systems, Man and Cybernetics*. IEEE. 2007, pages 2372–2377. DOI: `10.1109/ICSMC.2007.4413666` (cited on page 15).

[119]   Ian Mahon, Stefan B Williams, Oscar Pizarro, and Matthew Johnson-Roberson. "Efficient view-based SLAM using visual loop closures". In: *Robotics, IEEE Transactions on* 24.5 (2008), pages 1002–1014. DOI: `10.1109/TRO.2008.2004888` (cited on page 15).

[120]   Vincent Brandou, Anne-gaelle Allais, Michel Perrier, Ezio Malis, Patrick Rives, Jozee Sarrazin, and Pierre-Marie Sarradin. "3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS". In: *OCEANS'07 IEEE Aberdeen Conf. Proceedings (IEEE)*. IEEE. 2007, pages 1–6. DOI: `10.1109/OCEANSE.2007.4302315` (cited on page 15).

[121]   G Bianco, A Gallo, F Bruno, and M Muzzupappa. "A comparison between active and passive techniques for underwater 3D applications". In: *International Archives of the Photogrammetry. In: Remote sensing and spatial information sciences* 38 (2011). DOI: `10.5194/isprsarchives-XXXVIII-5-W16-357-2011` (cited on page 15).

[122] Maxime Ferrera, Vincent Creuze, Julien Moras, and Pauline Trouvé-Peloux. "AQUALOC: An underwater dataset for visual–inertial–pressure localization". In: *The International Journal of Robotics Research* 38.14 (2019), pages 1549–1559. DOI: `10.1177/0278364919883346` (cited on page 15).

[123] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age". In: *IEEE Transactions on robotics* 32.6 (2016), pages 1309–1332. DOI: `10.1109/TRO.2016.2624754` (cited on page 15).

[124] Takafumi Taketomi, Hideaki Uchiyama, and Sei Ikeda. "Visual SLAM algorithms: A survey from 2010 to 2016". In: *IPSJ Transactions on Computer Vision and Applications* 9.1 (2017), page 16. DOI: `10.1186/s41074-017-0027-2` (cited on page 15).

[125] Y Kwon. "A camera calibration algorithm for the underwater motion analysis". In: *ISBS-Conference Proceedings Archive*. Volume 1. 1. 1999 (cited on page 15).

[126] Alejo Concha, Paulo Drews-Jr, Mario Campos, and Javier Civera. "Real-time localization and dense mapping in underwater environments from a monocular sequence". In: *OCEANS 2015-Genova*. IEEE. 2015, pages 1–5. DOI: `10.1109/OCEANS-Genova.2015.7271476` (cited on page 15).

[127] Jun Zhang, Viorela Ila, and Laurent Kneip. "Robust Visual Odometry in Underwater Environment". In: *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*. IEEE. 2018, pages 1–9. DOI: `10.1109/OCEANSKOBE.2018.8559452` (cited on page 15).

[128] Liam Paull, Sajad Saeedi, Mae Seto, and Howard Li. "AUV navigation and localization: A review". In: *IEEE Journal of Oceanic Engineering* 39.1 (2013), pages 131–149. DOI: `10.1109/JOE.2013.2278891` (cited on page 15).

[129] Ryan M Eustice, Oscar Pizarro, and Hanumant Singh. "Visually augmented navigation for autonomous underwater vehicles". In: *IEEE Journal of Oceanic Engineering* 33.2 (2008), pages 103–122. DOI: `10.1109/JOE.2008.923547` (cited on page 15).

[130] Matthew Johnson-Roberson, Oscar Pizarro, Stefan B Williams, and Ian Mahon. "Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys". In: *Journal of Field Robotics* 27.1 (2010), pages 21–51. DOI: `10.1002/rob.20324` (cited on page 15).

[131] Ayoung Kim and Ryan M Eustice. "Real-time visual SLAM for autonomous underwater hull inspection using visual saliency". In: *IEEE Transactions on Robotics* 29.3 (2013), pages 719–733. DOI: `10.1109/IROS.2010.5653696` (cited on page 15).

[132] Sharmin Rahman, Alberto Quattrini Li, and Ioannis Rekleitis. "SVIn2: An Underwater SLAM System using Sonar, Visual, Inertial, and Depth Sensor,"". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019, pages 1861–1868. DOI: `10.1109/IROS40897.2019.8967703` (cited on page 15).

[133]   Florian Shkurti, Ioannis Rekleitis, Milena Scaccia, and Gregory Dudek. "State estimation of an underwater robot using visual and inertial information". In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2011, pages 5054–5060. DOI: `10.1109/IROS.2011.6048793` (cited on page 15).

[134]   Antoni Burguera, Francisco Bonin-Font, and Gabriel Oliver. "Trajectory-based visual localization in underwater surveying missions". In: *Sensors* 15.1 (2015), pages 1708–1735. DOI: `10.3390/s150101708` (cited on page 15).

[135]   Hauke Strasdat, J Montiel, and Andrew J Davison. "Scale drift-aware large scale monocular SLAM". In: *Robotics: Science and Systems VI* 2.3 (2010), page 7. DOI: `10.15607/RSS.2010.VI.010` (cited on page 15).

[136]   Robert Schattschneider, Guillaume Maurino, and Wenhui Wang. "Towards stereo vision slam based pose estimation for ship hull inspection". In: *OCEANS'11 MTS/IEEE KONA*. IEEE. 2011, pages 1–8. DOI: `10.23919/OCEANS.2011.6106988` (cited on page 15).

[137]   Joaquim Salvi, Yvan Petillot, and Elisabet Batlle. "Visual SLAM for 3D large-scale seabed acquisition employing underwater vehicles". In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2008, pages 1011–1016. DOI: `10.1109/IROS.2008.4650627` (cited on page 15).

[138]   Stephen M Chaves, Enric Galceran, Paul Ozog, Jeffrey M Walls, and Ryan M Eustice. "Pose-Graph SLAM for Underwater Navigation". In: *Sensing and Control for Autonomous Vehicles*. Springer, 2017, pages 143–160. DOI: `10.1007/978-3-319-55372-6_7` (cited on page 15).

[139]   Pep Lluis Negre Carrasco, Francisco Bonin-Font, and Gabriel Oliver Codina. "Stereo Graph-SLAM for autonomous underwater vehicles". In: *Intelligent Autonomous Systems 13*. Springer, 2016, pages 351–360. DOI: `10.1007/978-3-319-08338-4_26` (cited on page 15).

[140]   Giorgio Grisetti, Rainer Kummerle, Cyrill Stachniss, and Wolfram Burgard. "A tutorial on graph-based SLAM". In: *IEEE Intelligent Transportation Systems Magazine* 2.4 (2010), pages 31–43. DOI: `10.1109/MITS.2010.939925` (cited on page 15).

[141]   Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. "ORB-SLAM: a versatile and accurate monocular SLAM system". In: *IEEE transactions on robotics* 31.5 (2015), pages 1147–1163. DOI: `10.1109/TRO.2015.2463671` (cited on page 15).

[142]   Nikola Mišković, Zoran Triska, Đula Nađ, and Zoran Vukić. "Guidance of a small-scale overactuated marine platform—Experimental results". In: *2011 Proceedings of the 34th International Convention MIPRO*. IEEE. 2011, pages 684–689 (cited on page 16).

[143]   Ricard Prados, Raul Garcia, Nuno Gracias, Javier Escartin, and Luka Neumann. "A novel blending technique for underwater gigamosaicing". In: *Oceanic Engineering, IEEE Journal of* 37.4 (2012), pages 626–644. DOI: `10.1109/JOE.2012.2204152` (cited on page 111).

[144]   Pulung Nurtantio Andono, I Purnama, and Mochamad Hariadi. "Underwater image enhancement using adaptive filtering for enhanced sift-based image matching". In: *Journal of Theoretical & Applied Information Technology* 52.3 (2013). DOI: `10.1117/12.2503527` (cited on page 115).

[145] Gustau Vivar, Carlos de Juan Fuertes, and Ruth Geli. "Cap del Vol, un producto, un barco y un comercio del Conventus Tarraconensis en época de Augusto". In: *Actas I Congreso de Arqueologia Naútica y Subacuática Espanola*. 2014 (cited on page 115).

[146] Mitch Bryson, Matthew Johnson-Roberson, Oscar Pizarro, and Stefan B Williams. "True color correction of autonomous underwater vehicle imagery". In: *Journal of Field Robotics* 33.6 (2016), pages 853–874. DOI: `10.1002/rob.21638` (cited on page 116).

[147] Derya Akkaynak and Tali Treibitz. "Sea-thru: A method for removing water from underwater images". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pages 1682–1691. DOI: `10.1109/CVPR.2019.00178` (cited on page 116).

[148] Codruta O Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, Laszlo Neumann, and Rafael Garcia. "Color transfer for underwater dehazing and depth estimation". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pages 695–699. DOI: `10.1109/ICIP.2017.8296370` (cited on page 116).

[149] Jie Li, Katherine A Skinner, Ryan M Eustice, and Matthew Johnson-Roberson. "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images". In: *IEEE Robotics and Automation letters* 3.1 (2017), pages 387–394. DOI: `10.1109/LRA.2017.2730363` (cited on page 116).

[150] Marino Mangeruga, Fabio Bruno, Marco Cozza, Panagiotis Agrafiotis, and Dimitrios Skarlatos. "Guidelines for underwater image enhancement based on benchmarking of different methods". In: *Remote Sensing* 10.10 (2018), page 1652. DOI: `10.3390/rs10101652` (cited on page 119).

[151] Rongxing Li, Chuang Tao, Weihong Zou, RG Smith, and TA Curran. "An underwater digital photogrammetric system for fishery geomatics". In: *International Archives of Photogrammetry and Remote Sensing* 31 (1996), pages 319–323 (cited on page 122).

[152] Mark Shortis. "Calibration techniques for accurate measurements by underwater camera systems". In: *Sensors* 15.12 (2015), pages 30810–30826. DOI: `10.3390/s151229831` (cited on page 122).

[153] C Brown Duane. "Close-range camera calibration". In: *Photogramm. Eng* 37.8 (1971), pages 855–866 (cited on pages 122, 127).

[154] J-Y Bouguet. *Camera calibration toolbox for matlab*. `http://www.vision.caltech.edu/bouguetj/calib_doc/index.html`. [Online; accessed 22-January-2018]. 2004 (cited on page 122).

[155] Mark Shortis. "Camera calibration techniques for accurate measurement underwater". In: *3D recording and interpretation for maritime archaeology* (2019), pages 11–27. DOI: `10.1007/978-3-030-03635-5_2` (cited on page 122).

[156] Christian Bräuer-Burchardt, Christoph Munkelt, Ingo Gebhart, Matthias Heinze, Stefan Heist, Peter Kühmstedt, and Gunther Notni. "A-Priori Calibration of a Structured Light Underwater 3D Sensor". In: *Journal of Marine Science and Engineering* 8.9 (2020), page 635. DOI: `10.3390/jmse8090635` (cited on page 122).

[157] R.F. Tusting and D.L. Davis. "Non-conventional techniques for sampling and collecting marine organisms". In: *Proceedings of the Pacific Congress on Marine Technology, PACON'86*. 1986, pages 12–18 (cited on page 123).

[158] F. M. Caimi and R. F. Tusting. "Application of lasers to ocean research and image recording systems". In: *Proceedings of the International Conference on LASERS*. Volume 86. STS Press McLean, Virginia. 1987, pages 518–524 (cited on page 123).

[159] R. F. Tusting and D. L. Davis. "Laser systems and structured illumination for quantitative undersea imaging". In: *Marine Technology Society Journal* 26.4 (1992), pages 5–12 (cited on page 123).

[160] Eric A Wan and Rudolph Van Der Merwe. "The unscented Kalman filter for nonlinear estimation". In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. IEEE. 2000, pages 153–158. DOI: 10.1109/ASSPCC.2000.882463 (cited on page 124).

[161] Richard Camilli, Paraskevi Nomikou, Javier Escartín, Pere Ridao, Angelos Mallios, Stephanos P Kilias, Ariadne Argyraki, Muriel Andreani, Valerie Ballu, Ricard Campos, et al. "The Kallisti Limnes, carbon dioxide-accumulating subsea pools". In: *Scientific reports* 5 (2015), page 12152. DOI: 10.1038/srep12152 (cited on page 124).

[162] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. "g2o: A general framework for graph optimization". In: *2011 IEEE International Conference on Robotics and Automation*. IEEE. 2011, pages 3607–3613. DOI: 10.1109/ICRA.2011.5979949 (cited on page 134).

[163] Sameer Agarwal, Keir Mierle, et al. *Ceres Solver*. http://ceres-solver.org (cited on page 134).

[164] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. "Deep learning vs. traditional computer vision". In: *Science and Information Conference*. Springer. 2019, pages 128–144. DOI: 10.1007/978-3-030-17795-9_10 (cited on page 152).