# Visual 3D Reconstruction in the Deep Sea

from Geometric and Radiometric Perspectives

*by*

**Yifan Song**

**Dissertation**

zur Erlangung des akademischen Grades Doktor der Ingenieurwissenschaften

*(Dr.-Ing.)*

der Technischen Fakultät

der Christian-Albrechts-Universität zu Kiel

und

der GEOMAR - Helmholtz-Zentrum für Ozeanforschung Kiel

eingereicht im Jahr 2023

Prüfungskommission:

1. Gutachter / Betreuer: Prof. Dr.-Ing. Kevin Köser

2. Gutachter: Prof. Dr.-Ing. Reinhard Koch

3. Gutachter: Prof. Dr. rer. nat. Sören Pirk

Vorsitz: Prof. Dr. rer. nat. Matthias Renz

Datum der mündlichen Prüfung: 02.05.2024

# Abstract

With the rapid development and increased availability of underwater imaging technologies, underwater vision systems are now widely applied in ocean research. While much of the photogrammetric and computer vision literature has focused on shallow water applications, there is a growing interest in deep sea mapping research. The vast majority of the seafloor, and Earth's surface, lies in the deep ocean below 200 meters depth, and this region remains largely unexplored. Imaging in the deep sea presents two major challenges. Firstly, cameras housed in waterproof enclosures must withstand the extremely high water pressure. Light travels through multiple media layers, leading to refraction, which hinders the direct use of standard photogrammetry techniques that rely on light traveling in straight lines. Secondly, light is absorbed and scattered in water, resulting in image quality degradation. On top of it, artificial illumination is essential for deep sea imaging, as sunlight does not penetrate to such depths. This artificial lighting introduces unique visual effects in the images that cannot be addressed using current shallow water image processing solutions. Furthermore, the strong absorption of electromagnetic radiation from navigation satellites in water prohibits the use of satellite signals, making navigation in deep water more challenging and less accurate. All these factors collectively contribute to the complexity and difficulty of visual 3D reconstruction in the deep sea. Addressing these challenges requires innovative and specialized approaches tailored to the unique conditions of deep sea environments.

This dissertation presents comprehensive solutions to address the challenges and problems encountered in deep sea visual 3D reconstruction, covering both hardware and software aspects. Deep sea cameras are typically protected in robust housings with thick transparent windows (e.g., glass or sapphire) to withstand the harsh conditions of saltwater and high pressures. To effectively tackle the issue of geometric refraction caused by multiple media interfaces, the dissertation analyzes the geometrical properties of dome refraction and explores the advantages of using dome port windows. A systematic approach to mitigate refraction effects simplifies subsequent procedures, enabling the application of standard photogrammetry pipelines. Additionally, considering that flat port interfaces are still commonly used in ocean exploration tasks, this dissertation discusses effective methods to remove flat refraction effects in underwater photogrammetry. By addressing geometric challenges, the proposed methods ensure more accurate and reliable 3D reconstructions in deep-sea environments.

Furthermore, the dissertation explores radiometric problems arising from light absorption and scattering in water, particularly in the context of artificial illumination in deep-sea imaging. It delves into physical models of underwater image formation, understanding the realistic rendering of underwater scenes, and considering image restoration as its inverse procedure. A general model suitable for restoring various types of underwater images, especially in deep-sea environments with complex artificial illumination, is developed. This model allows for the simultaneous removal of artificial lighting patterns and the restoration of true colors, providing significant benefits for visual mapping purposes.

Moreover, the dissertation demonstrates several successful applications of the techniques developed within its scope. It first showcases the effectiveness of the proposed dome port centering techniques in various camera system designs, ensuring their suitability for accurate 3D underwater reconstruction. Then several underwater 3D reconstruction cases are presented, demonstrating the practical applications of the methods described in this dissertation. Lastly, the dissertation

highlights the usage of the proposed deep sea image simulation techniques for improving complex illumination configurations within imaging systems to enhance the quality of captured images.

In summary, this dissertation offers comprehensive solutions to overcome unique geometric and radiometric challenges in deep-sea visual 3D reconstruction. The proposed methods have been applied to various systems and tasks, enhancing the accuracy and efficiency of deep sea imaging systems for underwater mapping and exploration.

# Zusammenfassung

Mit der rasanten Entwicklung und der zunehmenden Verfügbarkeit von Unterwasserbildgebungstechnologien werden Unterwasserbildverarbeitungssysteme nun in großem Umfang in der Meeresforschung eingesetzt. Während sich ein Großteil der Fachliteratur zu Photogrammetrie und Computer Vision auf Anwendungen in flachen Gewässern konzentriert, wächst das Interesse an der Erforschung von Tiefseegebieten. Der größte Teil des Meeresbodens, und damit der Erdoberfläche, befindet sich in der Tiefsee unterhalb von 200 Metern und diese Region ist noch weitgehend unerforscht. Die Bildgebung in der Tiefsee stellt zwei große Herausforderungen dar. Zum einen müssen die in wasserdichten Gehäusen untergebrachten Kameras dem extrem hohen Wasserdruck standhalten. Da das Licht mehrere Medienschichten durchläuft kommt es zu Brechungen, die den direkten Einsatz von Standard-Fotogrammetrietechniken, die auf einer geradlinigen Ausbreitung des Lichts basieren, behindern. Zweitens wird Licht im Wasser absorbiert und gestreut, was zu einer Verschlechterung der Bildqualität führt. Darüber hinaus ist eine künstliche Beleuchtung für Tiefseeaufnahmen unerlässlich, da das Sonnenlicht nicht bis in solche Tiefen vordringt. Diese künstliche Beleuchtung führt zu einzigartigen visuellen Effekten in den Bildern, die mit den derzeitigen Lösungen für die Bildverarbeitung in flachen Gewässern nicht erfasst werden können. Außerdem verhindert die starke Absorption der elektromagnetischen Strahlung von Navigationssatelliten im Wasser die Nutzung von Satellitensignalen, was die Navigation in der Tiefsee erschwert und ungenauer macht. All diese Faktoren zusammengenommen tragen zur Komplexität und Schwierigkeit der visuellen 3D-Rekonstruktion in der Tiefsee bei. Die Lösung dieser Herausforderungen erfordert innovative und spezialisierte Ansätze, die auf die besonderen Bedingungen der Tiefsee zugeschnitten sind.

In dieser Dissertation werden umfassende Lösungen für die Herausforderungen und Probleme der visuellen 3D-Rekonstruktion in der Tiefsee vorgestellt, die sowohl Hardware- als auch Softwareaspekte umfassen. Tiefseekameras sind in der Regel in robusten Gehäusen mit dicken transparenten Fenstern (z. B. Glas oder Saphir) untergebracht, um den rauen Bedingungen des Salzwassers und des hohen Drucks standzuhalten. Um das Problem der geometrischen Refraktion, das durch mehrere Medienschnittstellen verursacht wird, effektiv anzugehen, werden in der Dissertation die geometrischen Eigenschaften der Kuppelrefraktion analysiert und die Vorteile der Verwendung von kugelförmigen Glasinterfaces untersucht. Ein systematischer Ansatz zur Minderung von Brechungseffekten vereinfacht die anschließenden Verfahren und ermöglicht die Verwendung von herkömmlichen Photogrammetrieverfahren. In Anbetracht der Tatsache, dass flache Glasinterfaces in der Meeresforschung immer noch häufig verwendet werden, werden in dieser Dissertation effektive Methoden zur Beseitigung von flachen Refraktionseffekten in der Unterwasserphotogrammetrie diskutiert. Durch die Lösung geometrischer Herausforderungen gewährleisten die vorgeschlagenen Methoden genauere und zuverlässigere 3D-Rekonstruktionen in Tiefseeumgebungen.

Darüber hinaus werden in der Dissertation radiometrische Probleme untersucht, die sich aus der Lichtabsorption und -streuung im Wasser ergeben, insbesondere im Zusammenhang mit der künstlichen Beleuchtung in der Tiefseebildgebung. Es wird auf physikalische Modelle der Unterwasserbildentstehung eingegangen, die realistische Abbildung von Unterwasserszenen betrachtet und die Bildrestaurierung als verwandtes Problem, aber in umgekehrter Richtung, betrachtet. Es wurde ein allgemeines Modell entwickelt, das sich für die Wiederherstellung verschiedener Arten

von Unterwasserbildern eignet, insbesondere in Tiefseeumgebungen mit komplexer künstlicher Beleuchtung. Dieses Modell ermöglicht die gleichzeitige Entfernung von künstlichen Beleuchtungsmustern und die Wiederherstellung der echten Farben, wodurch sich erhebliche Vorteile für die visuelle Kartierung ergeben.

Darüber hinaus werden in der Dissertation mehrere erfolgreiche Anwendungen der in ihrem Rahmen entwickelten Techniken demonstriert. Zunächst wird die Wirksamkeit der vorgeschlagenen Techniken zur Zentrierung von Kuppelöffnungen in verschiedenen Kamerasystemen aufgezeigt, um deren Eignung für eine genaue 3D-Unterwasserrekonstruktion zu gewährleisten. Dann wird die Verwendung der vorgeschlagenen Tiefseebildsimulationstechniken zur Verbesserung komplexer Beleuchtungskonfigurationen in Bildgebungssystemen hervorgehoben, um die Qualität der aufgenommenen Bilder zu verbessern. Schließlich werden mehrere Fälle von 3D-Unterwasser-Rekonstruktionen vorgestellt, die die praktische Anwendung der in dieser Dissertation beschriebenen Methoden demonstrieren.

Zusammenfassend bietet diese Dissertation umfassende Lösungen zur Bewältigung der einzigartigen geometrischen und radiometrischen Herausforderungen bei der visuellen 3D Rekonstruktion in der Tiefsee. Die vorgeschlagenen Methoden wurden in verschiedenen Systemen und Funktionen eingesetzt, um die Genauigkeit und Effizienz von Tiefseebildgebungssystemen für die Unterwasserkartierung und -erkundung zu verbessern.

# List of Abbreviations

Area Of Interest (AOI)
Attenuation-Curve Prior (ACP)
Autonomous Underwater Vehicle (AUV)
Background Light (BL)
Color-Lines Prior (CLP)
Contrast Limited Adaptive Histogram Equalization (CLAHE)
Dark Channel Prior (DCP)
Doppler Velocity Log (DVL)
Field Of View (FOV)
Ground Truth (GT)
Haze-Lines Prior (HLP)
Henyey-Greenstein (HG) function
Human Occupied Vehicle (HOV)
Independent Component Analysis (ICA)
Inertial Measurement Unit (IMU)
Inherent Optical Property (IOP)
Joint Trilateral Filter (JTF)
Light Detection And Ranging (LiDAR)
Light Emitting Diode (LED)
Maximum Intensity Prior (MIP)
Modulation Transfer Function (MTF)
Multi-View Stereo (MVS)
non Single View Point (nSVP)
Point Spread Function (PSF)
Radiation Intensity Distribution (RID)
Remotely Operated Vehicle (ROV)
Signal-to-Noise Ratio (SNR)
Simultaneous Localization And Mapping (SLAM)
Structure from Motion (SfM)
the Atmospheric Fog (AF) Model
the Jaffe-McGlamery (J-M) Model
Underwater DCP (UDCP)
Underwater Vehicle (UV)
Unmanned Underwater Vehicle (UUV)
Volume Scattering Function (VSF)

# Symbols and Notation

absorption coefficient: $a$
backscatter coefficient: $b_b$
beam attenuation coefficient: $c$
camera matrix: $\mathbf{K}$
camera optical center $\mathbf{C}$
decentering offset vector $\mathbf{v}_{off}$
dome center $\mathbf{o}$
dome sphere $\mathbf{D}$
focal length: $f$
irradiance: $E$
n-dimensional Euclidean space $\mathbb{R}^n$
n-dimensional projective space $\mathbb{P}^n$
normal vector of the surface: $\mathbf{n}$
pixel intensity: $I$
point in the local camera coordinate: $\mathbf{X}_c$, and its homogeneous coordinate: $\widetilde{\mathbf{X}_c}$
point in the pixel coordinate: $\mathbf{x}$, and its homogeneous coordinate: $\widetilde{\mathbf{x}}$
point in the world coordinate: $\mathbf{X}$, and its homogeneous coordinate: $\widetilde{\mathbf{X}}$
refractive intersection point $\mathbf{I}$
rotation matrix: $\mathbf{R}$
scattered light angle from the initial direction: $\psi$
scattering coefficient: $b$
scattering phase function: $\widetilde{\beta}_{vsf}$
translation vector: $\mathbf{t}$
unit sphere $\mathbf{Q}$.
viewing ray direction $\mathbf{L}$
volume scattering function: $\beta_{vsf}$
wavelength: $\lambda$

# Contents

# 1 Introduction

乘风破浪会有时，直挂云帆济沧海。

——《行路难》[唐] 李白

I will mount a long wind some day and break the heavy waves, And set my cloudy sail straight and bridge the deep, deep sea. —— Li Bai (AD 701–762)

Optical imaging stands out as an attractive technology for sensing and exploration due to its high resolution, interpretability, and non-contact data collection capabilities. The rapid progress in optical imaging has paved the way for excellent quality photogrammetric surveys both on land and in space, with hundreds of satellites and airborne imaging platforms continuously updating high-resolution imagery, optical imaging plays a pivotal role in modern society, enabling humans to even map the surface of the Moon and Mars. However, significant knowledge gaps remain regarding our own planet, human knowledge about the underwater world is relatively limited compared to what is known about terrestrial environments.

Over 70% of Earth's surface is covered by water, with more than 90% of it being deeper than 200 meters, shrouded in perpetual darkness. However, the exploration of the deep sea remains severely limited due to the challenges of accessing these depths. It necessitates specialized equipment and vehicles capable of withstanding extremely high water pressure and functioning in navigation satellite-denied environments. Moreover, limited visibility underwater restricts deep sea imaging to a few meters range, resulting in small-footprint coverage and time-consuming expeditions that demand meticulous energy budgeting, planning and coordinating.

Figure 1.1: GEOMAR AUV ANTON (Girona 500) performing subsea visual mapping tasks in the darkness with its own lighting offshore. The co-moving light source creates a light cone in water, illuminates the seafloor non-homogeneously and forms up an artificial pattern in the image.

Figure 1.2: Examples of four different types of platforms used for deep-sea mapping. From left to right: ROV (GEOMAR KIEL 6000), AUV (GEOMAR ABYSS Tiffy), HOV (GEOMAR JAGO) and towed platform (OFOS frame).

In addition to these technical challenges, optical imaging in the deep ocean faces specific obstacles. Cameras protected in housings with transparent windows must withstand the extremely high water pressure, causing refraction distortion as light travels through multiple media layers. Consequently, the standard photogrammetry theory, which relies on light traveling in straight lines, is no longer applicable. Adequate artificial lighting is also necessary to illuminate the scenes in perpetual darkness, introducing water attenuation, scattering, and lighting patterns that degrade the quality of 3D mapping. Effective image restoration algorithms are indispensable in removing these unwanted water effects and producing high-quality data products.

Given the unique characteristics and challenges presented by the deep sea environment, current visual mapping solutions are not directly applicable. This thesis thoroughly analyzes the current challenges and explores solutions for achieving high-quality visual 3D mapping in the deep sea.

## 1.1 Deep Sea Exploration Platforms

To perform visual mapping in the deep ocean, imaging systems have to be brought to location and navigated to scan a survey area. The dynamic platforms used for deep sea operations can be classified into four basic types: Remotely Operated Vehicles (ROVs), Autonomous Underwater Vehicles (AUVs), Human Occupied Vehicles (HOVs), and towed platforms (Fig. 1.2). They can be further categorized into two groups based on their power supply: cabled and uncabled platforms.

Cabled platforms, such as ROVs (Drap et al. 2015; Johnson-Roberson et al. 2010), are connected to operating ships or surface stations through underwater cables. All control commands and signals are transmitted between the platforms and operators via these cables. Towed platforms, on the other hand, can either be remotely powered and transmit signals directly to the support vessel via cables or operate independently of the ship. Additionally, more passive, towsleds are also often used for deep ocean imaging: they can either be remotely powered and transmit signals directly to the support vessel via cables (Barker et al. 1999; Lembke et al. 2017; Purser et al. 2018), or operate independently of the ship (Fornari et al. 2003; Jones et al. 2009). Uncabled platforms, including AUVs (Iscar et al. 2018; Kunz et al. 2008a; Singh et al. 2004b; Yoerger et al. 2007) and HOVs, are untethered underwater vehicles that rely on battery power for operation. AUVs are unoccupied underwater robots fully controlled by their onboard computers, while HOVs are crewed craft that transport a few passengers directly underwater for limited periods. They have limited

deployment endurance that mainly depends on the platforms' energy budget (or living supplies for HOVs).

Because of the poor underwater visibility conditions, all these platforms have to be operated close to the seafloor, leading to small footprints and mapping speeds of a hectare per hour or less (Kwasnitschka et al. 2016). Also deep dives to great depths takes several hours, making large-scale benthic visual maps require very long missions. ROVs, HOVs and towed platforms all demand labor intensive operation, and when cables are in the water, careful coordination is essential to avoid the surface vessel's propellers. Safety guidelines often restrict using more than one cabled device simultaneously, making parallelization difficult. AUVs offer more flexibility, as multiple fully automatic vehicles can work together or in parallel for extensive deep ocean mapping tasks.

## 1.2 Overview of Deep Sea Visual Mapping

Deep sea imaging has a long history since the Second World War, with (Harvey 1939) conducting pioneering work using a pressure chamber to endure two miles of water depth. Early systems were primarily focused on inspection and exploration, but later evolved to include mapping purposes. Examples of early applications include deep sea photo mosaics used to visualize the sunken submarine *Thresher* (Ballard 1975) and the famous sunken ship *Titanic* (Ballard et al. 1987). At that time, digital image processing was not available, researchers manually pieced photos together to create larger mosaics. In modern times, quantitative underwater visual mapping has found a wide range of applications in deep sea scenarios, including: (1) geological mapping; (Escartín et al. 2008; Yoerger et al. 2000) created mosaics for hydrothermal vents and spreading ridges, assessments of ferromanganese-nodule distribution (Peukert et al. 2018) (2) biological surveys; (Corrigan et al. 2018; Lirman et al. 2007; Ludvigsen et al. 2007; Simon-Lledó et al. 2019; Singh et al. 2004a) used them to map benthic ecosystems and species. (3) in archaeology; (Ballard et al. 2002; Bingham et al. 2010; Foley et al. 2009; Johnson-Roberson et al. 2017) documented ancient shipwrecks via mosiacs. (Gracias et al. 2000; Gracias et al. 2003) applied charted mosaics for later (4) navigation purposes. (5) underwater structure inspection; (Shukla et al. 2016) produced mosaics to inspect underwater industry infrastructure.

Early works mainly demonstrate 2D subsea mosaicing in relatively small areas, achieved through image stitching (Eustice et al. 2002; Marks et al. 1995; Pizarro et al. 2003; Vincent et al. 2003). At that time, lighting issues have already been considered, compensation of the light patterns have also been demonstrated later in some large area mapping tasks (Prados et al. 2012; Singh et al. 2004c). More recently, 3D photogrammetric reconstruction techniques using structure from motion (SfM) (Hartley et al. 2004; Maybank et al. 1992) or simultaneous localization and mapping (SLAM) (Durrant-Whyte et al. 2006) have enabled advanced 3D reconstruction for deep sea mapping (Drap et al. 2015; Johnson-Roberson et al. 2017; Johnson-Roberson et al. 2010; Jordt et al. 2016).

## 1.3 Deep Sea Imaging System

When the first deep sea imaging systems were designed, they already comprised the basic components: camera, pressure housing, and artificial illumination.

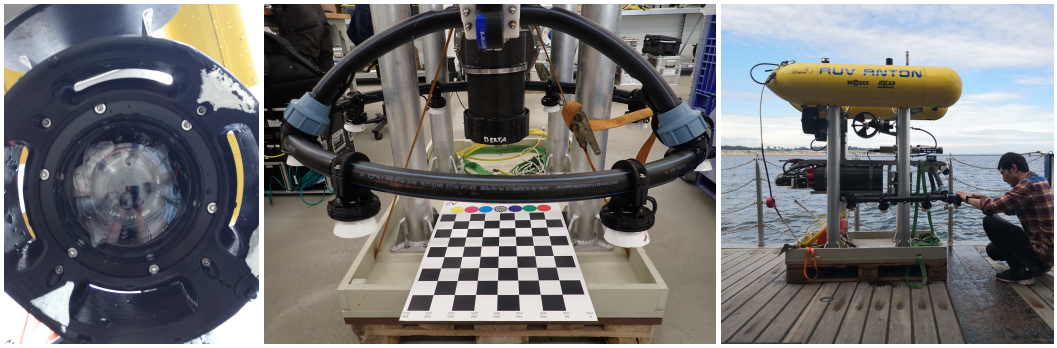



Figure 1.3: The Deep Sea Imaging System consists of three key components: a high-resolution underwater camera for capturing detailed images, a robust housing designed to protect the camera from extreme ocean conditions, and adequate artificial lighting, typically powered by LEDs, to illuminate the scene in the deep sea's perpetual darkness. Form left to right: an advanced camera enclosed within a dome port housing, the imaging system with a ring-shaped lighting comprising eight high-performance LEDs, integration of the entire system onto an AUV platform.

The evolution of cameras has seen a transition from analog to smart high-definition digital cameras, with 4k resolution becoming the standard in the market. High-resolution underwater cameras play a crucial role in capturing quality images, enabling precise 3D reconstruction down to millimeter accuracy. The housing serves as a critical component, safeguarding the camera from the extreme deep ocean conditions. It is constructed using durable materials like titanium or aluminum, to endure the immense pressure and corrosive seawater. Its front is equipped with an optically transparent window, often made of sapphire or tempered glass, ensuring that light can pass through and reach the camera's lens without compromising its integrity. In the deep sea, where sunlight cannot penetrate, additional artificial light sources are essential to illuminate the scene. Co-moving light sources project illumination patterns onto the seafloor, generating unique effects that are different from homogeneous sunlight. Adequate lighting is crucial for obtaining high-quality images, and modern deep sea systems are increasingly employing lightweight, energy-efficient light emitting diode (LED) lighting over traditional Xenon strobes. To achieve better energy efficiency, a timing circuit is often incorporated to seamlessly triggers the flash in coordination with the camera.

Most deep sea imaging systems did not prioritize 3D reconstruction, leading to cameras suffering from strong refractive distortions at the housings. However, for accurate 3D reconstruction, it is essential for the camera not only to "see" the subsea but also to capture images according to a well-understood photogrammetry model. Geometric properties and hardware design considerations for accurate 3D reconstruction in deep sea imaging systems are discussed in Part I, while lighting issues related to radiometric aspects are explored in Part II.

## 1.4 Main Contributions

The contributions of this thesis have been previously published in the following works, for which I am either the main author or have significantly contributed:

1. Song, Y., Köser, K., Kwasnitschka, T., and Koch, R. (2019). "Iterative refinement for underwater 3d reconstruction: Application to disposed underwater munitions in the baltic sea". *ISPRS-international archives of the photogrammetry, remote sensing and spatial information sciences*, 42, pp. 181-187.

2. She, M., Y. Song, J. Mohrmann, and K. Köser (2019). "Adjustment and Calibration of Dome Port Camera Systems for Underwater Vision". In *Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019*. Springer International Publishing, 42, pp. 79-92.

3. Song, Y., D. Nakath, M. She, F. Elibol, and K. Köser (2021). " Deep Sea Robotic Imaging Simulator". In *Pattern Recognition. ICPR International Workshops and Challenges*. Cham: Springer International Publishing, pp. 375-389.

4. Song, Y., J. Sticklus, D. Nakath, E. Wenzlaff, R. Koch, and K. Köser (2021). "Optimization of Multi-LED Setups for Underwater Robotic Vision Systems". In *Pattern Recognition. ICPR International Workshops and Challenges*. Cham: Springer International Publishing, pp. 390-397.

5. Köser, K., Y. Song, L. Petersen, E. Wenzlaff, and F. Woelk (2021). "Robustly Removing Deep Sea Lighting Effects for Visual Mapping of Abyssal Plains". *arXiv preprint arXiv:211-0.00480*.

6. She, M., D. Nakath, Y. Song, and K. Köser (2022). "Refractive geometry for underwater domes". *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, pp. 525–540.

7. Song, Y., Nakath, D., She, M., and Köser, K. (2022). "Optical imaging and image restoration techniques for deep ocean mapping: a comprehensive survey". *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 90(3), pp. 243-267.

8. Song, Y., She, M., and Köser, K. (2022). "Virtually throwing benchmarks into the ocean for deep sea photogrammetry and image processing evaluation". *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, pp. 353-360.

9. Song, Y., She, M., and Köser, K. (2024). "Advanced Underwater Image Restoration in Complex Illumination Conditions". *ISPRS Journal of Photogrammetry and Remote Sensing*, 209, pp. 197-212.

In detail, this thesis makes substantial contributions across three key areas:

Refractive Geometry **Refractive Geometry** The thesis conducts a comprehensive exploration of dome port refractive geometry, unveiling various geometric insights. My collaboration in (She et al. 2022a) involved significant contributions to the analysis and validation of properties related to dome geometry. Recognizing the potential of dome ports to physically counteract refraction effects at the hardware level, a joint effort with other authors in (She et al. 2019) resulted in the innovation of a mechanical adjustment method for aligning the camera behind the

dome port. Simultaneously, a calibration approach was proposed to estimate any remaining decentering offset. In cases where imaging systems employ flat port housing — where refraction is inevitable — an iterative refinement approach was presented in (Song et al. 2019) for underwater 3D reconstruction from images distorted by flat port refraction.

**Underwater Image Formation** The underwater image formation can be categorized into four types based on illumination conditions (Song et al. 2024). An adapted Jaffe-McGlamery model was introduced to describe deep sea image formation under multi-directional artificial illumination. This model, detailed in (Song et al. 2021a), not only simulated deep-sea images for an underwater robotic simulator but was also employed to transform real-world scenes into deep-sea scenarios for benchmark synthesis (Song et al. 2022b). Additionally, the thesis demonstrates that the image formation model can estimate or optimize parameters in the model, such as illumination configurations for an AUV-based deep sea imaging system, thereby improving the quality of captured images (Song et al. 2021b).

**Underwater Image Restoration** The thesis provides a comprehensive review of underwater image restoration, focusing on deep-sea mapping, and categorizes approaches based on their employed image formation approximations (Song et al. 2022a). Based on a novel 3D lookup table structure, the thesis proposes a general image formation model capable of simultaneously mitigating water and artificial lighting effects for various types of underwater images (Song et al. 2024). Extensive experiments validated the effectiveness of this approach, accompanied by a detailed analysis of underwater image correspondences and constraints essential for estimating the parameters of the lookup table. Later, a simplified version of the aforementioned model, specifically designed for the restoration of blank and flat seafloor images, was introduced in (Köser et al. 2021). During this phase, I was contributing to part of the theory development and experiments.

Throughout my PhD journey, there have been several other published works to which I have also contributed as a co-author are not discussed in this thesis:

1. Zhan, K., Song, Y., Fritsch, D., Mammadov, G., and Wagner, J. (2020). "Computed tomography data colouring based on photogrammetric images". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp.361-368.

2. Nakath, D., M. She, Y. Song, and K. Köser (2021). “In-Situ Joint Light and Medium Estimation for Underwater Color Restoration”. In *CVF International Conference on Computer Vision Workshops*. pp. 3724-3733.

3. Nakath, D., She, M., Song, Y., and Köser, K. (2022). "An Optical Digital Twin for Underwater Photogrammetry: GEODT—A Geometrically Verified Optical Digital Twin for Development, Evaluation, Training, Testing and Tuning of Multi-Media Refractive Algorithms". *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 90(1), pp. 69-81.

4. She, M., Weiß, T., Song, Y., Urban, P., Greinert, J., and Köser, K. (2022). "Marine bubble flow quantification using wide-baseline stereo photogrammetry". *ISPRS Journal of Photogrammetry and Remote Sensing*, 190, pp. 322-341.

5. She, M., Song, Y., Nakath, D., and Köser, K. (2023). "Efficient Large-scale AUV-based Visual Seafloor Mapping". *arXiv preprint arXiv:2308.06147* (submitted to: Journal of Field Robotics).

# Part I

# Geometry

# 2 Refractive Geometry

As water pressure increases by about 1 atmosphere for every 10 meters of depth, underwater camera systems designed for deep-sea applications must contend with the challenging environment of saltwater and high pressures. To protect these systems, they are typically enclosed in a housing equipped with a thick transparent window, often made of materials such as glass or sapphire. When light rays from objects submerged in water interact with this protective window, they pass through it and enter the air-filled interior of the housing. However, due to the differing optical densities of water, glass, and air, these rays change direction as they cross the interfaces when the angles are non-orthogonal. To achieve accurate visual reconstruction of the seafloor, it is crucial to have a comprehensive understanding of the geometry of the underwater imaging system. This chapter delves into the theories and solutions related to refractive effects, providing insights into how to eliminate these effects for precise underwater 3D reconstruction.

## 2.1 Basic Concepts

### 2.1.1 Projective Geometry and Camera Models

Camera projections are fundamental to the process of creating a two-dimensional (2D) representation of a three-dimensional (3D) world. They are instrumental in helping us understand the 3D structure of objects depicted in images. Visual 3D reconstruction, in essence, is the method used to infer the 3D structure of objects from the information captured in 2D images.

To facilitate the linear handling of transformations involving points and image plane, the Euclidean space $\mathbb{R}^n$ is expanded by one dimension accommodate points at infinity, which results in the projective space $\mathbb{P}^n$. In this extended space, the Euclidean space is embedded as the plane where $\omega = 1$. This concept is widely employed in computer vision. Consequently, a 3D point $\mathbf{X}(X,Y,Z)^\top$ in $\mathbb{R}^3$ can be represented using homogeneous coordinates as $\widetilde{\mathbf{X}}(X,Y,Z,1)^\top$. Similarly, the corresponding 2D point $\mathbf{x}(u,v)^\top$ can be expressed in homogeneous coordinates as $\widetilde{\mathbf{x}}(u,v,1)^\top$.

The projective geometry (Hartley et al. 2004) of mapping a 3D point to a 2D camera image is often described using a pinhole camera model. In this model, the view of a 3D scene is captured by projecting each 3D point onto the 2D image plane through a perspective transformation, resulting in its corresponding pixel on the image. This transformation essentially reduces the 3D world to a 2D image, a process where one dimension is lost. The distortion-free projective transformation, mapping a 3D point represented in world coordinates as $\mathbf{X}$ to its corresponding 2D pixel in image coordinates $\mathbf{x}$, follows the pinhole camera model and can be defined as:

$$s \cdot \widetilde{\mathbf{x}} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\widetilde{\mathbf{X}}. \tag{2.1}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and translation vector that transform from the world coordinates to the local camera coordinates, which are also named extrinsic parameters. The parameter $s$ denotes an arbitrary scaling factor which is not an inherent part of the camera model. The camera matrix, denoted as $\mathbf{K}$, is responsible for converting 3D points from the local camera coordinate system into 2D pixel coordinates. It comprises intrinsic parameters including the focal lengths $f_x$ and $f_y$ (in pixel units), as well as the principal point $(c_x, c_y)$ which marks the perspective center projected onto the image plane:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.2}$$

In practical optical systems, it is commonplace for lenses to exhibit specific forms of distortion. In the realm of camera modeling, a more comprehensive approach is often required to account for the idiosyncrasies associated with real-world lens systems. Two primary types of distortion are considered within this context: radial distortion and tangential distortion. Radial distortion emerges due to the imperfection of actual lenses, which deviate from the ideal pinhole model. It is as if we are mapping the image plane onto the curved surface of the lens itself. Radial distortion is typically categorized into two forms: pincushion distortion and barrel distortion, describing the deformation in the image, whether it is inward or outward. Tangential distortion, on the other hand, arises when the lens assembly is not perfectly centered over and aligned parallel to the image plane.

These distortion characteristics are typically determined through a common calibration procedure, often involving planar chessboard patterns (Zhang 2000). This thesis adopts the widely used distortion model from (Duane 1971), which accounts for three radial parameters $(k_1, k_2, k_3)$ and two tangential parameters $(p_1, p_2)$. The camera model is extended as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x x' + c_x \\ f_y y' + c_y \end{bmatrix}. \tag{2.3}$$

where

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy \end{bmatrix}. \tag{2.4}$$

with

$$r = \sqrt{x^2 + y^2}. \tag{2.5}$$

Here, $(x, y)$ represent the normalized image coordinates, computed from its local camera coordinates $\widetilde{\mathbf{X}_c} = (X_c, Y_c, Z_c, 1)^\top = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \widetilde{\mathbf{X}}$:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X_c / Z_c \\ Y_c / Z_c \end{bmatrix}. \tag{2.6}$$

For the sake of readability and without loss of generality, the subsequent contents will assume that pixel coordinates have already been corrected for distortion.

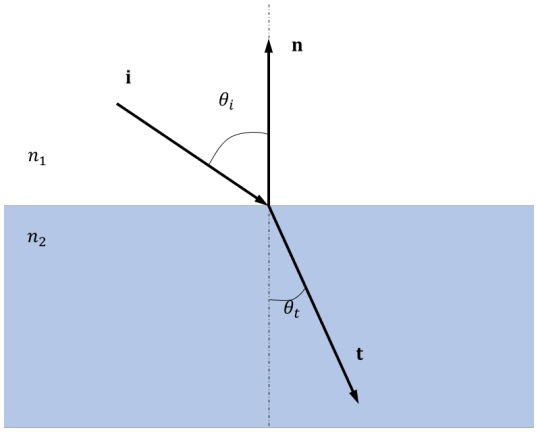



Figure 2.1: Get refracted ray follow Snell's law.

The process described above, where a projective camera maps a 3D point to a 2D image point, is known as forward projection. Conversely, when provided with an image point, it can determine the set of points in 3D space that map to this point, forming a ray (named back projection). To uniquely determine the position of the point in 3D, additional depth information ($Z_C$, also referred to as $depth$ in this thesis) is required:

$$\mathbf{X}_c = Z_c \frac{\mathbf{K}^{-1}\widetilde{\mathbf{x}}}{\|\mathbf{K}^{-1}\widetilde{\mathbf{x}}\|_2}. \tag{2.7}$$

Here the operator $\|\cdot\|_2$ indicates the Euclidean norm of the vector.

### 2.1.2 Refractive Geometry

In the context of deep-sea exploration, where water pressure increases by approximately 1 atmosphere for every 10 meters of depth, camera systems are commonly enclosed in protective housings featuring thick, transparent windows made of materials such as glass or sapphire. These housings shield the equipment from the corrosive effects of saltwater and the immense pressures experienced at depth. In underwater photography, light rays alter their directions as they traverse the interfaces between media with different optical densities, adhering to Snell's law:

$$\frac{\sin(\theta_i)}{\sin(\theta_t)} = \frac{n_1}{n_2}. \tag{2.8}$$

where $n_1$ and $n_2$ denote the refractive indices of the two media involved (typically 1.0 for air and 1.33 for water, although these values can slightly vary based on composition). $\theta_i$ and $\theta_t$ represent the angle of incidence and refraction, respectively. Snell's law is the foundation for calculating the refractive back projection from 2D to 3D in the local camera coordinate system, the refracted ray vector $\mathbf{t}$ can be derived as follows:

$$\mathbf{t} = \frac{n_1}{n_2}\mathbf{i} + (\frac{n_1}{n_2}\cos\theta_i - \sqrt{1-\sin^2\theta_t})\mathbf{n}. \tag{2.9}$$

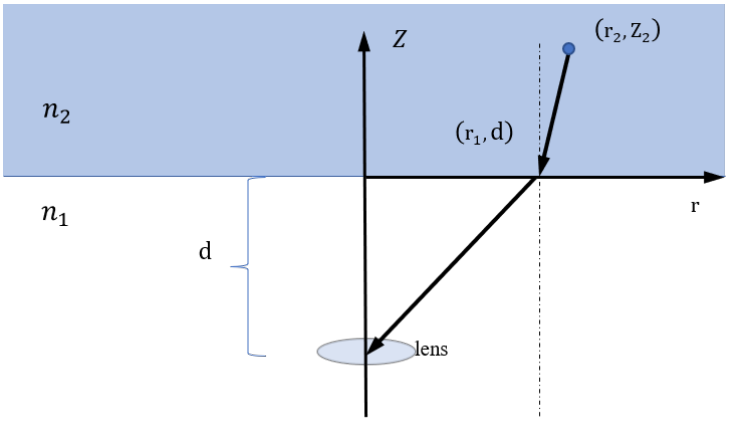



Figure 2.2: The ray from a 3D point in the medium intersects the refraction plane, modified from (Treibitz et al. 2011).

where **i** stands for the incident ray vector and **n** represents the normal vector of the refraction interface. According to Equation 2.8, $\sin\theta_t$ can be calculated from $\sin\theta_i$, leading to the following derived formula:

$$\sin^2\theta_t = (\frac{n_1}{n_2})^2 \cdot \sin^2\theta_i = (\frac{n_1}{n_2})^2(1-\cos^2\theta_i). \tag{2.10}$$

where $\cos\theta_i$ is the cosine of the supplementary angle between two known vectors **i** and **n**, can be formulated by the dot product of these vectors:

$$\cos\theta_i = -\mathbf{i}\cdot\mathbf{n}. \tag{2.11}$$

The previously mentioned formulas demonstrate that, for each ray in 3D space, the refracted ray can be directly calculated based on the normal vector of the refraction interface and the refractive indices of the two media.

Estimating the 2D projection of a 3D point on the refraction plane can be challenging because the intersection point on the interface is unknown, and the path of the ray cannot be directly defined. Nonetheless, this process still adheres to Fermat's principle, which dictates that light follows the path between two points that requires the least time to traverse.

As illustrated in Figure 2.2, the 3D point, embedded in the medium, passes through the flat interface and is refracted towards the center of the lens. Exploiting the symmetry around the camera's optical axis Z in this model, the 3D coordinate $\mathbf{X}_c$ can be expressed in radial coordinates as $((r_2, Z_2)$. Consequently, the travel time of the optical path $L$ can be formulated as:

$$L = n_2\sqrt{(r_2-r_1)^2+{Z_2}^2} + n_1\sqrt{{r_1}^2+d^2}. \tag{2.12}$$

where $d$ denotes the distance from the camera center to the refraction interface. The solution minimizes the traveling distance according to its partial derivatives (Glaeser et al. 2000):

$$\frac{\partial L}{\partial r_1} = n_2 \frac{r_1 - r_2}{\sqrt{(r_2 - r_1)^2 + {Z_2}^2}} + n_1 \frac{r_1}{\sqrt{{r_1}^2 + d^2}} = 0. \tag{2.13}$$

## 2.2 Models Considering Underwater Refraction

Over the past three decades, various concepts have been proposed to address underwater refraction effects within the camera model. The simplest approach involves employing the perspective camera model and adjusting intrinsic parameters from in-air calibration results to accommodate underwater conditions. A seminal study by (Fryer et al. 1986), comparing calibration results both above and under water, concludes that refraction can be compensated for by multiplying the in-air calibrated focal length by the index of refraction of water and the radial distortion coefficient is also adjusted accordingly. Similarly, (Lavest et al. 2000) reaches a similar conclusion about the focal length but employs a different model for computing radial distortion. Early underwater 3D reconstruction applications often directly adopt the perspective camera model for underwater camera calibration based on these findings (Bryant et al. 2000; Harvey et al. 1998; Kang et al. 2012; Shortis 2015).

However, it is crucial to note that the symmetric behavior of refraction effects using radial distortion holds true only when the camera viewing direction is perpendicular to the planar interface. As the refractive distortion is contingent on the distance between the scene and the camera, this approach introduces systematic errors (Jordt-Sedlazeck et al. 2012; Treibitz et al. 2011). The underwater calibration results obtained using the normal perspective camera model are only valid for points at the same distance as the calibration images.

A more general ray-based camera model, which accounts for refraction effects independently for each pixel, eliminates the systematic errors. The raxel camera model was initially introduced by (Grossberg et al. 2001; Grossberg et al. 2005), and later, (Narasimhan et al. 2005b) applied this model in underwater vision applications. Building upon the raxel model, (Sturm et al. 2004) proposed a generic concept for camera calibration, while (Ramalingam et al. 2006) further developed a generic SfM framework based on this camera model. Subsequently, (Chari et al. 2009) examined multi-view relations in this context.

However, describing refraction effects using independent 3D rays for each pixel introduces the risk of over-parameterization, resulting in a high degree of freedom. In (Łuczyński et al. 2017b), they introduced a modified model that utilizes a stored lookup table to compensate for distortion on a per-pixel basis at a fixed distance. To address the over-parameterization issue, (Wolff 2007) presented a simplified solution by using the axial camera model to characterize the refractive camera. In this approach, the camera is viewed as a non-single viewpoint (nSVP) camera with a caustic. In this case, light rays intersect at a common axis instead of single point and a refractive camera can be approximated by several perspective cameras corresponding to different areas of the image. Drawing from the axial model, as outlined in (Telem et al. 2010), the study involved mapping each 3D point through perspective projection and calculating the correct intersection with the axis.

Research on underwater image refraction can be broadly classified into two categories: flat port and dome port refraction. Notably, flat port refraction has garnered more extensive attention compared to dome port cases. In the realm of flat port refraction, (Maas 1995) introduced a linearized correction model for multi-media photogrammetry at flat interfaces. (Glaeser et al. 2000)

conducted an in-depth analysis of refraction effects, providing efficient formulas and parametric equations for flat refraction. This study also contributed to the development of a general theory of refractions. Recognizing the unique characteristics of cameras behind flat ports, (Treibitz et al. 2011) identified them as nSVP cameras. Subsequently, (Agrawal et al. 2012) demonstrated axial camera properties for flat refractive cameras. This work revealed that the projection of 3D points into the camera could be computed using analytical forward projection, solving a 12th-degree polynomial. Expanding on these findings, (Jordt-Sedlazeck et al. 2012) employed an analysis-by-synthesis approach for flat port underwater camera calibration. Additionally, inspired by the work of (Ramalingam et al. 2006), (Jordt-Sedlazeck et al. 2013) proposed a method utilizing a virtual camera, facilitating efficient bundle adjustment in underwater scenarios.

In cases involving dome port housings, when the optical center of the camera aligns with the center of the dome port, the refractive effect becomes negligible (Kotowski 1988; Menna et al. 2016; She et al. 2019). However, in scenarios with decentered dome ports, refraction persists. Attempts to directly apply the standard camera radial distortion model for calibrating dome-refracted cameras result in systematic errors. (Menna et al. 2016; Nocerino et al. 2021) address this issue by introducing additional filtering or correction mechanisms for the perspective model calibration results, to mitigate systematic errors. Highlighting the non-linear geometry of dome refractive setups, (Kunz et al. 2008b) provides insights into offset computation through a ray-tracing model, although no implementation is detailed. Building on this foundation, (She et al. 2022a) conducts a more in-depth analysis of dome refractive geometry and presents calibration methods for accurately estimating this offset. Additionally, leveraging a similar axial model as (Telem et al. 2010), (Elnashef et al. 2022) formulates a calibration approach specifically tailored for decentered dome configurations.

## 2.3 Refractive Geometry for Underwater Domes

The two primary interfaces commonly used in underwater imaging systems are flat ports and dome ports. While there are other constructions, such as pressure-proof deep ocean lenses used directly in the water or cylindrical windows, these have not seen extensive use in deep seafloor mapping. In the case of flat ports, only the ray that is perpendicular to the interface remains un-refracted. This refraction significantly reduces the camera's FOV underwater. However, dome ports offer a different scenario. If the optical center of the camera is precisely aligned with the center of the dome, incoming principal rays will not experience refraction (as shown in Figure 2.3).

A well-centered dome port in underwater imaging systems provides several advantages over flat ports. It eliminates the refraction effect, preserving the camera's FOV and focal length, which is particularly crucial for subsea mapping. Therefore, compared to flat port systems, a camera behind a dome port creates a larger footprint on the seafloor and requires fewer photos to cover the same area from the same flying altitude. Additionally, images captured by a wide FOV lens tend to perform better in pose estimation (Streckel et al. 2005; Zhang et al. 2016), a critical factor in the satellite-denied deep sea environment with challenging external localization. Furthermore, dome ports exhibit less chromatic aberration and can achieve sharper images (Menna et al. 2017).

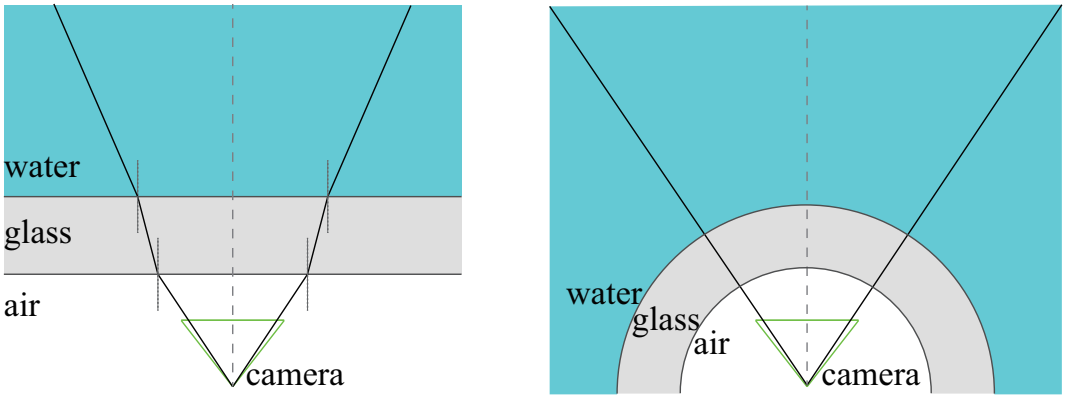



Figure 2.3: Left: Incoming light rays are refracted at the flat glass port. Right: when the entrance pupil of the camera is precisely positioned at the center of the dome, the principal rays are not refracted because they all pass through the air-glass-water interface orthogonally. The complete system can be considered as a normal pinhole camera.

Beyond their optical properties, devices designed for the deep sea must be mechanically stable to operate in high-pressure underwater environments. Flat port glass thickness requirements do not scale well with their diameter and the water depth, limiting the use of flat ports in the deep ocean to very small sizes or requiring extremely thick glass. Spherical ports (dome ports) are geometrically more stable since their spherical shape evenly distributes and withstands water pressure from different directions. Therefore, they require much thinner glass to withstand the same pressure compared to flat ports.

This thesis primarily concentrates on investigating refractive geometry related to underwater dome ports. This is because the refraction at a flat interface can be viewed as a special case of spherical interface refraction where the radius of the sphere is infinitely large. All the theorems derived can be readily applied to cases involving flat ports.

### 2.3.1 Decentered Dome Geometry

Positioning the camera center precisely at the center of the dome port presents challenges because both centers are imaged and invisible. For a decentered dome system where the centers are not perfectly aligned, the refraction effect remains.

The transparent window, such as glass, can be considered to have almost zero thickness (thin dome model), or, in particular for deep-sea housings, which endure several hundred bars of pressure, may have a window with several millimeters in thickness (thick dome model). These windows have an optical density denoted as $n_{glass}$. The exact refractive indices of seawater, glass, and air depend on their compositions and materials (This thesis generally assume $n_{air} < n_{water} \leqslant n_{glass}$). The refractive dome geometry of these two types of setups are illustrated in Figure 2.4. The vector that originates from the dome center $\mathbf{o}$ to the camera optical center $\mathbf{C}$ is known as the decentering offset vector, denoted as $\mathbf{v}_{off}$. The line that passes through the dome center and the camera center is referred to as the refraction axis, with its direction defined as $\mathbf{v}_{off} = \mathbf{C} - \mathbf{o}$. The refraction axis intersects the (thin) dome surface at two points, which we will call the refraction

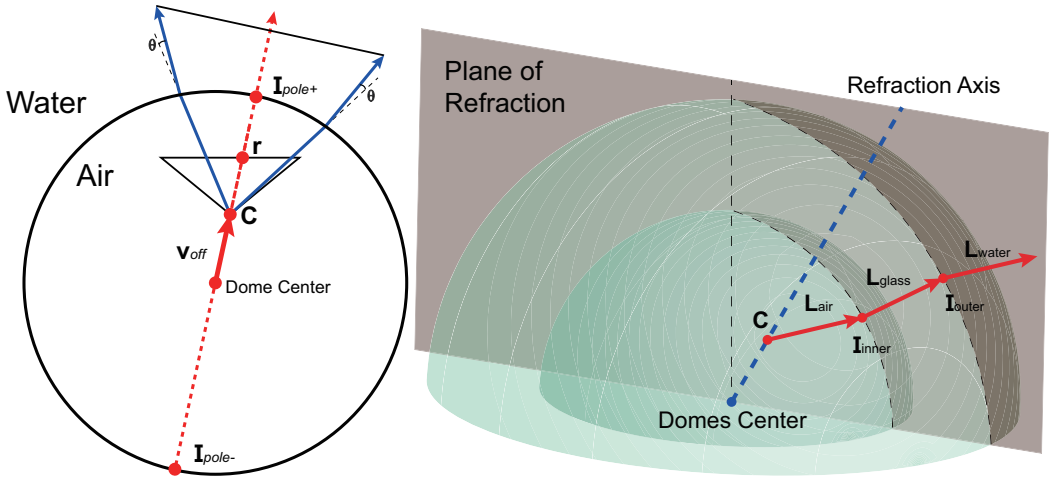



Figure 2.4: Sketches of both dome refraction setups (left: thin dome, right: thick dome). In both cases, the camera center is not located at the center of the dome (the origin). Only the ray passing through both the dome center the camera center will not undergo refraction at the sphere. This ray lies on the refractive axis, which is defined by the decentering offset vector $\mathbf{v}_{off}$. Others rays from the camera will not continue straight but rather be refracted. For each viewing ray, its refracted ray and the line from the origin to the dome center (refraction axis) all lie in one plane, called the plane of refraction.

poles. This thesis distinguishes between the pole closer to the camera center, labeled as the positive refraction pole ($\mathbf{I}_{pole+}$), and the pole further away as the negative refraction pole ($\mathbf{I}_{pole-}$).

In Figure 2.4 on the right, in the case of a thick dome setup, when considering the path of light from the camera optical center, through the glass dome, and to an object in the water, the entire path can be divided into three segments: the air segment with $\mathbf{L}_{air}$, the glass segment with direction $\mathbf{L}_{glass}$, and the water segment with a viewing ray direction of $\mathbf{L}_{water}$. The path of light intersects the air-glass interface at $\mathbf{I}_{inner}$ and the glass-water interface at $\mathbf{I}_{outer}$. When tracing a light ray from the camera optical center to its intersection at the inner interface, the following lemma can be derived:

**Lemma 1.** *The surface normal $\mathbf{n}_{inner}$ of the inner interface at $\mathbf{I}_{inner}$ is a linear combination of $\mathbf{v}_{off}$ and $\mathbf{L}_{air}$.*

*Proof.* The surface normal $\mathbf{n}_{inner}$ is defined along the vector originating from the dome center $\mathbf{o}$ to the intersection point $\mathbf{I}_{inner}$. Given that $\mathbf{o}$, $\mathbf{C}$, and $\mathbf{I}_{inner}$ form a triangle, which can be expressed as: $\overrightarrow{\mathbf{oC}} + \overrightarrow{\mathbf{CI}}_{inner} + \overrightarrow{\mathbf{I}_{inner}\mathbf{o}} = 0$. With $\overrightarrow{\mathbf{oC}} = \mathbf{v}_{off}$, $\overrightarrow{\mathbf{CI}}_{inner}$ being parallel to $\mathbf{L}_{air}$, and $\overrightarrow{\mathbf{I}_{inner}\mathbf{o}}$ parallel to $\mathbf{n}_{inner}$, we can derive: $\mathbf{n}_{inner} = c_1\mathbf{v}_{off} + c_2\mathbf{L}_{air}$. □

**Lemma 2.** *For the entire light path from an object in the water to the camera, all the light ray segments $\mathbf{L}_{air,glass,water}$, their normals at the inner and outer dome intersection points, and the refraction axis, all exist within a single plane.*

*Proof.* As per Lemma 1, it has been established that $\mathbf{n}_{inner} = c_1\mathbf{v}_{off} + c_2\mathbf{L}_{air}$, implying they lie in the plane defined by $\mathbf{v}_{off}$ and $\mathbf{L}_{air}$. In line with Snell's law, the refracted light ray can be expressed

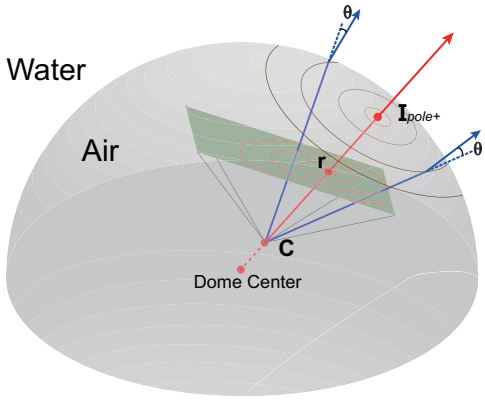



Figure 2.5: The iso curves are circles on the dome's inner surface, each formed by incident rays with the same angle. These circles share a common feature: all their centers align with the refraction axis. When these iso curves are projected onto the image, they are conic sections.

as a linear combination of the incident light ray and the surface normal (refer to Equation 2.9). The $\mathbf{L}_{glass}$ segment can also be determined as a linear combination of $\mathbf{L}_{air}$ and $\mathbf{n}_{inner}$, indicating they are all situated within the same plane. Since $\mathbf{L}_{air}$ and $\mathbf{n}_{inner}$ are not parallel and they can uniquely define a single plane in 3D space, thus confirming that all of these vectors reside in the same plane. Similarly, the normal vector at the outer dome intersection point $\mathbf{n}_{outer}$ can also be verified to lie in this plane. □

In the flat refractive case (Agrawal et al. 2012), this plane is referred to as the plane of refraction. Subsequently, all viewing rays originating from the camera that intersect at the inner interface of the dome and possess identical incident angles are grouped into iso-curves, as shown in Figure 2.5. This results in:

**Theorem 1.** *Iso curves, formed by the incident rays with the same incident angle, are essentially circles on the inner surface of the dome, and their midpoints all lie on the refraction axis. The projection of these iso curves onto the image plane results in conic sections, all of which share a common focus.*

*Proof.* Due to the refraction axis crossing the dome center, it is evident that rays with the same angle relative to the axis will have same incident angles on the inner surface of the dome. These rays collectively form a 3D circle that exhibits symmetry around the refraction axis. As the camera center also lies on this axis, each circle combined with the camera center forms a cone within 3D space. When these cones intersect with the image plane, they create conic sections. Given that these cones originate from the same vertex at the camera center and have the same cone axis, it follows that the conic sections share a common focal point. This focal point corresponds to the intersection of the refraction axis with the image plane. □

The subsequent exploration focuses on the displacement arising from variations in the optical densities of different media. First assume that there is no water or dome port between the camera and the object, capturing the object as if it were "in-air". The pixel coordinates $\widetilde{\mathbf{x}}_a$ of an observed 3D point can be directly computed using Equation 2.1: $\widetilde{\mathbf{x}}_a \simeq \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\widetilde{\mathbf{X}}$. Upon introducing

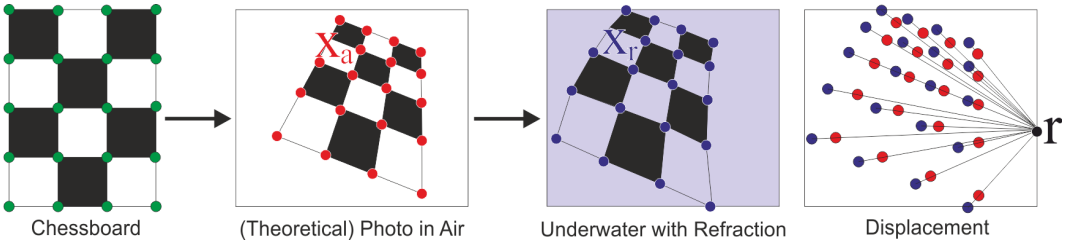



Figure 2.6: "In-air" observation $\mathbf{x}_a$ of any 3D point and its underwater refracted position $\mathbf{x}_r$ define a line passing through the refraction center $\mathbf{x}_r$ ((She et al. 2022a)). A chessboard (left image, corners marked in green) is photographed from an oblique viewpoint (second image). The red dots show the projected pixel positions without refraction $\mathbf{x}_a$. When the camera and chessboard are submerged in water behind a decentered dome, the light rays experience refraction (third image), and the projected pixel positions $\mathbf{x}_r$ are marked in blue. In the right image, displacement vectors generated by corresponding $\mathbf{x}_a$ and $\mathbf{x}_r$ points intersect at a single point, which corresponds to the refraction center $\mathbf{r}$.

water and a decentered dome port between the camera and the object, refraction occurs, leading the same 3D point to be imaged at a different pixel position $\mathbf{x}_r$. Specifically, only 3D points aligned with the refraction axis remain unrefracted, projecting onto the refraction center $\mathbf{r}$ in the image.

**Theorem 2.** *The "in-air" observation of a 3D point and its underwater "refracted" observation form a line through the refraction center* $\mathbf{r}$.

*Proof.* For any underwater 3D point $\mathbf{X}$ observed by the camera within a decentered dome, Lemma 2 asserts that all segments of the light path, including the 3D point and the camera pinhole, lie within a plane along with the refraction axis. The intersection of this plane with the image plane creates a line which crosses the refraction center $\mathbf{r}$. □

This implies that refraction essentially occurs along a line that connects the unrefracted "in-air" observation and the refraction center (as illustrated in Figure 2.6). When we consider the entire dome camera system submerged in water as a unique camera system, we can derive the following:

**Theorem 3.** *The decentered dome port camera system is an axial camera.*

*Proof.* Based on Lemma 2, it becomes evident that each water segment along a light path lies within a plane jointly with the refraction axis. For any 3D point which is not at infinity, its water segment within the light path will have a Euclidean intersection with the axis. As a result, all water segments of the paths leading to the pinhole, meaning the viewing rays in the water intersect the refraction axis. Hence, it can concluded that the overall system behaves like an axial camera. A similar conclusion regarding flat port systems can be found in (Treibitz et al. 2011). □

**Theorem 4.** *In the thin dome port camera system in water, all the rays from the camera are refracted towards the positive refraction pole.*

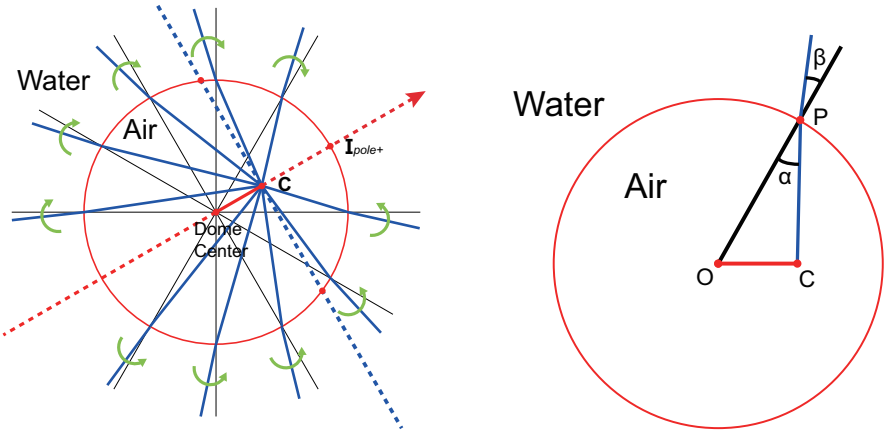



Figure 2.7: Left: Refraction of various blue rays from the camera center happens towards the positive refraction pole. Right: Maximum change of direction happens to the rays that enter the camera center from the plane perpendicular to the axis.

*Proof.* As established in Lemma 2, all segments of a light path share a common plane with the axis. the incoming ray changes its direction to the outgoing ray at the refraction interface. According to Snell's law, when transitioning from a less dense medium to a denser one, rays are refracted toward the surface normal. As all surface normals cross the dome center, also given that the camera is closer to the positive refraction pole, the incident rays from the camera are thus always refracted towards the positive refraction pole (see also Figure 2.7, Left). □

Hence, in the image, the 2D displacement direction with respect to the refraction center (either inward or outward) depends on the cosine of the angle between the viewing ray direction and the decentering offset vector. When this cosine is positive (referred to as forward decentering), refracted light rays change toward the refraction center, resulting in pincushion-like distortion in the image. Conversely, if the cosine is negative (known as backward decentering), refracted light rays diverge from the refraction center, leading to barrel-like distortion in the image. This phenomenon elucidates the underwater dome distortion effects commonly observed by underwater photographers and empirically reported e.g. in (Menna et al. 2016).

**Theorem 5.** *In the thin dome port model the maximum change of direction (refraction at sphere) happens to the rays that approach the camera center inside the plane perpendicular to the refraction axis.*

*Proof.* Given a unit circle $\odot O$, a point $C$ inside the circle has the distance $k \in [0,1]$ to the circle center $O$. A ray from $C$ intersects $\odot O$ at $P$ with the incidence angle $\angle\alpha$ and an outgoing angle $\angle\beta$ (see Figure 2.7, Right). The change of direction can be represented as $\angle diff = \angle\alpha - \angle\beta$. According to Snell's Law, $n_{air} \sin\alpha = n_{water} \sin\beta$, change of direction in range $[0, \pi/2)$ can be rewritten to:

$$\angle diff = \alpha - \arcsin\left(\frac{n_{air}}{n_{water}} \sin\alpha\right), \quad (\alpha \in [0, \pi/2) \tag{2.14}$$

Its first derivative is:

$$\frac{\partial \angle diff}{\partial \alpha} = 1 - \frac{1}{\sqrt{1-(\frac{n_{air}}{n_{water}}\sin\alpha)^2}} \cdot \frac{n_{air}}{n_{water}}\cos\alpha$$
$$= 1 - \sqrt{\frac{1}{\frac{n_{water}^2 - n_{air}^2}{n_{air}^2} \cdot \frac{1}{\cos^2\alpha} + 1}} > 0 \quad (2.15)$$

Since the derivative is strictly positive, $\angle diff$ is monotonically increasing ($\alpha \in [0, \pi/2)$) and does not have local maxima. Then the problem of finding a point $P$ on the circle which has the largest changes of direction $\angle diff$ is equivalent to find $P$ which has the largest incident angle $\angle\alpha$. Then, according to the Law of Sines,

$$\frac{\overline{OC}}{\sin\alpha} = \frac{\overline{OP}}{\sin\angle PCO}, \quad (2.16)$$

since $\overline{OC}$ and $\overline{OP}$ are fixed, and $\sin\alpha$ is monotonically increasing in the range $[0, \pi/2]$, $\angle\alpha$ must have its largest value when $\sin(\angle PCO)$ reaches its maximum 1. As $\angle PCO \in [0, \pi)$, it follows that when $\angle PCO = \pi/2$, $\sin(\angle POC) = 1$. Therefore, the maximum incident angle $\angle\alpha$ on the circle happens when $\overline{PC} \perp \overline{OC}$.

Since, for the sphere all refractions happen in a plane of refraction, which always include the axis, we can subdivide the sphere surface into circles that include the poles, and in each of them consider the problem only as a 2D problem inside the specific plane of refraction. As shown above, in each of them the maximum change of direction happens perpendicular to the axis.

□

This means that the ray which meets the camera center perpendicular to the refraction axis has suffered from the largest angular change (see Figure 2.7, Left), whereas the ray on the axis is not at all refracted. This is an important finding for setting up experiments to observe or to calibrate the decentering. Note that the largest effect in the image also depends on the orientation of the camera, since the angular resolution of a pinhole camera increases towards the boundaries: Lateral (left-right, or up-down, in the camera coordinate system) decentering will provide a much clearer signal-to-noise ratio (SNR) of refraction effects vs. corner detector uncertainty, as compared to forward-backward decenterings.

### 2.3.2 Refractive Dome Projection

According to Equation 2.9, the refracted ray can be computed from the incident ray and the normal at the refraction surface. For a homogeneous 2D pixel point $\tilde{\mathbf{x}}$ within the underwater image, its viewing ray, which starts from the camera optical center, is determined within the local camera coordinate system as follows:

$$\mathbf{L}_{air} = \frac{\mathbf{K}^{-1}\tilde{\mathbf{x}}}{\|\mathbf{K}^{-1}\tilde{\mathbf{x}}\|_2} \quad (2.17)$$

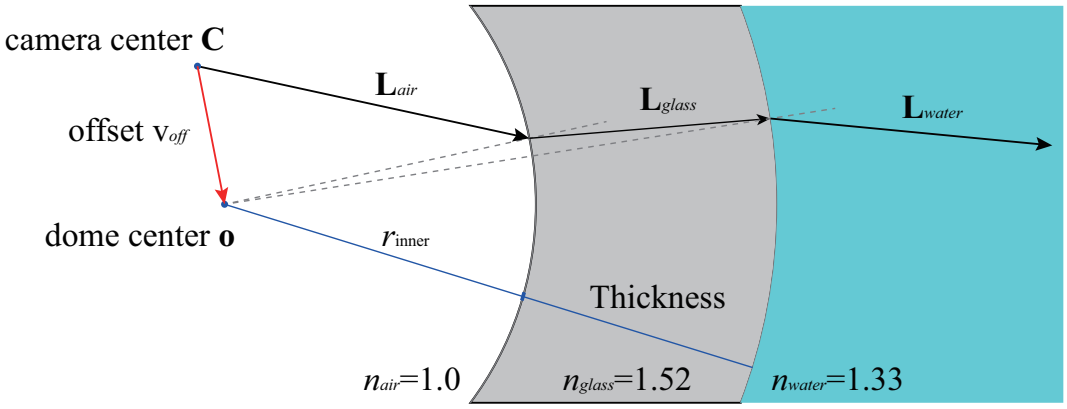



Figure 2.8: Refraction of viewing ray in misaligned dome port underwater.

If the camera center is set off from the dome center by $\mathbf{v}_{off} = (\delta X_{dome}, \delta Y_{dome}, \delta Z_{dome})$, the viewing ray from the image is refracted twice on the air-glass and glass-water interfaces, as shown in Figure 2.8. When light is refracted, the point of intersection on the interface can be determined by finding where the viewing ray intersects with the sphere. These spheres can be mathematically represented as implicit quadric surfaces (Hartley et al. 2004). Specifically, all 3D points $\widetilde{\mathbf{X}}_u$ located on a unit sphere satisfy the equation $\widetilde{\mathbf{X}}_u^\top \mathbf{Q} \widetilde{\mathbf{X}}_u = 0$, where the unit sphere is represented using a diagonal matrix notation, denoted as $\mathbf{Q} = diag\{1,1,1,-1\}$. The inner and outer spheres of the dome $\mathbf{D}$ then can be transformed from the unit dome sphere as follows:

$$\begin{aligned} \mathbf{D}(r, \mathbf{v}_{off}) &= (\mathbf{H}^{-1})^\top \mathbf{Q} \mathbf{H}^{-1}. \\ \text{where} \quad \mathbf{H}(r, \mathbf{v}_{off}) &= \begin{bmatrix} r & 0 & 0 & C_x + \delta X_{dome} \\ 0 & r & 0 & C_y + \delta Y_{dome} \\ 0 & 0 & r & C_z + \delta Z_{dome} \\ 0 & 0 & 0 & -1 \end{bmatrix}. \end{aligned} \tag{2.18}$$

Here $r$ indicates the radius of inner or outer sphere of the dome.

The ray-sphere intersection point $\widetilde{\mathbf{I}}_{inner}$ on the air-glass interface is satisfying:

$$\widetilde{\mathbf{I}}_{inner}^\top \mathbf{D}_{air} \widetilde{\mathbf{I}}_{inner} = 0 \quad \text{with} \quad \widetilde{\mathbf{I}}_{inner} = \mathbf{C} + \lambda_{air} \overrightarrow{\mathbf{L}}_{air} \tag{2.19}$$

which boils down to a single quadratic equation in $\lambda_{air}$. Once the intersection is determined, the normal vector can be derived and the refracted ray $\overrightarrow{\mathbf{L}}_{glass}$ can be calculated by Equation 2.9. For the glass-water interface, the intersection point $\widetilde{\mathbf{I}}_{water}$ can be calculated in the same way by:

$$\widetilde{\mathbf{I}}_{outer}^\top \mathbf{D}_{glass} \widetilde{\mathbf{I}}_{outer} = 0 \quad \text{with} \quad \widetilde{\mathbf{I}}_{outer} = \widetilde{\mathbf{I}}_{inner} + \lambda_{glass} \overrightarrow{\mathbf{L}}_{glass}. \tag{2.20}$$

enabling the computation of the outer interface point and the direction of the refracted ray within the water. It is used to calculate the refractive dome back projection, which transforms a 2D pixel into a 3D ray underwater.

When discussing refractive projection from 3D points into an image plane, it is important to note that there is no compact representation for the projection from a 3D point $\mathbf{X}_i$, into a decentered dome port camera. Therefore, this thesis employs an iterative approach to compute the projection, similar in spirit to (Grossberg et al. 2005; Kunz et al. 2008b). The iterative process for *refractive projection* essentially seeks to minimize the Euclidean distance between the back-projected 3D ray $\hat{\mathbf{L}}_i$ in water (obtained through the aforementioned back-projection approach) and the actual 3D coordinates $\mathbf{X}_i$:

$$\mathbf{x}_{i,proj} = \text{argmin} \, |\text{dist}(\mathbf{X}_i, \hat{\mathbf{L}}_i(\mathbf{x}_i, \mathbf{K}, \mathbf{R}, \mathbf{t}, \mathbf{v}_{off}, \mathbf{D}_{air}, \mathbf{D}_{glass}))|. \tag{2.21}$$

where $\text{dist}(\cdot)$ refers to the perpendicular distance between the 3D point and the back-projected ray. This distance can be computed using the following formula:

$$\text{dist}(\mathbf{X}_i, \hat{\mathbf{L}}_i) = \frac{\|[\mathbf{X}_i - \mathbf{I}_{outer}] \times \hat{\mathbf{L}}_i\|_2}{\|\hat{\mathbf{L}}_i\|_2} \tag{2.22}$$

Here, operator $\times$ indicates the cross product.

The optimization process can be described as follows: Initially, the 2D position $\mathbf{x}_i$ is computed by directly projecting the given 3D point into the image plane using the standard perspective model. Subsequently, a ray is traced backward from this initial 2D position to 3D space, taking into account multiple layers of refraction. The residual is determined as the Euclidean distance between the 3D point and the back-projected ray. This residual is then minimized using the Gauss-Newton algorithm to obtain the originally sought 2D projection $\mathbf{x}_{i,proj}$.

In underwater calibration cases, when employing a chessboard as the calibration target, the projection of chessboard corners can be simplified by working within the chessboard coordinate system, where all 3D points are situated on the plane with Z = 0. The intersection point $\hat{\mathbf{X}}_i$ of the back-projected ray with the Z = 0 plane, is essentially the back-projected chessboard corner in 3D coordinates. We can further simplify optimization process into one that minimizes the point-point distance:

$$\mathbf{x}_{i,proj} = \text{argmin} \, \|\mathbf{X}_i - \hat{\mathbf{X}}_i(\mathbf{x}_i, \mathbf{K}, \mathbf{R}_c, \mathbf{t}_c, \mathbf{v}_{off}, \mathbf{D}_{air}, \mathbf{D}_{glass})\|^2. \tag{2.23}$$

## 2.4 Adjustment and Calibration of Dome Port Camera Systems

As outlined at the beginning of this section, the dome port camera system offers the potential to eliminate refraction effects if the camera optical center aligns precisely with the dome center. This section introduces a novel mechanical adjustment procedure that relies on visual feedback from straight lines both above and below the water surface. This procedure enables highly accurate alignment of the camera center within the dome. It has proven to be effective for a wide range of underwater photogrammetry applications and has been implemented in several underwater camera system designs.

Applications demanding an exceptionally high level of precision, necessitate consideration of even the tiniest offsets between the camera optical center and the dome center. This thesis also

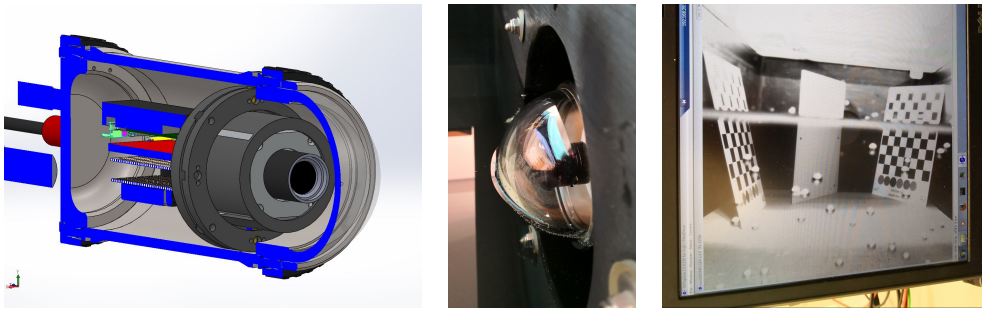


Figure 2.9: Left: Technical drawing of deep sea housing with mechanics for camera adjustments towards the dome port. Center: dome port (half underwater) attached to tank. Right: View through dome (after adjustment), with straight lines crossing the surface.

present a chessboard-based calibration method. This method utilizes pairs of images taken both underwater and above water to estimate any remaining offsets between the camera center and the dome center. This fine-tuning approach allows for further correction of refraction correction.

### 2.4.1 Mechanical Adjustment of Cameras

Using the pinhole camera model with underwater dome port cameras requires centering the lens with the dome port as good as possible. The key assumption here is that the lens does not exhibit a caustic in air (see (Grossberg et al. 2005)), i.e. it can be considered a pinhole. Within the three possible degrees of freedom (3D offset of the pinhole from dome center), the positioning in camera forward/backward direction poses the largest challenge, as both the lens and the dome port pressure housing are typically rotationally symmetric and their axes can be aligned (already by construction) with high precision. This section proposes a method for mechanically aligning the camera in the difficult forward/backward direction, using the concept of optical feedback control. In principle, this through-the-lens approach allows to adjust the lens, until refraction in this particular camera-lens-dome system becomes neglectable or cannot be observed. Measuring the error in pixels enables also to easily transfer the concept to other lenses that require different spatial alignment accuracy in millimeters without actually having to know that accuracy requirement.

For the mechanical alignment this thesis proposes to mount the camera at the flange of the dome port as depicted in Figure 2.9 (left). In our design the distance between the camera and the dome can be varied using a screw mechanism (moving 1mm backward/forward per screw rotation). Besides this option, many other constructions are possible where the lens is moved in forward/backward direction while staying centered in the other directions.

Then the dome port should be positioned at a water tank (see Figure 2.9, Center), such that the dome is half-way underwater and looks parallel to the water surface. In case the camera is centered perfectly, no refraction will occur and the underwater part and the above water part of the image will be consistent.

Straight lines that cross the water surface will simply continue. Figure 2.10 shows the corresponding images of a chessboard. The task of finding the correct camera position means to find the position at which straight lines remain straight across the water surface. By manually adjusting the mechanics and viewing the live images, one can determine the correct position easily.

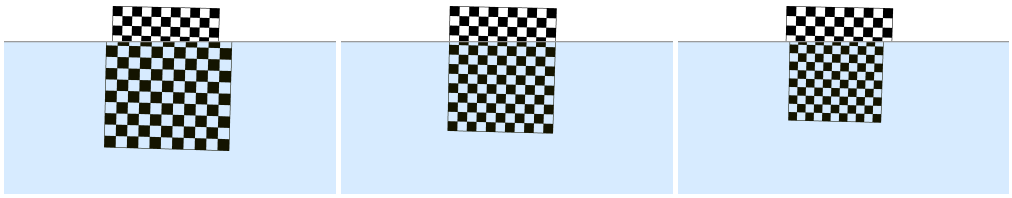


Figure 2.10: Simulated Chessboard images with parts above and below the water line. The lens is misaligned forward (left), aligned (center), misaligned backwards (right).

### 2.4.2 Calibration of Remaining Offsets

To compute the actual 3D offset of a camera from the dome center, an air-water image pair of a chessboard at the same position and orientation is acquired. In case the camera is perfectly centered, the corners in the two images will be exactly at the same location. Displacements indicate that there is refraction due to a centering offset (as the pose stayed the same). Assuming Gaussian noise on the detected chessboard corners, the estimation of the 3D offset $\mathbf{v}_{off}$ can be formulated in the Gauss-Markov model, essentially minimizing the energy:

$$E(\mathbf{v}_{off}) = \sum_{i \in \Omega} \|\mathbf{x}_{i,r} - \mathbf{x}_{i,proj}(\mathbf{x}_{i,a}, \mathbf{v}_{off}, \mathbf{K}, \mathbf{R}_c, \mathbf{t}_c, \mathbf{D}_{air}, \mathbf{D}_{glass})\|^2 \tag{2.24}$$

where $\mathbf{x}_{i,r}$ and $\mathbf{x}_{i,a}$ represent the detected coordinates of chessboard corners from underwater and in-air images, respectively. $\mathbf{R}_c$ and $\mathbf{t}_c$ contain the pose information of the camera relative to the chessboard. This information can be derived from a set of 3D-2D correspondences in the in-air condition using standard pose estimation techniques. $\mathbf{x}_{i,proj}$ is the projection of the chessboard point through the dome port in the underwater case, computed using Equation 2.23.

To initiate the optimization process, an initial hypothesis of zero offset can be assumed, where the projection results in the in-air corner positions. The offset parameters are then determined through Gauss-Newton optimization, utilizing the Gauss-Markov model, as described in (McGlone 2004). Derivatives necessary for the optimization can be computed through finite differences approximation.

### 2.4.3 Evaluation of the Calibration

The proposed dome calibration method was first tested using real data to validate the accuracy of the camera-to-dome center offset estimations. This experiment involved a low-cost webcam housed in an acrylic dome with a radius of 77.45mm and a thickness of 2.4mm.

The in-air calibration of the webcam was performed using a standard toolbox (Bouguet 2004). During the experiments, the webcam was placed at different positions in the dome port (see Figure 2.11), which are described in terms of the initial position close to ① the center of the dome port: ② offset 15mm along Z-axis, ③ offset 15mm along Y-axis, ④ offset -15mm along Y-axis, ⑤ offset 20mm along X-axis. The calibration results are shown in Table 2.1.

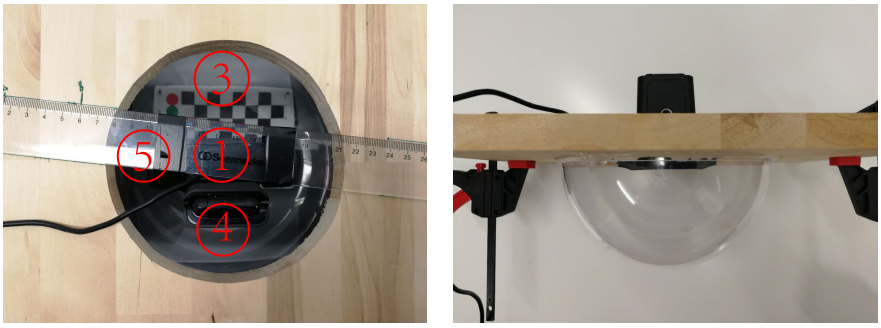



Figure 2.11: Experimental settings of the webcam with an acrylic dome. Left: Top view with test positions, Right: Side view.

.

Table 2.1: Calibration results of the low cost webcam with acrylic dome ports

| | offsets | | | residual[pixel] | |
|---|---|---|---|---|---|
| | $\delta_x$[mm] | $\delta_y$[mm] | $\delta_z$[mm] | before calibration | after calibration |
| Test ① | 0.8071 | 3.3953 | 1.8441 | 5.6855 | 0.8051 |
| Test ② | 1.5459 | 4.5547 | 15.7268 | 7.8347 | 0.6356 |
| Test ③ | -0.6054 | 17.6406 | -0.7525 | 30.1930 | 1.6123 |
| Test ④ | 1.1253 | -11.7163 | 5.2843 | 19.2762 | 0.8741 |
| Test ⑤ | -14.3336 | 0.3173 | 0.1010 | 23.8291 | 0.5925 |

The experiments demonstrated that the proposed calibration algorithm yields reasonable offset values. Utilizing these computed offsets results in a significant reduction in the re-projection error residuals, and there is a strong agreement between the computed offsets and the physically measured offsets. Remaining differences can be explained by the inaccurate experiment setup and the imperfection of the cheap acrylic dome which is not an optical instrument.

The proposed technique has also been applied to a stereo camera system (two Basler cameras with $1280 \times 1024$ resolution, also see Section 5.1.2) with high quality glass domes (radius 50.10mm and thickness 7mm). The cameras were first mechanically adjusted according to the proposed approach of Section 2.4.1. The result of the mechanical adjustment procedure can be seen in Figure 2.9 (right), the calibration and evaluation results are shown in the Table 2.2.

The results show that the mechanical adjustment procedure can precisely align the camera center with the dome center, achieving sub-millimeter accuracy. To assess the accuracy of the esti-

Table 2.2: Calibration results of the stereo cameras with glass dome ports. Data courtesy of (She et al. 2019)

| | offsets | | | residual[pixel] | |
|---|---|---|---|---|---|
| | $\delta_x$[mm] | $\delta_y$[mm] | $\delta_z$[mm] | before calibration | after calibration |
| Master Camera | 0.0490 | 0.5033 | -0.2967 | 2.9552 | 0.2847 |
| Slave Camera | -0.2431 | 0.0012 | 0.0324 | 6.5177 | 0.4120 |

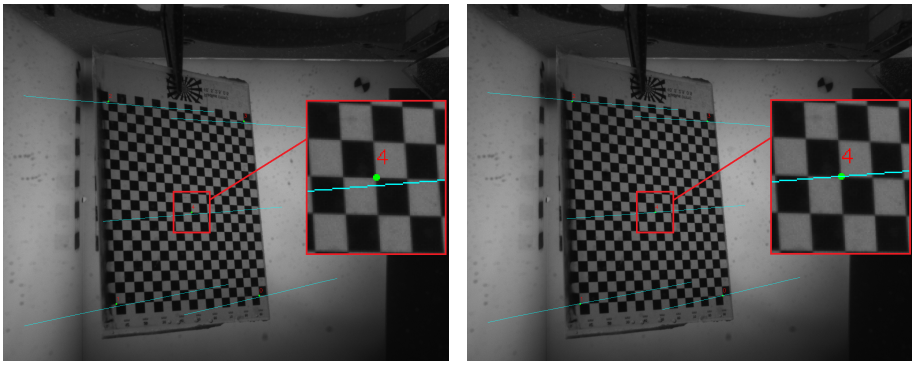


Figure 2.12: Epipolar lines of chessboard corners from right camera plotted in left camera image. Left: before offset calibration, Right: after offset calibration. Image courtesy of (She et al. 2019)

mated remaining offset, the stereo camera system is first calibrated in air and then submerged in water. The epipolar lines of some chessboard corners are plotted from the left image to the right image (refer to Figure 2.12, Left). When using the in-air calibration results to directly compute the epipolar lines from the right camera image to the left camera image, noticeable errors are observed, and the epipolar lines do not perfectly intersect the corners due to the refraction effect caused by remaining offset. When taking into account the estimated decentering offset of the dome port in epipolar line computation, the epipolar lines almost perfectly intersect the corresponding chessboard corners in the right image (see Figure 2.12, Right). This demonstrates that the offset calibration is both accurate and valuable.

## 2.5 Iterative Refinement for Underwater 3D Reconstruction

A well-centered dome port camera system effectively functions as a standard pinhole camera, allowing the utilization of the standard 3D reconstruction procedures for visual mapping. For imaging systems that cannot avoid refraction, such as underwater cameras with flat ports, this thesis introduces an iterative refinement approach to rectify refraction effects for underwater 3D reconstruction. This approach leverages approximate geometry to compensate for water refraction effects in images and then brings the new images into the next iteration of 3D reconstruction until the changes of resulting depth maps become negligible. Assuming that the refractive geometry parameters have been obtained through calibration, the complete reconstruction pipeline process for the proposed methods can be described as follows:

1. First, obtain the standard calibration of the camera in air.
2. Import the original underwater images into a standard photogrammetry processing pipeline (in this thesis, the processing is performed using commercial software Agisoft Metashape) to reconstruct the initial 3D model with fixed camera intrinsics from Step 1.
3. Utilize the depth maps rendered from the initial 3D model to correct for the refraction effects in the underwater images.

4. Remap the depth maps to the original image frame.

5. Iteratively reconstruct the 3D model from updated refraction corrected images and generate the new depth maps to compute the new refraction corrected images. Then, remap current depth maps to the original image frame for comparison until the depth map values in successive iterations have converged.

6. Once the final geometry-refined images are obtained, the color information can be adjusted based on the depth maps, and the reconstruction result is updated to obtain the final 3D reconstruction product (More details about underwater color restoration will be discussed in the Chapter 4).

At first glance it might appear that using the original images to reconstruct the scene in the first iteration would be inconsistent. Indeed, if there is prior geometry information at the beginning (known ground plane, maximum viewing distance, detected markers, etc.), the images before the first iteration could be undistorted with respect to this geometry. Not undistorting them means assuming that the entire scene is close to the glass interface of the camera (refraction can be ignored). Which of the priors should be used such that the algorithm will converge to the correct 3D scene layout depends on the setting and needs further evaluation. For the initial experiments reported in this contribution, it applies the "close scene" assumption at the beginning.

For practical reasons (holes, noise, artifacts), all the depth maps mentioned in this paper are rendered from the photogrammetric reconstruction result rather than the raw depth maps from dense matching as they are more complete and consistent. Another assumption is if there is no depth information in some image area, then the object is assumed to be on a plane which exceeds visibility distance (here it is set to 15m). To fill small holes, this thesis uses a superpixel segmentation of the image and fill missing depth values by interpolating between neighboring pixels of the same segment. The workflow of the whole processing is illustrated in Figure 2.13.

### 2.5.1 Geometric Refinement Processing

The geometric refinement process comprises two primary components: refraction correction and image frame conversion. Refraction correction is responsible for rectifying the original underwater images to mitigate the refraction effects, while image frame conversion involves remapping the depth map from the current image frame back to the original image frame, facilitating a comparison of depth value changes for loop decision-making.

#### Refraction Correction

Refraction correction involves accessing the known depth value for each pixel, back-projecting the ray along the refracted path to obtain the 3D point, and then projecting this 3D point onto the 2D image plane using the standard pinhole model to obtain its unrefracted pixel coordinates. The target coordinates for each pixel are non-integer values, necessitating a scattered interpolation process to interpolate the pixel values for each pixel in the target image. First, a Delaunay triangulation is created using the scattered data points as vertices. Subsequently, natural neighbor interpolation (Sibson 1981) is applied to interpolate the values at each pixel center. The pseudo-code in Algorithm 1 and Figure 2.14 outline the complete procedure for image refraction correction:

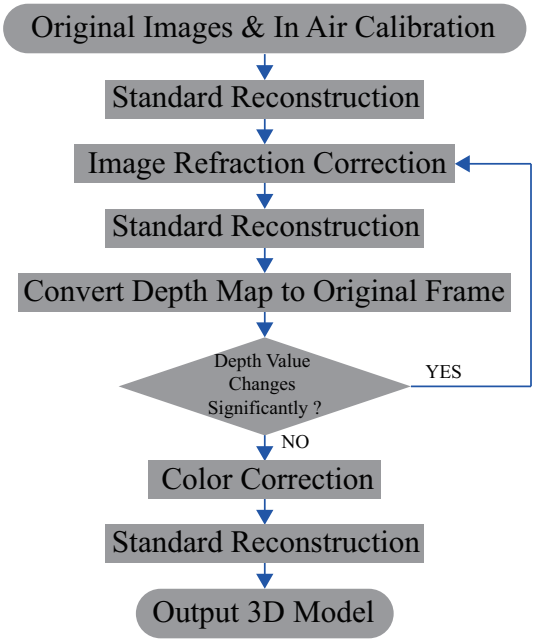



Figure 2.13: Workflow of the proposed iterative refinement 3D reconstruction.

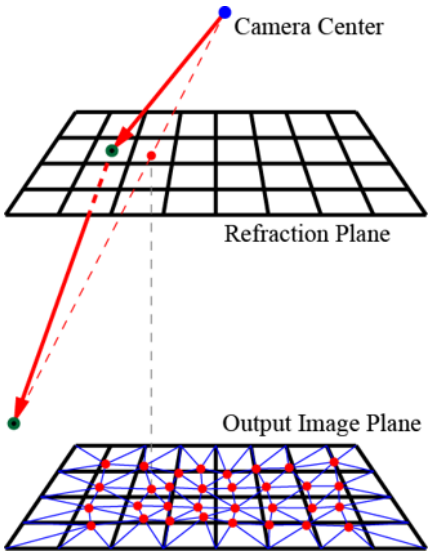



Figure 2.14: Remove the refraction effect from images. Each pixel is back-projected along the refracted path to obtain a 3D point. This 3D point is then projected onto the 2D image plane using a straight line under the pinhole model. The resulting pixel coordinates, along with their intensity values, are considered as scattered data, which are employed to interpolate the pixel intensities for each pixel in the new refraction-free image.

**Algorithm 1** Refraction correction algorithm in a underwater image

**Require:** underwater image with refraction, depth map $D$, camera intrinsics, distance to refraction interface $d_0$, refraction indices

1: **for** each pixel **do**
2: Generate the original ray from the camera center according to Equation 2.17.
3: Calculate the refracted ray according to Equation 2.9.
4: Compute the 3D point along refracted ray with the depth of $(D(\mathbf{x}) - d_0)$.
5: Project the 3D point to image plane using the standard pinhole model, save as the target coordinates.
6: Create the Delaunay triangulation from the target coordinates with their corresponding depth values and color information.
7: Interpolate each pixel value (color and depth) for the output images.
**return** refraction corrected color image with corresponding depth map in the same image frame.

#### Image Frame Remapping

The pixel coordinates in the image typically changed after the refraction correction and the current exported depth map cannot be directly compared with the one from the previous iteration. Also, the correction processing must use the original underwater images, so that the conversion of the new depth map from the refraction corrected image frame to the original image frame is needed. There are two solutions to solve this problem. One solution is applying inverse transformation of the refraction correction. During the refraction correction procedure, the target coordinates for each pixel in the original image have been calculated, which can be stored in a transformation matrix. The transformation matrix records the target pixel coordinates for each pixel, which also can be used for inverse transformation from corrected image frame to original image frame. The image transforming function by using transformation matrix has been implemented in OpenCV cv::remap function (Bradski 2000).

Another solution is to project the 3D points under the pinhole camera model and to minimize the light travelling time (see Equation 2.13), to estimate the intersection point on the refraction plane and to derive the supposed pixel coordinate in the refraction scene. Afterwards, apply a procedure similar to the one which is described in Algorithm 1 to interpolate the pixel values for the output images.

### 2.5.2 Verification on Test Dataset

To validate the proposed geometry refinement approach, an underwater test dataset with ground truth (GT) information is required. Ideally, this dataset should include images with refraction effects and accurate position and geometry information, which can be challenging to obtain in practice. In this section, a pre-built 3D model was used as the GT data, and a set of images was generated from this model. Subsequently, one layer of air-water interface refraction effects were added to these images by applying refractive ray-tracing according to depth maps. The iterative

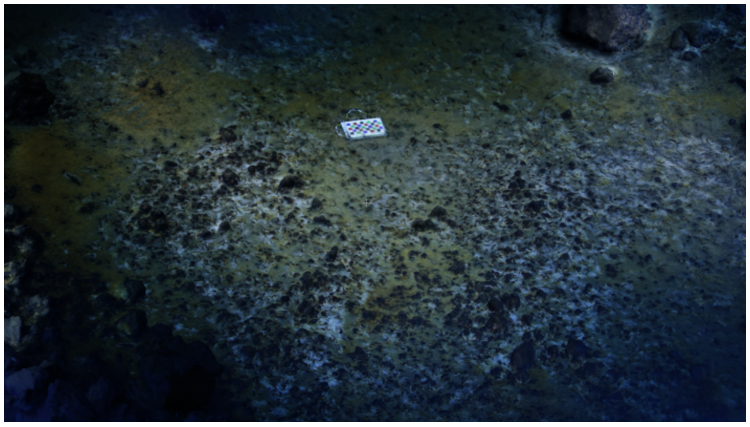


Figure 2.15: The reconstructed 3D model of a deep-sea floor area was used as the ground truth for synthesizing images with refraction, allowing for the evaluation of the proposed methods.

refinement approach was then evaluated on these test images to assess the improvements in 3D reconstruction.

### Simulation of Underwater Flat Port Fraction

In computer graphics, refraction effects have been extensively utilized to simulate underwater scenes (Hu et al. 2006; Sedlazeck et al. 2011; Wyman 2005). To validate the proposed approach, a 3D model was created using underwater footage obtained during a research cruise to the Niua South hydrothermal vent field (Kwasnitschka et al. 2016) (as depicted in Figure 2.15). Because the camera used to capture these images was well centered inside a dome port, these images were directly employed to generate the 3D model within the standard reconstruction pipeline. Subsequently, images were synthesized with refraction according to Algorithm 2:

First, a standard graphics rendering engine was utilized to render a set of ground truth images without refraction, as well as their depth maps, and then these "as in air" images were converted into refractive underwater images according to the corresponding depth maps. The basic of the refraction simulation algorithm is ray-casting, which finds the first intersection for each ray casted from the observer (camera). Refractive ray-casting additionally computes the refracted ray from the original casted ray for further intersection calculation. To implement the ray-casting, the depth map was converted to a 3D triangle mesh net, and then the ray-triangle intersection was calculated by using the Möller-Trumbore intersection algorithm (Möller et al. 2005). The pseudo-code in Algorithm 2 and Figure 2.16 describe the algorithm of the refractive ray-casting approach to convert the in-air image to an underwater (refracted) scene.

### Accuracy Evaluation

The evaluation is conducted on the 31 simulated images with refraction effects obtained from the aforementioned 3D model. The in-air calibration result was pre-defined by rendering the images on the graphics engine. The simulated images with refraction were then imported into the iterative refinement reconstruction workflow. After two iterations, the generated depth map

**Algorithm 2** Add refraction effect on an in-air image

**Require:** in-air image with depth map, camera intrinsics, distance to refraction plane, refraction indices

Convert Depth Map to a regular triangle net.
Get the min&max values from the Depth Map.
**for** each pixel **do**
    Generate the original ray from the camera center.
    Calculate the refracted ray according to Equation 2.9.
    Back project the pixel with min&max depth along the refracted ray and get two 3D points.
    Project the points to the image plane to form a 2D line.
    Get all the triangle faces which touch the line in 2D.
    **for** each triangle face **do**
        Check if the refracted ray intersects the triangle face in 3D by using Möller-Trumbore intersection algorithm.
        **if** ray hits the face **then**
            Select the intersection point with minimum depth.
        **else**
            Back project the point along refracted ray with two times of maximum depth.
        Project the point to the image to get the refracted pixel position.
**return** image with refraction effect.

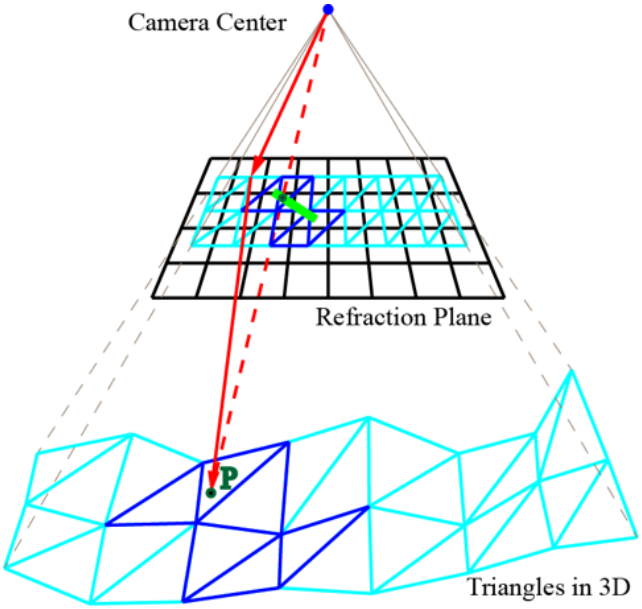



Figure 2.16: Simulating the refraction effect begins by creating a 3D mesh based on the in-air image. Subsequently, refracted rays are generated for each pixel in the output image, and their intersection points on the 3D mesh are computed. These intersection points are then projected onto the image plane using the pinhole model. Finally, color values are interpolated from the in-air image and assigned to the corresponding output image pixels.

values have already converged. Figure 2.17 illustrates the example intermediate results after the first iteration.

Figure 2.18 illustrates the differences between a refraction-free GT image and the refraction-corrected image in adjacent iterations. The mean absolute intensity errors between these two images are 2.5599 for the first iteration and 2.4271 for the second iteration, within the range of [0,255]. This indicates that the iterative refinement process effectively reduces the differences between the corrected images and the GT images taken in air without any refraction effects.

The GT model serves as a reference for evaluating the quality of the 3D models obtained in different iterations. The analysis focuses on the area covered by all models, and the statistics for the models from each iteration are examined. As shown in Figure 2.19 and Table 2.3, the absolute distances improve in subsequent iterations following the refraction refinement process. The refinement processing converged after two iterations.

In addition to assessing the quality of the 3D models in each iteration, an evaluation of the 3D reconstruction using a photogrammetry pipeline with auto-calibration mode was conducted. As depicted in Figure 2.20, the photogrammetry pipeline with auto-calibration mode yielded a satisfactory model with acceptable accuracy. This mode estimates the camera intrinsic parameters using a virtual camera, and the remaining refraction effects are compensated by the distortion parameters. However, this result could presumably be influenced by the fact that the selected evaluation area is located in the center of each image, where refraction effects are less severe compared to the pixels on the image boundary. The proposed approach consistently outperforms the auto-

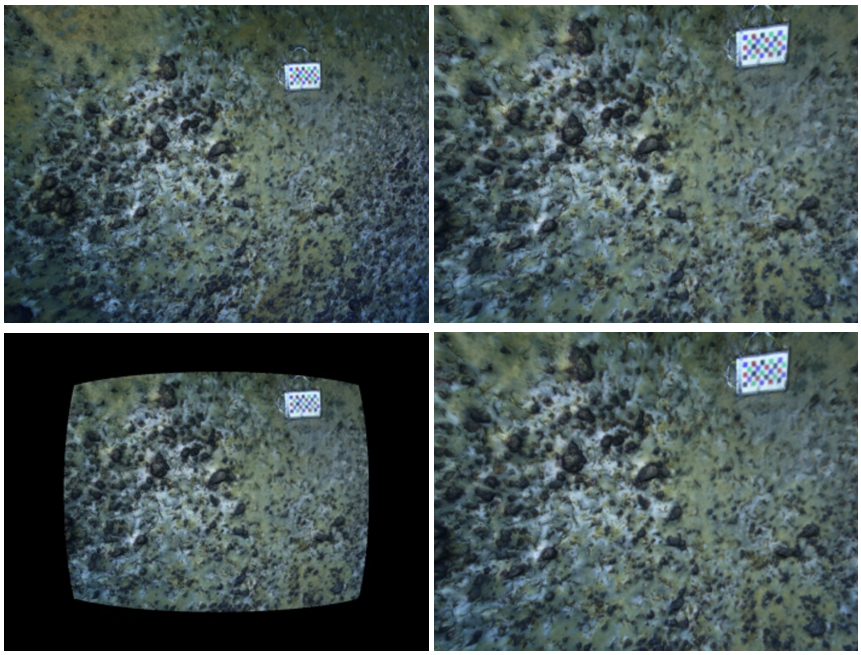


Figure 2.17: Rendered ground truth image (top left) from the pre-build 3D model, rendered refracted image (top right), refraction corrected image (bottom left), remapped image from refraction corrected image frame to original input image frame (bottom right). Please note that the bottom right color image is not needed during the workflow, only the remapped depth map under the same image frame is used.

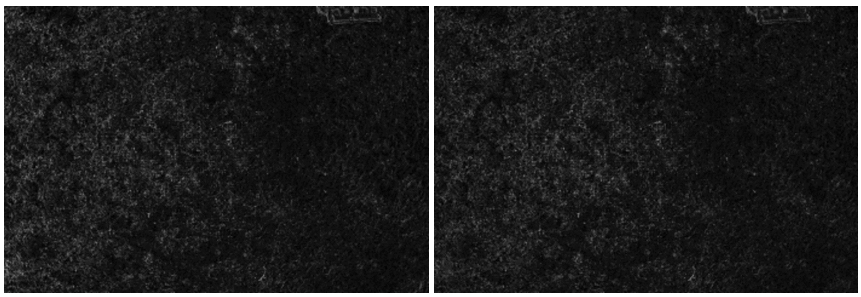


Figure 2.18: Absolute intensity differences (cropped center area) between the ground truth images and the refraction corrected images in each iterations (left: after first iteration, right: after second iteration). For a better visualization, all the values have been amplified with the factor of 10.

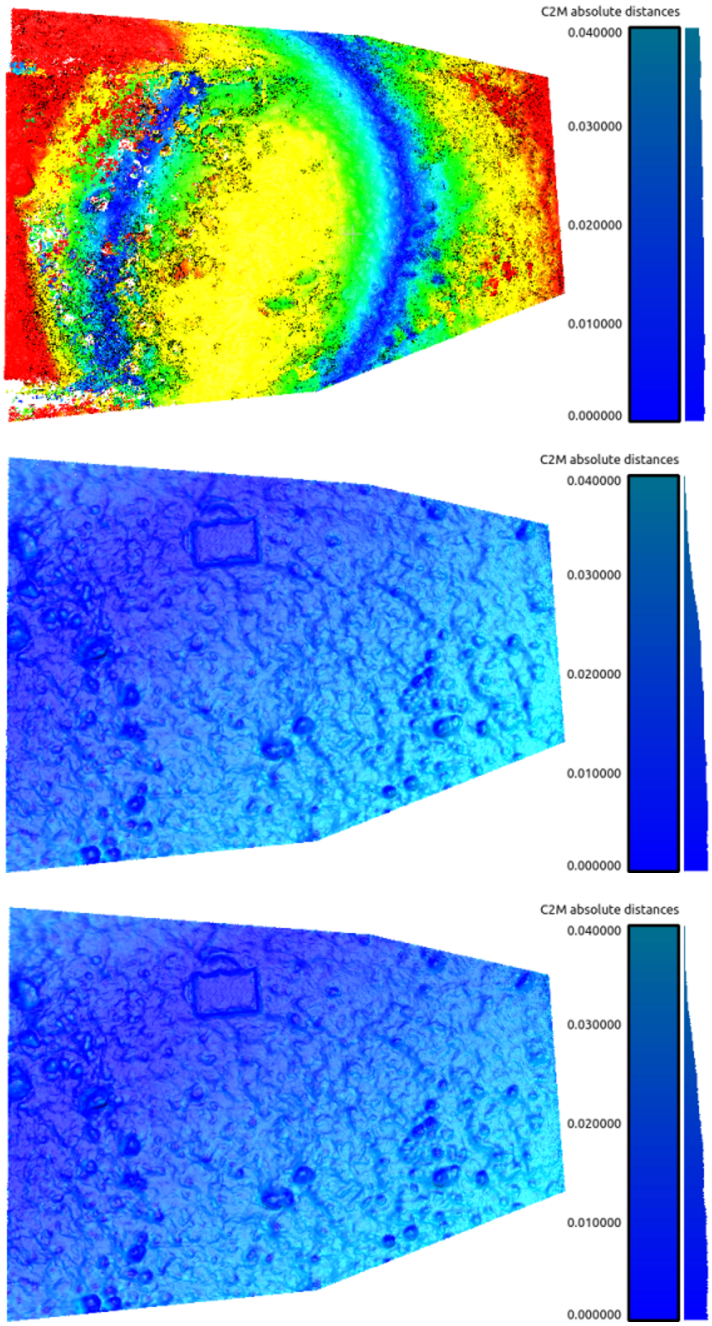



Figure 2.19: 3D reconstruction quality after each iteration. From top to bottom: iteration 0, 1, 2. (The green , yellow, red color in the first picture indicates that the error of this model is much larger than the color bar's range: green (0.06m, 0.12m], yellow (0.12m, 0.18m] , red (0.18m, $+\infty$). After two iterations, the reconstruction quality has already reached a reasonable level of accuracy.

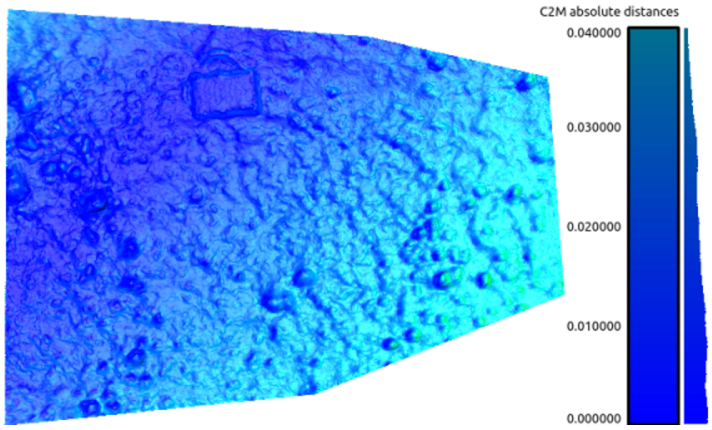



Figure 2.20: Evaluation of a 3D reconstruction standard photogrammetry pipeline with auto calibration setting.

Table 2.3: Assessment metrics for 3D reconstruction in each iterations and methods

| | Error [m] | |
|---|---|---|
| | mean | std. |
| Iteration 0 | 0.157335 | 0.102416 |
| Iteration 1 | 0.014616 | 0.009554 |
| Iteration 2 | 0.013680 | 0.008942 |
| AutoCalib | 0.016991 | 0.012032 |

calibration photogrammetry result in terms of accuracy across various metrics (see Table 2.3). Another advantage of employing the proposed approach for 3D reconstruction is that the estimated camera extrinsics can be directly applied in other applications, such as underwater vehicle navigation. This capability is not achievable with the auto-calibration photogrammetry result, as slight changes in the camera's pose are required to compensate for some of the "unmodelled refraction" errors.

# Part II

# Radiometry

# 3 Underwater Image Formation

Understanding underwater image formation is pivotal as it directly impacts the quality, precision, and interpretability of visual data captured in aquatic environments. A well-defined and accurate model that effectively captures the intricacies of light propagation underwater not only serves in the estimation of unknown water property parameters for image restoration, but also aids in synthesizing underwater images for diverse applications such as system simulations and training data generation. This chapter delves into the intricate process of pixel formation in underwater images from the radiometric perspective. Optical fundamentals of light in water presented in this chapter are based on the research from (Mobley et al. 2010; Mobley 1994).

## 3.1 Light in Water

When light interacts with a medium, it exhibits two potential effects. The light can be absorbed, transforming its energy into another form like heat or the energy within a chemical bond. This phenomenon is termed absorption. On the other hand, the light can deviate from its original trajectory, a phenomenon known as scattering. From an energy conservation perspective, as a light beam traverses through a medium, the initial incident power of the beam is divided into absorbed energy within the volume of the medium, energy scattered out of the beam in all directions and the residual light transmitted through the volume without any change in direction. In the context of natural aquatic environments, the composition of water is intricate, comprising a blend of dissolved and particulate elements. These constituents exhibit a noteworthy range of optical characteristics, their combined types and concentrations contributing to significant variations in optical properties. As a result, the optical attributes of natural waters manifest considerable discrepancies across both temporal and spatial dimensions, often deviating markedly from the optical characteristics of pure water.

Accurately discerning the individual optical properties of individual components within natural water is challenging, given that the optical behavior of water is a combined outcome influenced by a variety of biological, chemical, and geological constituents. Often, a single factor is employed to encapsulate the integrated effects of all these constituents within the water. Among the various descriptors for detailing the optical characteristics of a medium, those properties that exclusively hinge on the medium itself are termed Inherent Optical Properties (IOPs). The absorption coefficient and the volume scattering function (VSF) are the two fundamental IOPs, describing how a medium both absorbs and scatters light. Understanding these two parameters allows for a comprehensive grasp of all the IOPs of the medium..

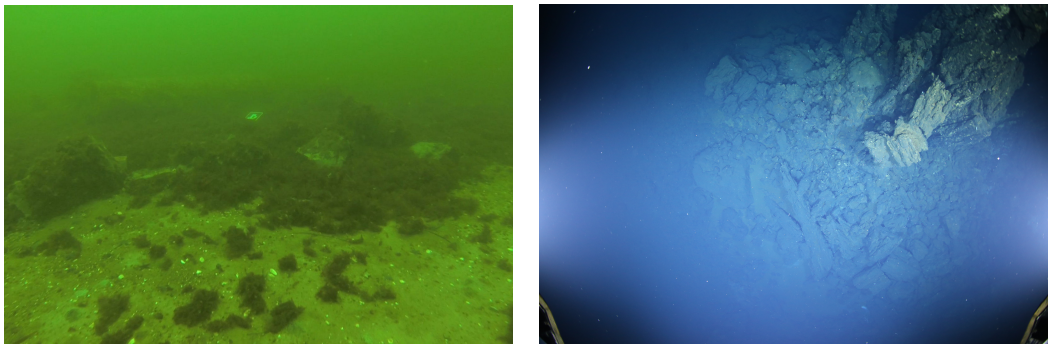


Figure 3.1: Varied water regions exhibit distinct colors as a result of varying concentrations and compositions of suspended and dissolved particles. Left: Coastal water in the Baltic Sea. Right: Deep sea water in the SE Pacific Ocean.

### 3.1.1 Physics

**Absorption** The interaction between light and water gives rise to a phenomenon known as absorption, which is heavily influenced by the wavelength. This interaction leads to a reduction in light intensity and a modification of its color. Consequently, the absorption effect in underwater images results in scenes appearing darker and undergoing a noticeable color shift.

The degree and spectral characteristics of absorption depend on various factors, including the concentration and composition of suspended and dissolved particles, as well as the properties of the water itself. Within the visible spectrum, pure water's absorption is relatively mild in the blue range but escalating towards ultraviolet, red, and infrared regions. Moreover, oceanic components such as phytoplankton, non-algal particles, and colored dissolved organic matter play a significant role in shaping water's absorption behavior. Each of these constituents exhibits unique absorption spectra, and the total optical absorption reflecting the cumulative impact of these influences. Changes in the concentration of these different components significantly impact the appearance of images. As shown in Figure. 3.1, in deep ocean waters, the absorption of light by water molecules and other dissolved substances tends to dominate, resulting in the stronger absorption of longer wavelengths (such as red and yellow) and a prevalence of shorter wavelengths (blue and green), contributing to a bluish appearance. Conversely, coastal waters often have elevated concentrations of phytoplankton, algae, and suspended particles. These microorganisms and particles selectively absorb light, but least for the green wavelengths, resulting in a more pronounced greenish tint to the water.

**Scattering** The phenomenon of scattering occurs when photons interact with a medium and deviate from their initial trajectory. Scattering is influenced by factors like refraction and the sizes of particles, but this thesis will not delve into the intricate physics of these aspects. A common parameter employed to describe scattering is the VSF. Because scattering occurs symmetrically around the axis defined by the incident light direction, the VSF is specifically dedicated to characterizing the distribution of scattered light energy in relation to the scattered angle $\psi$ from the initial direction, representing as a one-dimensional function. Here, the scattered angle $\psi$ spans from 0 to 180 degrees, defining the angle between the incoming and outgoing light rays. Once the

VSF is established (denoted as $\beta_{vsf}$), the total scattering coefficient $b$, which encompasses all directions of scattered light, can be calculated by integrating the VSF over all angles across a sphere, as expressed in the equation:

$$b \equiv 2\pi \int_0^{\pi} \beta_{vsf}(\psi) \sin(\psi) d\psi. \tag{3.1}$$

Scattering can be further categorized into two types: forward and backward scattering, based on the direction of the scattered light. The backscatter coefficient, denoted as $b_b$, pertains to the scattering of light approximately opposite to its source. This encompasses the energy accumulated over a semi-sphere, which measures the quantity of light that scatters within the angular range of $\psi$ from 90 to 180 degrees. It can be determined through the following formula:

$$b_b \equiv 2\pi \int_{\frac{\pi}{2}}^{\pi} \beta_{vsf}(\psi) \sin(\psi) d\psi. \tag{3.2}$$

The scattering is often described using the phase function form $\widetilde{\beta}_{vsf}$, which is the normalized VSF relative to the total scattering $b$:

$$\widetilde{\beta}_{vsf}(\psi) \equiv \frac{\beta_{vsf}(\psi)}{b}. \tag{3.3}$$

Numerous analytical models of phase functions have been developed to depict the angular distribution of scattered photons interacting with particles of various sizes and properties. For instance, the Mie phase function (Mie 1976) and the Rayleigh phase function (Strutt 1871) are employed to characterize light scattering by small spherical particles, finding extensive application in atmospheric research. Another prevalent model is the Henyey-Greenstein (HG) phase function, initially designed for simulating interstellar dust cloud scattering, and subsequently embraced in various domains owing to its simplicity. HG phase function has the unique feature that its free "g"-parameter value can signify either forward or backward scattering. To mitigate this limitation, the two-term HG phase function, a linear combination of two HG phase functions, was proposed (Haltrin 1999; Haltrin 2002). Additionally, the Fournier-Forand phase function (Fournier et al. 1994) and its subsequent form (Fournier et al. 1999) have gained prominence, particularly in oceanography.

Beyond analytical models, actual measurements of ocean waters have also been incorporated. A seminal contribution by (Petzold 1972), presenting meticulous scattering measurements that comprehensively span almost the entire range of scattering angles. These measurements encompass various scenarios, including in clear, coastal, and turbid ocean water (see Figure 3.2).

**Attenuation** In underwater light transportation, both absorption and scattering processes jointly contribute to diminishing the received light. Scattering not only causes light to change direction but also amplifies the impact of absorption, especially as light travels longer distances through the medium. A widely employed IOP is the beam attenuation coefficient, which encompasses both the absorption and scattering coefficients. At a specific wavelength $\lambda$, the beam attenuation coefficient is defined as $c(\lambda) = a(\lambda) + b(\lambda)$. Beer-Lambert Law states that the attenuation of light passing through a material is directly proportional to both the concentration

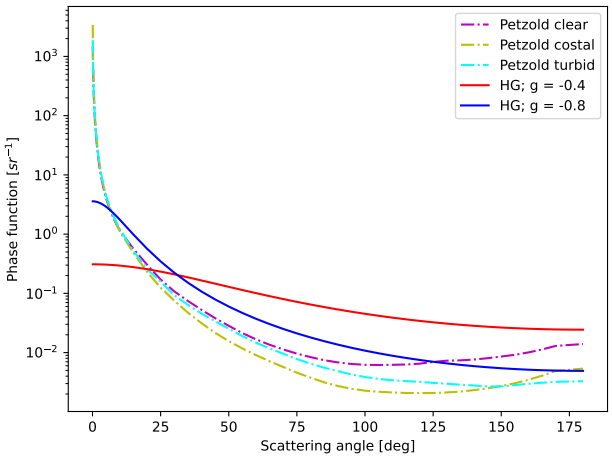



Figure 3.2: The volume scattering phase functions for three types of ocean water, as measured by (Petzold 1972), are juxtaposed with the analytical H-G phase function model, with two different parameters settings.

of the material (refers to $c$) and the length of the path $d$ that the light traverses within it. This relationship can be mathematically expressed as:

$$E(\lambda, d) = E(\lambda, 0) \cdot e^{-c(\lambda) \cdot d}. \tag{3.4}$$

Water attenuation directly signifies the decrease in light intensity during its passage through water, making it a crucial consideration in underwater image formation models. Jerlov conducted pioneering work (Jerlov 1968) by measuring and categorizing Earth's water into fourteen distinct spectra, which are commonly referred to as Jerlov water types. Considering that images are often stored in RGB three-channel format, Akkaynak further constrained the range of oceanic attenuation coefficients across a wide bandwidth (Akkaynak et al. 2017), particularly for applications in computer vision.

### 3.1.2 Artificial Light in Water

In deep sea imaging, artificial illumination plays a pivotal role. The light originates from the artificial sources attached to the vehicle and interacts with the water body in front of the camera, yielding distinct visual effects in the images, particularly in the backscatter component. As a result, developing an accurate underwater image formation model for deep sea scenarios necessitates supplementary information about the artificial light sources, including their corresponding poses and properties.

To model artificial lighting in underwater image formation, two key factors must be taken into account. Firstly, the spectral composition of artificial light sources must be considered. These sources, based on different technologies and materials, generate light with distinct spectra compared to sunlight. This spectral dependency on wavelength $\lambda$ can be obtained from the LED's

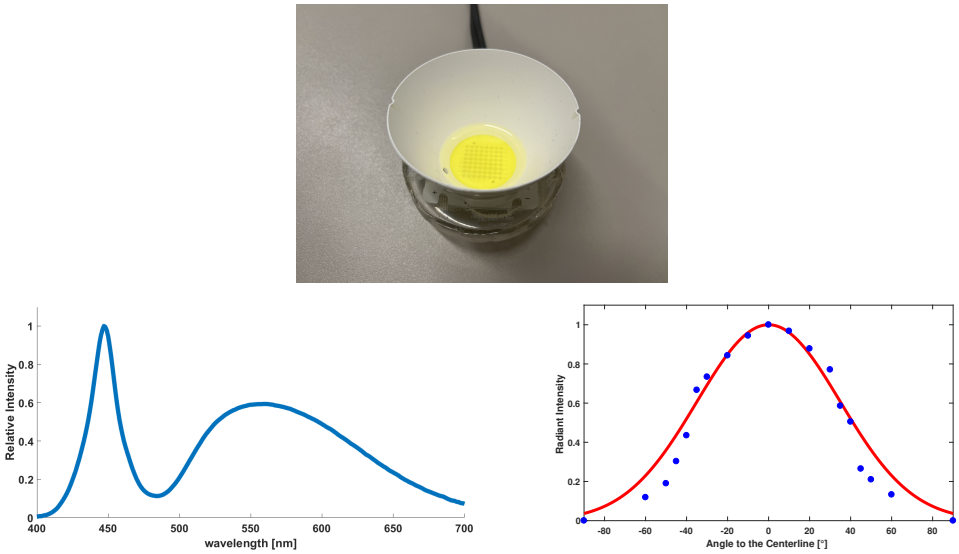



Figure 3.3: Top: An underwater LED light source featuring a BXRA-56C9000 LED array and a reflector, cast in resin. Bottom left: Spectrophotometer measurement displaying the color spectrum of the light source. Bottom right: Radiation characteristics of the light source, where the blue dots represent the lab measurement of the light source underwater, and the red line indicates its approximation using a scaled Gaussian function ($\sigma = 35°$). Data measurements courtesy of Jan Sticklus.

color spectrum curve (refer to Figure 3.3, Bottom left), often provided by the manufacturer or measured using a spectrophotometer.

Secondly, one must take into account the directionality of the light. While sunlight can be approximated as diffuse and parallel, underwater artificial light sources are predominantly directional, such as spotlights. These sources possess specific radiance distribution patterns that significantly impact deep sea image formation. This thesis focuses on spotlights, a common choice for underwater vehicles. Typically, these light sources emit the highest intensity along their central axis, with intensity diminishing as angles deviate from this axis. This angular characteristic is formulated by a radiation intensity distribution (RID) curve (Figure 3.3, Bottom right). Moreover, previous work (Bryson et al. 2016) approximated the RID using a Gaussian function, and its parameters then can be estimated during optimization. In this Gaussian model, the radiance along each light ray can be computed as follows:

$$E(\theta,\lambda) = E(0,\lambda)\cdot e^{-\frac{1}{2}\cdot\frac{\theta^2}{\sigma^2}}. \tag{3.5}$$

Where $E(\theta,\lambda), E(0,\lambda)$ are the relative light irradiance at angle $\theta$ and the maximum light irradiance along the central axis respectively.

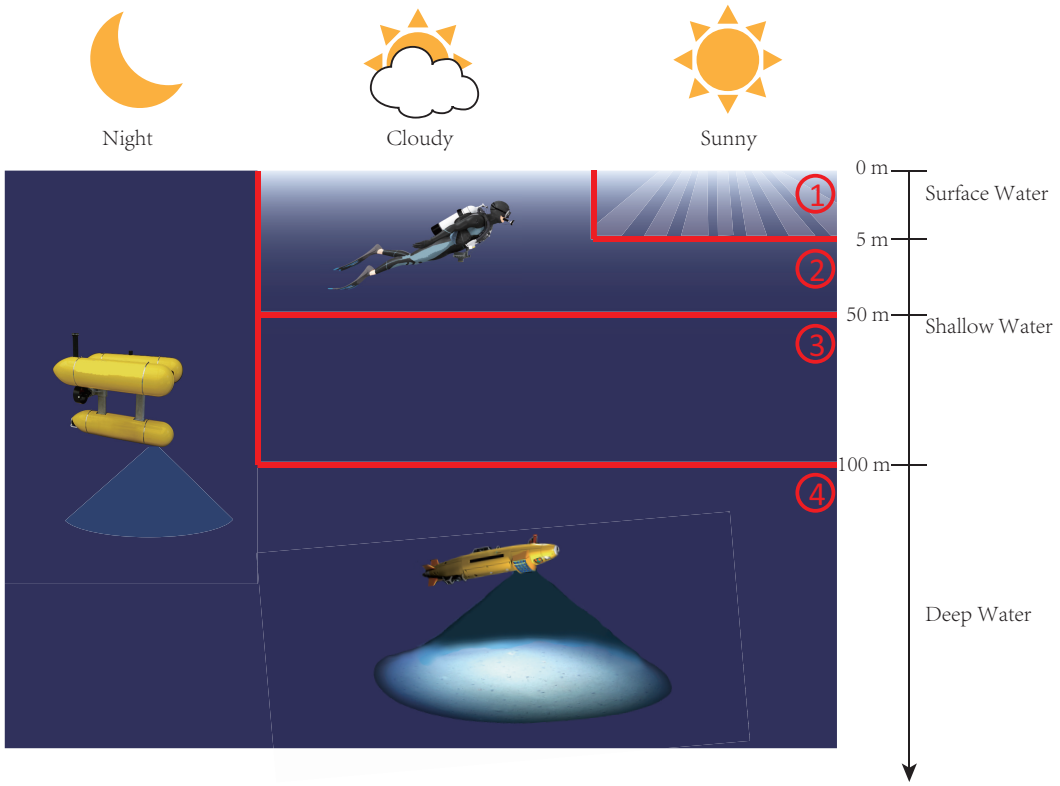



Figure 3.4: Different underwater image formations under different illumination conditions.

## 3.2 Underwater Image Formation

Underwater image formation encompasses the processes involved in capturing scenes beneath the water's surface, incorporating various optical phenomena and challenges unique to underwater environments. A comprehensive understanding of these factors is crucial for the development of effective imaging techniques, correction methods, and technologies. These advancements are aimed at enhancing the quality of underwater photographs and videos, thereby facilitating further analysis and applications derived from these images.

Furthermore, a well-defined underwater image formation model can serve as a valuable tool for tasks such as estimating 3D geometry underwater or determining water optical properties. This section delves into underwater image formation models under different illumination conditions. It also conducts a detailed investigation into image formation in deep water environments, contributing to the synthesis of realistic deep underwater images.

### 3.2.1 Underwater Images with Different Illumination Conditions

In underwater imaging, the physical-based underwater image formation can be classified into four categories according to their illumination conditions (see Figure. 3.4) and commonly used models:

- **Type I**: *Surface water model*. This model describes the image formation in the water surface region where the scene is completely illuminated by sunlight. Its basic image formation model is similar to Type II, but strong sunlight is refracted dynamically at waves of the water surface, producing additional caustic patterns in the scene. The caustic patterns are constantly changing due to the water surface and it is challenging to predict the caustic pattern in the water based on physical models, as it requires information such as water surface normal, water depth, geometry of the scene and the relative position of the sun.
- **Type II**: *Shallow water model*. This model is by far the most popular model which has been widely applied in underwater image processing methods. It descends from atmospheric dehazing which has originally been used to recover the depth cues from images affected by haze or fog. In this model the underwater image formation is composed by direct attenuated light and ambient light(backscatter). The sunlight first travels from the water surface to the seafloor, and is then reflected to the camera. The scattering of the sunlight in the first path, known as veiling light, depends on the water depth, but the attenuated light in the same region of water is relatively homogeneous, allowing for it to be approximated as the background color. The attenuation of object intensity is only considered in the second path, resulting in the corresponding image formation model becoming a weighted linear combination of object intensity and background color (backscatter).
- **Type III**: *Mixed model*. This type of model combines characteristics of both Type II and IV models. While the natural sunlight is not enough to illuminate the scene, the ambient illumination is not completely dark, and thus additional artificial illumination is required to supplement the illumination.
- **Type IV**: *Deep water model*. When the region is devoid of sunlight, the scene is illuminated solely by artificial light sources co-moving with the camera. In this image formation model, the signal is still a sum of direct and backscattered light (forward scattering effect is often approximated as an extra smoothing of the signal). However, the attenuation of light in water now needs to consider the path from the artificial light source to the object and then to the camera. It must also be taken into account that the artificial light sources have spectrum different from sunlight. The total backscatter in the scene is no longer represented by a single, uniform background color. Instead, it is an integral of water body scattering along each viewing ray, which depends on the configuration of the artificial illumination and the water properties such as the VSF. The most popular model is the Jaffe-McGlamery model.

Example images for each type are illustrated in Figure 3.5.

### 3.2.2 Underwater Image Formation Models

Several approximations to the low-level physical model have been proposed in the previous literature, including assuming a macroscopic atmosphere-like fog model for shallow water, a single scatter-model for artificial light sources and numerical/discretized simulation of the full problem using Monte-Carlo-based methods. Two popular physical models are the Atmospheric Fog (AF) Model (Equation 3.6) and the Jaffe-McGlamery (J-M) Model (Jaffe 1990; McGlamery 1980) (see Figure 3.6).

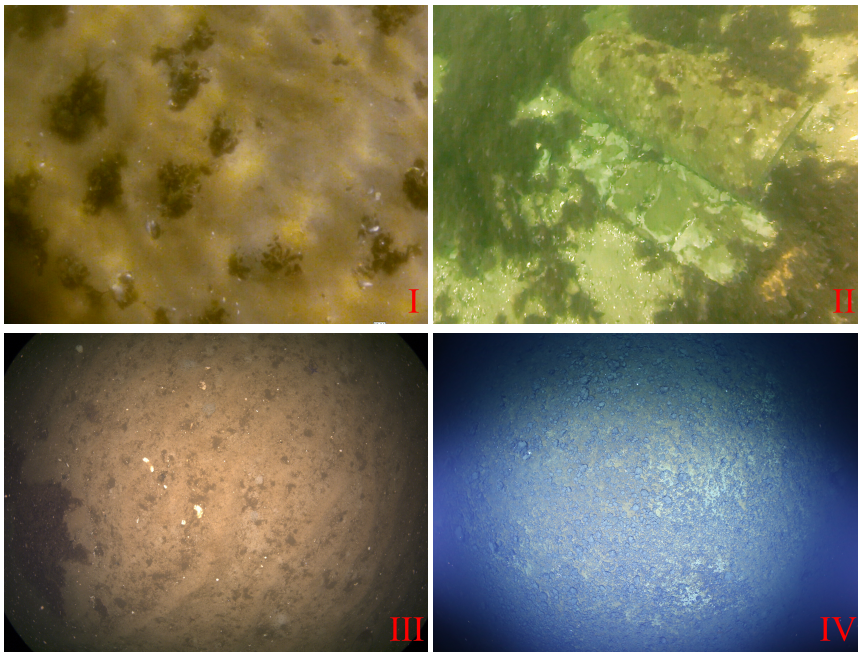



Figure 3.5: Examples of underwater seafloor images captured under different illumination conditions, each corresponding to a different image formation model. **I**: In surface water where the strong sunlight creates a dynamic caustic pattern. **II**: In shallow water where the illumination is relatively homogeneous due to the abundant sunlight. **III**: in the twilight zone where the sunlight is severely attenuated and additional artificial light is used to illuminate the scene. **IV**: In complete darkness in the depth ocean and is illuminated solely by artificial light sources.

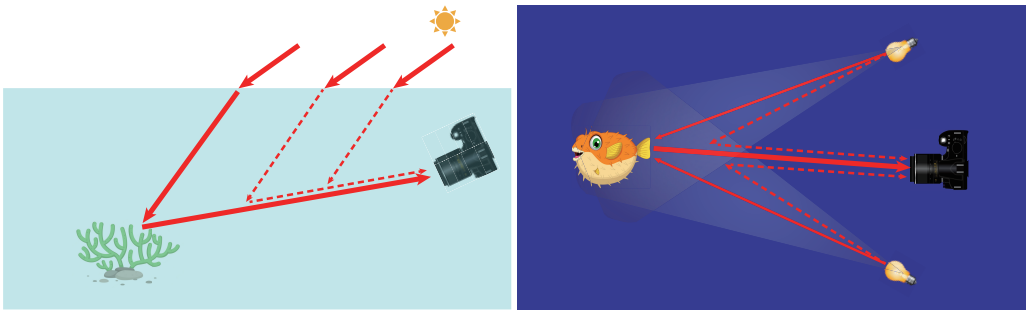


Figure 3.6: Two popular underwater image formation models used in underwater image restoration. Left: Shallow water image formation with homogenous illumination from the sunlight. Right: Deep water image formation under artificial illumination.

The AF Model stands out as the predominant underwater image formation approximation, finding extensive application in shallow water scenarios. Derived from the atmospheric scattering model (Cozman et al. 1997), this model characterizes the underwater image $I(\mathbf{x})$ as a linear combination of the attenuated signal and the backscatter. The AF Model (and its modifications) is widely used due to its simplicity. It assumes the scene is illuminated homogeneously and the total backscatter is defined by a background light (also named water color, veiling light *et al.*), which depends on the water. Each pixel in an underwater image is described as a weighted combination of the true color and the background light, and the underwater color is interpolated from these two values:

$$I(\mathbf{x}) = T(\mathbf{x}) \cdot I_0(\mathbf{x}) + (1 - T(\mathbf{x})) \cdot B_\infty. \tag{3.6}$$

where $I_0(\mathbf{x})$ represents the object color without any perturbation at pixel location $\mathbf{x}$ and $B_\infty$ denotes the "pure" water color. The transmission map $T$ is often expressed by $T(x) = e^{-\eta d(x)}$, which comprises the water attenuation effect. Here $\eta$ is the attenuation coefficient and $d$ is the scene distance.

The advantage of the AF Model is that it only contains a few parameters (no integral involved) and does not require many redundant observations from multiple images. The information extracted from a single image is sufficient to estimate these unknown parameters. However, the drawback of this model is also obvious: $B_\infty$ is not able to describe complex total backscatter patterns, especially for Type III and IV images.

The J-M Model is a more complex underwater image formation model that addresses the manifold scattering pattern cased by artificial point light sources. In this model, the light reaching the camera is conceptualized as the sum of three distinct components (direct signal, forward scattering, and backscatter), all subject to the influences of light attenuation and scattering. Unlike the AF Model, the J-M Model necessitates the consideration of attenuation across the entire light path—from the source to the scene and then to the camera. Additionally, the representation of backscatter in the J-M Model differs; rather than being portrayed as a uniform background color, it is expressed as the cumulative scattered light along each camera viewing ray within the viewing frustum. As a result, the J-M Model is better suited for scenarios where the scene is illuminated by artificial light sources. Its efficacy is particularly pronounced when leveraging knowledge about

the relative geometry between the camera, the underwater scene, and the light sources. Further elaboration on the intricacies of the J-M Model will be provided in Section 3.3.1.

## 3.3 Deep Sea Image Simulation

Deep sea images are often marred by degradation caused by scattering, attenuation, and the influence of artificial light sources, resulting in a markedly distinct appearance compared to images in shallow water or on land. This impairs transferring current vision methods to deep sea applications. The recent trend to employ machine learning methods for various vision tasks even increases the performance gap between underwater vision and approaches on land, given that machine learning typically demands substantial training data for optimal results. Unfortunately, the scarcity of suitable underwater images, particularly from deep sea environments, coupled with accurate ground truth data, acts as a bottleneck for the development of methods in this domain. One potential avenue to ameliorate this challenge is the simulation of deep sea images, including the incorporation of illumination, attenuation, and scattering effects, which could serve as valuable resources for developmental and training purposes in the realm of underwater visual perception. Moreover, due to the intricacies and costs associated with developing and testing deep sea imaging and illumination systems, a simulation system for generating deep sea images can prove invaluable for aiding system design, mission planning, and related endeavors.

This thesis presents a physical model-based image simulation solution. It utilizes in-air texture images along with corresponding depth maps as inputs to simulate underwater image sequences captured by UVs in deep ocean environments. The primary goal is to provide a solution that achieves realistic rendering while maintaining the unique characteristics of deep sea images. Additionally, it has to be lightweight and efficient, allowing for interactive simulations without relying on resource-intensive GPU computations. This adaptability makes it easily integrable into other simulation platforms.

### 3.3.1 Adapted Jaffe-McGlamery Model

As discussed in the preceding section, the AF model is inadequate for modeling the intricate artificial lighting effects in deep sea scenarios, where the J-M model proves to be more suitable. The choice of employing the J-M model is motivated not only by its suitability for depicting underwater effects in the presence of artificial lighting, but also by its simplified form. The model's parameters are differentiable, making them estimable for various potential applications, such as estimating water IOPs, scene depth, or lighting parameters.

Expanding upon the foundational structure of the J-M model, which dissects the incoming light recorded by the camera ($E_{total}$) into three fundamental constituents: the direct signal ($E_{direct}$), the forward-scattering ($E_{fs}$) and the backscatter ($E_{bs}$) components:

$$E_{total} = E_{direct} + E_{fs} + E_{bs}. \tag{3.7}$$

Building upon the models proposed by (Jaffe 1990; Sedlazeck et al. 2011), this thesis makes a substantial advancement by incorporating the artificial lighting models detailed in Section 3.1.2 into the existing framework. These models encompass the complete spectrum of illumination as

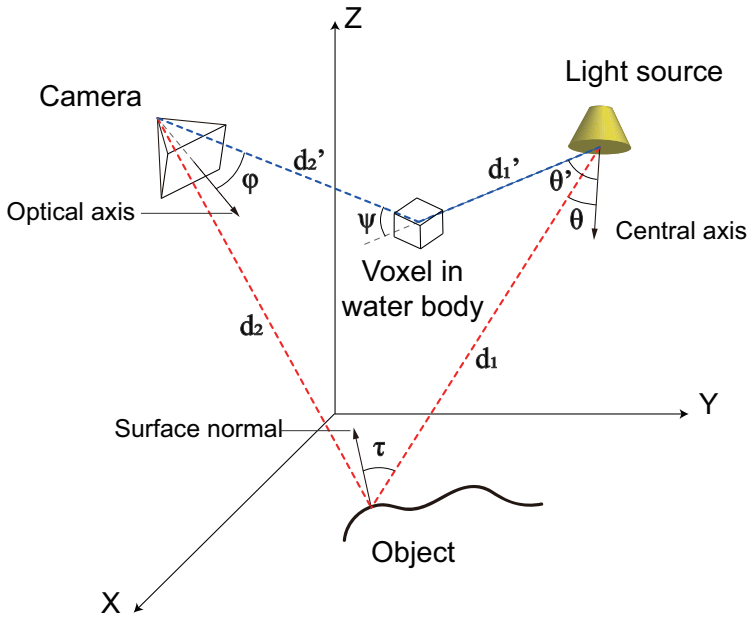



Figure 3.7: The adapted J-M model with spotlight. Red dashed lines delineate the path of the direct signal's light transportation, while blue dashed lines illustrate the trajectory of scattered light. Figure adapted from (Sedlazeck et al. 2011).

well as intricate parameters such as the orientation and angular attributes of multiple spotlight sources. The adapted J-M model, purposefully tailored for the simulation of deep-sea imagery, as well as the corresponding parameters are depicted in Figure 3.7. The computation of the three components of the model is intertwined with the configurations of the light sources as follows:

**Direct Signal** As light traverses through water, it undergoes attenuation, with the extent of irradiance loss contingent upon both the distance covered and the inherent properties of the water. When delving into the propagation of artificial light from sources, this attenuation becomes a pervasive phenomenon, occurring throughout the entire trajectory spanning from the light sources to the underwater object and subsequently to the camera. Given the spotlight's characteristic point source nature, the application of the Inverse Square Law becomes imperative. This principle captures the quadratic decline in light irradiance over distance from the originating point source. Furthermore, the object reflection model, an assumption that light is uniformly diffused in all directions across the object surface (Lambertian surface), is incorporated, the irradiance of the reflected light hinges solely on the angle formed between the incident light and the surface normal. The computation of the direct signal, encompassing the interplay of artificial light sources, attenuation effects, and reflection model, can be expressed as follows:

$$
\begin{aligned}
E_{direct}(\mathbf{x},\lambda) &= \sum_{i=1}^{n} E_{direct_i}(\mathbf{x},\lambda) \\
E_{direct_i}(\mathbf{x},\lambda) &= I_0(\mathbf{x},\lambda)\cdot E(\theta_i,\lambda)\frac{e^{-c(\lambda)\cdot(d_{1_i}+d_2)}}{d_{1_i}^2}\cos\tau_i.
\end{aligned}
\tag{3.8}
$$

Here, $E_{direct}(\mathbf{x},\lambda)$ denotes the cumulative irradiance reaching the pixel $\mathbf{x}$ within the image. It results from the aggregation of the direct signal components $E_{direct_i}$($\mathbf{x}$ originating from $i$th the artificial light sources. $I_0$ indicates the albedo of the object. $E(\theta,\lambda)$ embodies the relative irradiance of light at a given angle $\theta$. The attenuation factor, represented as $c$, quantifies the magnitude of irradiance loss through a particular type of water at the specific wavelength $\lambda$. $d_{1_i}$ represents the distance from the $i$th light source to the object, while $d_2$ signifies the distance from the object to the camera. The variable $\tau$ denotes the incident angle formed between the light ray emitted by the light source and the surface's normal vector. Note that the denominator, which delineates the quadratic falloff of irradiance characteristic for point light sources, explicitly pertains to the path $d_1$. This occurs because as $d_2$ increases, each pixel in the camera will essentially receives the light from an expanded surface area, which effectively compensates for each other, counterbalancing the effects.

**Forward Scatter** In the J-M model, scattering phenomena are partitioned into two distinct components: forward scatter and backscatter. Forward scatter characterizes the scattering phenomenon wherein light deviates by an exceedingly small angle, leading to a perceptible loss of scene sharpness in the resultant images. This thesis approximates the influence of forward scatter through the utilization of a Gaussian filter (similar to (Schechner et al. 2004)), denoted as $g(\overline{d_2})$, with the dimensions of the filter mask contingent upon the local average depth of the scene $\overline{d_2}$. The forward scattering model from light sources to the scene is omitted as the RID curve associated with the light is usually very smooth (e.g. modeled as a Gaussian function like Equation 3.5). This smooth curve obviates the need for marginal additional smoothing to account for the forward scatter effect during this path. The calculation of forward scatter can be perceived as the smoothed direct signal, achieved through a convolution operation denoted by $*$:

$$
E_{fs}(\mathbf{x},\lambda) = E_{direct}(\mathbf{x},\lambda) * g(\overline{d_2}). \tag{3.9}
$$

**Backscatter** Among the three components that constitute the J-M model, the backscatter component emerges as the most intricate due to its inherently cumulative nature. It refers to light rays that interact with the water and are subsequently scattered backwards to the camera. This leads to a "veiling light" effect in the medium. Following the approach presented in (McGlamery 1980), the 3D viewing frustum located in front of the camera can be discretized by slicing it into several slabs with certain thicknesses. The backscattered irradiance on each individual slab can then be accumulated along the viewing rays until they intersect with the object's surface. This meticulous process is undertaken to compose the backscatter component within the resulting image.

$$\begin{cases} E'_i(\mathbf{x},\lambda) = E'(\theta_i,\lambda)\frac{e^{-c(\lambda)(d'_{1_i}+d'_2)}}{d'^2_{1_i}} \\ E'_{fs_i}(\mathbf{x},\lambda) = E'_i(\mathbf{x},\lambda) * g(\overline{d'_2}) \\ E_{bs_i}(\mathbf{x},\lambda) = \sum_{j=1}^{n} \beta_{vsf}(\psi)[E'_{i,j}(\mathbf{x},\lambda) + E'_{fs_i}(\mathbf{x},\lambda)]\Delta z_j. \end{cases} \tag{3.10}$$

Equation 3.10 encapsulates the process for computing the backscatter component originating from each individual light source $i$. In this equation, $n$ denotes the number of slabs within the water body situated between the object and the camera. $E'_{i,j}(\mathbf{x},\lambda)$ represents the direct irradiance reaching a given slab ($j$), and $d'_{1_i}$ and $d'_2$ signify the distances from a voxel within the slab to the respective light source and camera. The forward scattering component $E'_{fs_i}(\mathbf{x},\lambda)$, emerges through convolution of $E'_i(\mathbf{x},\lambda)$ with the Gaussian filter $g(\overline{d'_2})$, where $*$ denotes the convolution operator. $E_{bs_i}(\mathbf{x},\lambda)$ is the resultant cumulative backscatter component, aggregating all backscattered light interactions within the water body along the viewing ray until it reaches the object's surface. Here, $\beta_{vsf}(\psi)$ accounts for the VSF, with $\psi$ representing the angle between the scattered light ray and its original direction, as visually depicted in Figure 3.7. $\Delta z_j$ corresponds to the thickness of the $j$th slab within the model. In this thesis, the VSF model can be achieved in one of two forms: it can either rely on real-world measurements as detailed in (Petzold 1972), which are stored as a lookup table for direct reference, allowing for the interpolation of scattered light irradiance based on this table; alternatively, an analytical phase function model, as expounded upon in Section 3.1.1, can be employed, with an additional scaling factor applied in proportion to the total scattering coefficient $b$.

**Camera Optical Model** The camera optical model consists of two fundamental aspects: the geometric model and the radiometric model. The geometric model encompasses typical camera projection geometry, often described by intrinsic camera parameters. Additionally, underwater camera system exhibits refractive distortions that can arise due to multi-media refraction effects. As previously discussed in Section 2.4, underwater dome ports can be adjusted to mitigate refraction. This thesis considers the inclusion of refraction as an optional step for underwater renderer. If necessary, it can be incorporated using the method introduced in Section 2.5.2.

The radiometric model of a real digital camera encompasses various optical effects attributed to the camera's lens system, such as vignetting and lens transmittance. The camera optical model also considers sensor characteristics, especially the conversion of analog light intensity to digital pixel values. This thesis adopts the optical effect model presented in (Jaffe 1990). This model is contingent on several factors, including the transmittance of the camera lens ($T_l$), the f-number ($f_n$), and the angle ($\phi$) between the incident light ray on the camera lens and the camera's optical centerline. Regarding the digital signal transformation, it is assumed that the camera exhibits a linear response function. Additionally, three other factors are incorporated into the model. These include $k_{wb}(\lambda)$, which represents white balance parameters affecting the relative intensity for different color channels, and $k_{rs}(\lambda)$, representing the camera's relative sensitivity for the RGB channels. Finally, there is a global scaling factor denoted as $s$, which encompasses various effects such as the absolute irradiance of the light source, ISO settings, exposure time of the camera, and the analog-to-digital (A/D) conversion process.

The final pixel value $I(\mathbf{x}, \lambda)$, can be computed from the total irradiance $E_{total}(\mathbf{x}, \lambda)$ (comprising direct signal and scattering components) that reaches the camera lens. This calculation incorporates the camera's optical and digital signal processing effects which is defined as follows:

$$I(\mathbf{x}, \lambda) = s \cdot k_{wb}(\lambda) \cdot k_{rs}(\lambda) \cdot \frac{cos^4 \phi T_l^2}{4 f_n} \cdot E_{total}(\mathbf{x}, \lambda). \tag{3.11}$$

However, these models find their essentiality primarily in contexts where image simulation is tailored for a specific real camera system, often serving purposes like system design or calibration. Therefore, they have been crafted to operate as optional components within the renderer. It will only come into play when the simulation involves real-world camera and lighting systems, making it a selectable feature rather than a default component.

### 3.3.2 Deep Sea Image Simulation from Virtual Scenes

This section demonstrates the implementation of the deep-sea robotic imaging simulator, utilizing the image formation model discussed earlier. The primary objective of this simulator is to facilitate the creation of authentic images portraying the deep-sea environment. These images are pivotal for the advancement of deep-sea robotic perception systems and the design of imaging technology tailored to underwater exploration. For seamless integration into Unmanned Underwater Vehicle (UUV) simulation platforms, the renderer is meticulously designed for lightweight performance and rapid rendering speeds. Simultaneously, it must encompass all the nuanced light and water effects required to produce realistic images. In pursuit of this equilibrium, several optimizations have been implemented within the rendering pipeline.

#### Optimizations for Rendering

In deep sea image simulation, one of the most computationally costly parts is the simulation of the backscatter component. Backscatter occurs when a light ray originating from a light source directly intersects the water body situated between the camera and the 3D scene, resulting in the scattering of light towards the camera. This phenomenon is cumulative in nature within the image. Calculating the backscatter component necessitates the computation of backscattered signals at numerous sampling points along the viewing ray. However, when the relative geometry between the camera and the light source remains fixed, and the water properties remain constant, the backscatter component essentially remains constant within the 3D viewing frustum situated in front of the camera . This means that if there are no objects present and only water exists within the camera's field of view, the resulting image will exhibit a relatively constant appearance, comprising the total backscattered light directed towards the camera. Once objects come into view within the scene, the water volume which contains backscatterred light is effectively partitioned, contingent upon the distance between the object and the camera. The portion of this volume that remains unobstructed is systematically accumulated to generate the image's backscatter component.

To this end, this thesis employs a structure where a 3D viewing frustum within the camera's field of view is sliced into several volumetric slabs (McGlamery 1975), each having certain thicknesses that are all parallel to the image plane (refer to Figure 3.8). Each slab rasterized into unit

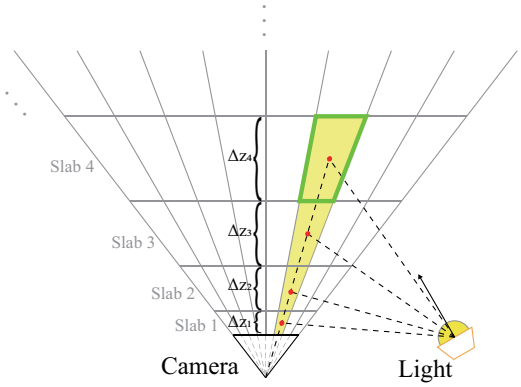



Figure 3.8: Pre-rendered backscatter field, each unit cell in the slab (green) stores the accumulated backscatter component (yellow) along the camera viewing ray.

cells according to the individual pixels in the resulting image. In order to expedite the rendering of backscatter components across an image sequence and to avoid redundant computations, a proactive approach is employed. Cumulative backscatter elements are pre-computed for each individual unit cell and then systematically stored within a dedicated 3D lookup table. Considering that the backscatter component for each pixel entails an integration across all illuminated slabs, with each slab's contribution scaled by its corresponding thickness along the viewing ray, the process of calculating backscatter for a pixel at a given depth $d$ is streamlined. This simplification is achieved by directly interpolating the value between the two nearest unit cells along the viewing ray.

During the pre-rendering phase, when analyzing the backscatter signals within the slabs, a notable observation emerges. In underwater scenarios involving UUVs, particularly when the object of interest is considerably more distant from the camera than the light source, the scattering in each slab becomes smoother and eventually disappears in the far field (see Figure 3.9 and 3.10), the appearance of the backscatter component in an image is predominantly influenced by the irradiance from the water volume in close proximity to the camera. What becomes apparent is that slabs located nearer to the camera exhibit more dynamic and intricate backscatter patterns. Conversely, slabs situated at greater distances tend to manifest similar and weaker patterns. These distinctive backscatter patterns are profoundly contingent upon the relative positioning of the light source(s) and exhibit substantial variation across various camera-lighting systems. This variability represents a significant contrast to shallow water scenarios, where even at considerable distances from the camera, abundant and uniform sunlight continues to be available.

To achieve the goal of accurately simulating the backscatter component with a limited number of slabs, this thesis introduces an novel approach: an adaptive slab thickness sampling function founded on the Taylor series expansion of the exponential function. The primary objective of this function is to achieve a more concentrated sampling of slabs near the camera and sparser sampling at greater distances under the limited total number of slabs. This adaptive sampling technique optimizes the allocation of computational resources, ensuring that the simulation remains efficient without compromising the fidelity of the backscatter component.

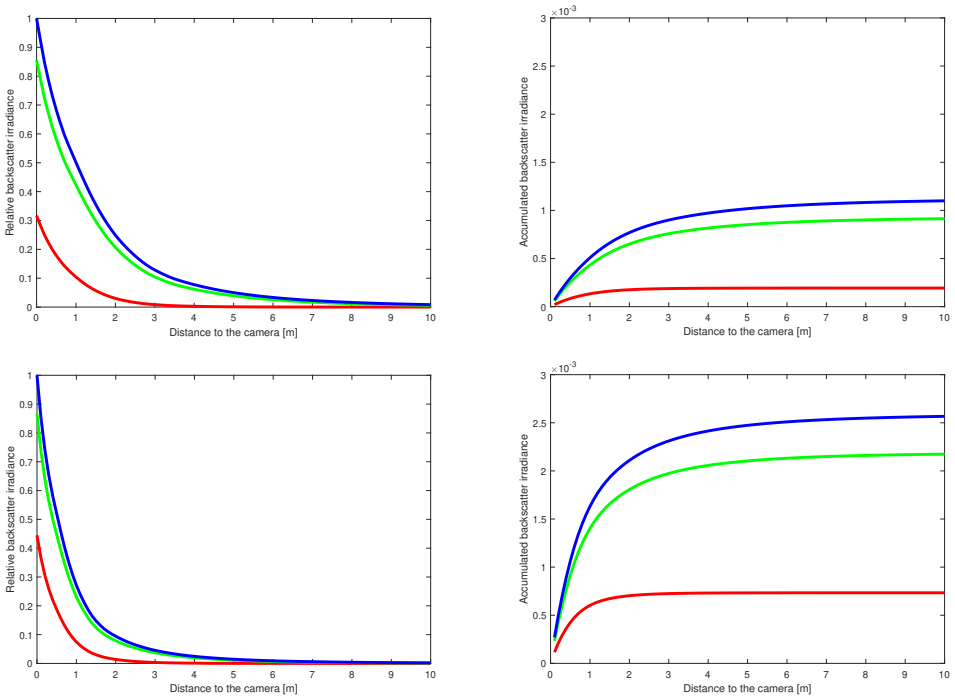



Figure 3.9: Ideal point light source positioned at 2m (top row) and 1m (bottom row) lateral distance from a camera in Jerlov water type II. Left: Irradiance observed by a camera pixel from scattering at given distances (horizontal axis). Right: Cumulative curve showing how much scattered light is collected along the viewing ray up to a certain distance. It can be seen that the curve saturates soon, i.e. most of the scattered light originates from the first few meters.

$$\Delta z_j = s \cdot \frac{n^{(j-1)}}{(j-1)!} \qquad (j = 1, 2, \ldots, n) \tag{3.12}$$

where $\Delta z_j$ represents the slab thickness of slab index $i$. The scale factor $s$ is defined as $s = 2.2 \cdot d_{max}/e^n$, with $d_{max}$ denoting the maximum depth of the scene field, which is subdivided into a total of $n$ slabs. Here, $e^n$ is the normalization factor for the Taylor series and the term $2.2 \cdot d_{max}$ guarantees that the slab thickness monotonically increases within the range $(1 < j < n)$. Furthermore, it ensures that $\sum_{j=1}^{n} \Delta z_j \approx d_{max}$ for cases where $n > 3$. This equation results in denser sampling of slabs in proximity to the camera, where dynamic changes in the backscatter pattern are more pronounced. As illustrated in Figure 3.11, under the lighting conditions described in its caption, the brightest spot should be located at the bottom right corner of the image. The utilization of the slab thickness sampling equation given by Equation 3.12 yields a more realistic and plausible rendering of backscatter effects compared to the equal-distance sampling approach.

The value of maximum depth of the scene $d_{max}$ is also an important factor that significantly influences both the quality and performance of backscatter rendering. Figure 3.9, it provides valuable insight by illustrating the normalized backscattered irradiance of voxels along the optical axis

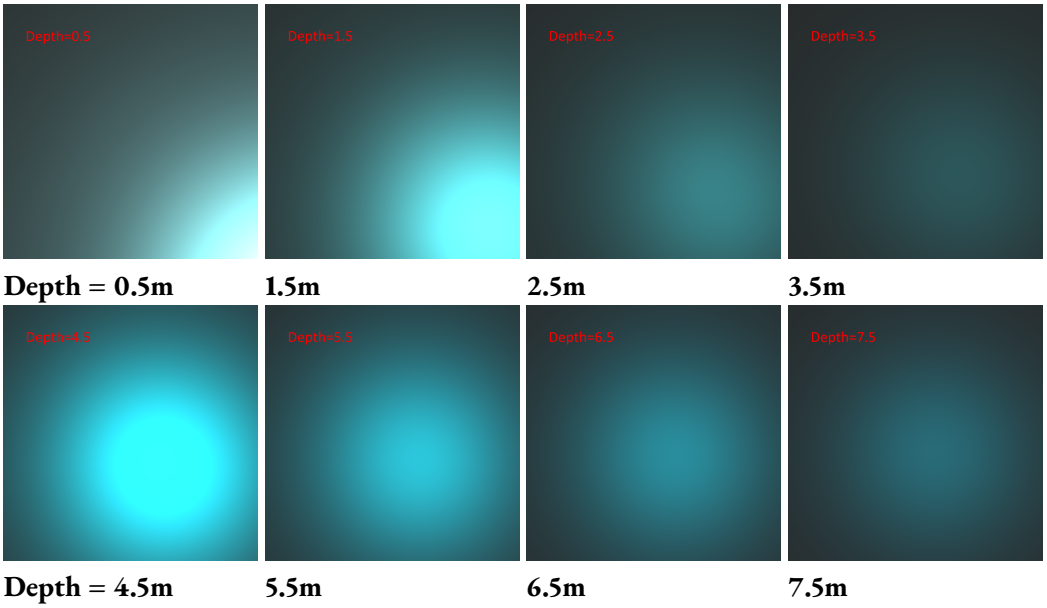



Figure 3.10: Backscatter components of different slabs from 0.5m to 7.5m depth with the same camera-light settings as Figure 3.11 (Second row images' intensities are amplified 10 times).

of the camera in deep ocean water. This figure serves as a reference for determining the appropriate $d_{max}$ when simulating underwater images under various conditions or settings.

### Rendering Pipeline

The complete workflow of deep sea image rendering from RGB-D images is outlined below and illustrated in Figure 3.12.

1. Establish the 3D backscatter lookup table, where each unit cell accumulates backscatter elements along the viewing ray from the camera, calculated using Equation 3.10.

2. Generate the direct signal component while accounting for attenuation and object surface reflection, as described in Equation 3.4.

3. Compute the forward scattering component by applying a Gaussian filter to smooth the direct signal.

4. Interpolate the backscatter component from the backscatter lookup table based on the depth values obtained from the depth map.

5. Assemble the underwater color image by combining the direct signal, forward scattering, and the backscatter component.

6. Optionally, introduce a refraction effect to the image.

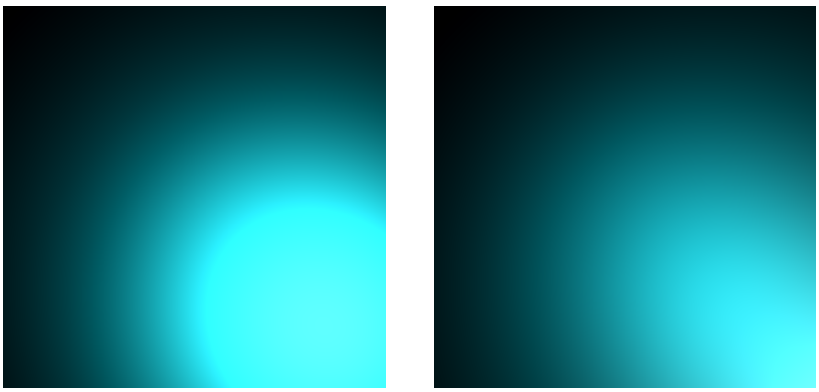


Figure 3.11: Rendering of backscatter component under the same setups ($d_{max} = 10$m, $n = 3$, single light which is at position (1m, 1m, 0m) in camera coordinate system and pointing parallel to the camera optical axis.) with different slab thickness sampling approaches. Left: by equal distance sampling, Right: by the sampling function introduced in Equation 3.12.

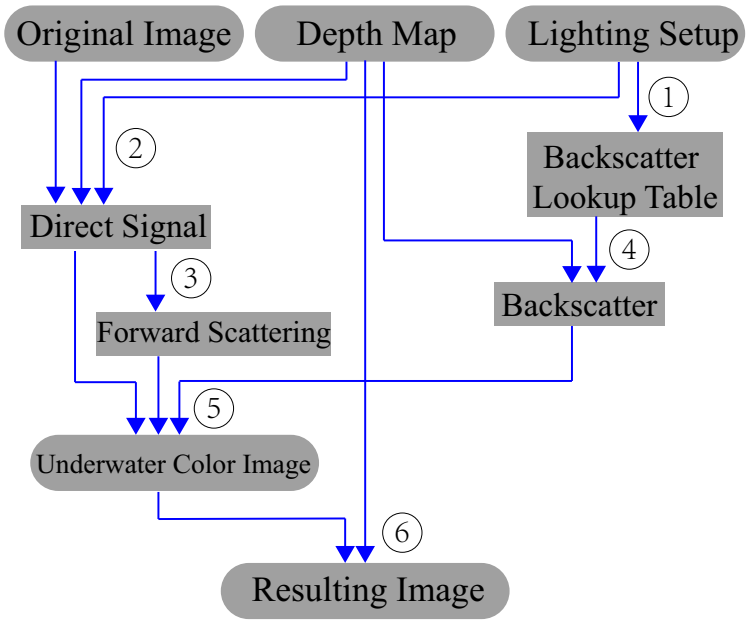
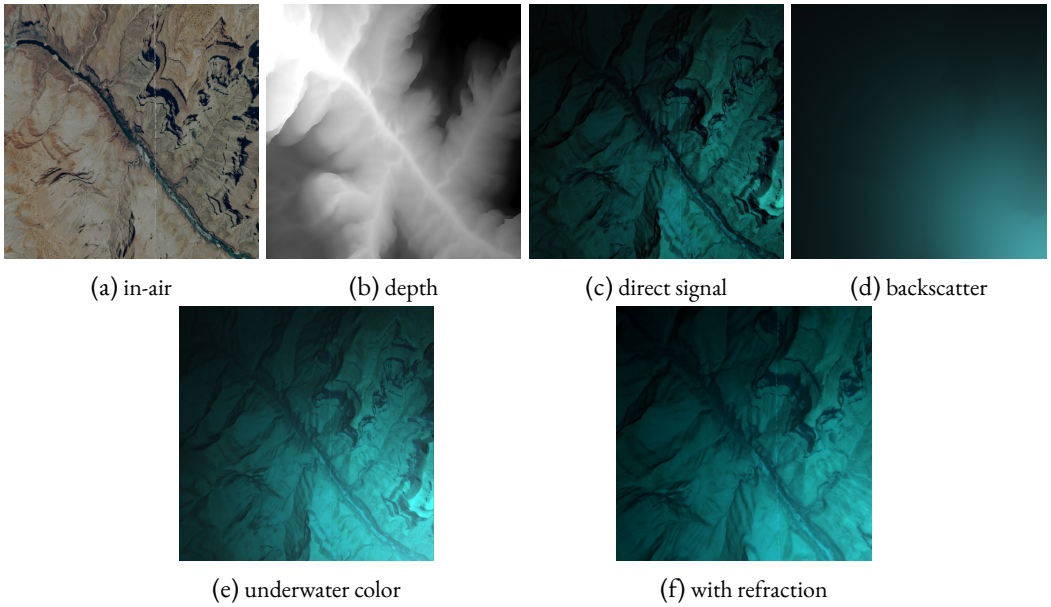



Figure 3.12: Deep Sea image rendering pipeline.



(a) in-air (b) depth (c) direct signal (d) backscatter

(e) underwater color (f) with refraction

Figure 3.13: Deep sea image simulation results.

Figure 3.13 illustrates the rendering results along with their corresponding intermediate stages: (a) and (b) represent the inputs obtained from the RGB-D sensor plugin in normal robotic simulation platforms. Subsequently, the direct signal (c) and backscatter (d) components are individually computed, afterwards the simulated underwater color image (e) is created by combining the direct signal, the smoothed direct signal (forward scattering), and the backscatter. In the end, on optional refraction effect is introduced to the underwater color image in (f) using the procedure outlined in Algorithm 2 in Section 2.5.2.

### 3.3.3 Integration in Robotic UUV Simulation Platform

A prominent application of deep sea image simulator is its integration into UUV simulation platforms. These platforms demand rendering that balances between being lightweight and comprehensive, especially for simulating the intricate effects of deep underwater lighting and viewing by the camera. Traditional ray-tracing solutions are often too resource-intensive to achieve real-time performance. By integrating the proposed simulator, developers can effectively develop, test, and coordinate the performance of underwater robotic systems without risking expensive hardware in real-world applications.

As cameras become increasingly integral to robotic systems, existing robotic simulators have started incorporating simulations of regular cameras and depth sensors, paving the way for potential extensions to underwater imaging scenarios. For instance, (Prats et al. 2012) introduced UWSim, a software tool designed for visualizing and simulating underwater robotic missions. UWSim features a camera system that renders images as seen by underwater vehicles, although it does not account for water-related effects. (Manhães et al. 2016) expanded the capabilities of the open-source robotics simulator Gazebo to include underwater scenarios, creating the UUV

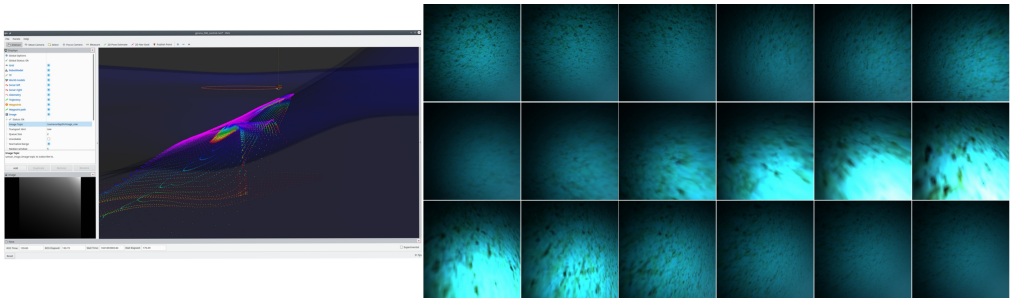


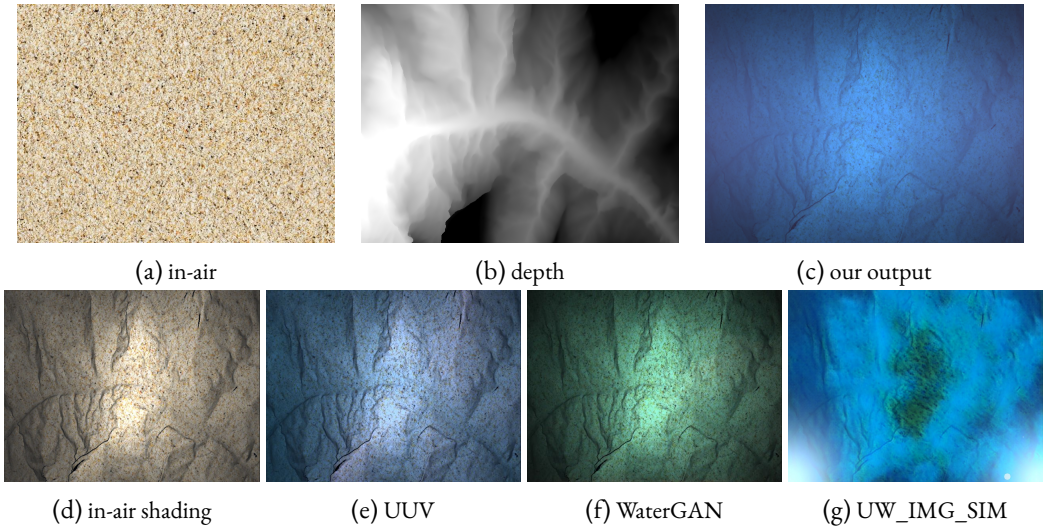
Figure 3.14: Left: camera path overview in simulator. Right: Rendered image sequence. Due to the physically correct model, already in the simulation we can see that some images will be overexposed with the settings chosen. Consequently, the exposure control algorithm of the robot can be adapted already after simulation without wasting precious mission time at sea.

Simulator. This simulator employs RGB-D sensor plugins to generate depth and color images, which are then transformed into underwater scenes using a simplified AF Model (Equation 3.6). Another underwater renderer, developed by (Álvarez-Tuñón et al. 2019), utilizes trained convolutional neural networks to perform style transfer on images generated from (Manhães et al. 2016). It additionally introduces forward scattering and haze effects, although it lacks a solid physical interpretation. (Allais et al. 2011) integrated the Ocean-Atmosphere Radiative Transfer (OSOA) model into their SOFI simulator. They developed lookup tables to construct the backscatter component. However, the OSOA model primarily describes sunlight transformation at the ocean--atmosphere interface, limiting its suitability to shallow water scenarios.

Gazebo, an open-source robotics simulator, stands out by offering a choice of four distinct physics engines for simulating robot mechanics and dynamics. It serves as a versatile platform for hosting various sensor plugins. Building upon Gazebo's foundation, the UUV Simulator (Manhães et al. 2016) extends its capabilities to underwater scenarios. This comprehensive simulator takes into account hydrodynamic and hydrostatic forces and moments, enabling the faithful simulation of vehicle dynamics in underwater environments. Furthermore, the UUV Simulator offers a suite of sensor plugins commonly deployed on UUVs, including sensors like Inertial Measurement Units (IMUs), magnetometers, sonars, multi-beam echo sounders, and camera modules. The deep sea image rendering pipeline introduced in this study can be seamlessly integrated into the UUV Simulator's camera plugin, initially designed to provide in-air and depth images. By leveraging these inputs, the deep sea image simulator transforms them into realistic deep sea-like images. It achieves interactive rendering speeds for $800 \times 800$ image sizes using OpenMP, all without the need for GPU acceleration, running smoothly on a 16-core CPU consumer hardware. The workspace interface and sample rendering results are shown in Figure 3.14.

### Deep Sea Image Simulator Evaluation

The quality of the deep sea image simulator was evaluated by comparing it with three state-of-the-art methods, all of which use in-air and depth (RGB-D) images as input to synthesize underwater images: UUV Simulator (Manhães et al. 2016), WaterGAN (Li et al. 2017b) and UW_IMG_SIM



(a) in-air (b) depth (c) our output

(d) in-air shading (e) UUV (f) WaterGAN (g) UW_IMG_SIM

Figure 3.15: Outputs of different underwater image simulators for the same scene.

(Álvarez-Tuñón et al. 2019). Due to the image size limitation in WaterGAN, all evaluated images were simulated at a fixed size of 640×480, although the proposed method does not have this limitation.

To generate realistic deep sea images resembling those shown in Figure 3.1, the camera-light setup was initialized as follows: two artificial spotlights placed 1 meter away from the camera on both the left and right sides, both tilted at 45° toward the image center. The real image used for comparison was taken in the Niua region (Tonga) in the South Pacific Ocean. According to the global distribution map of Jerlov water types from (Johnson 2012), the water in this region falls under type IB, with corresponding attenuation parameters of (0.37, 0.044, 0.035)[$m^{-1}$] for RGB channels. The simulation comparisons are presented in Figure 3.15. An in-air virtual scene featuring a sand texture was created as the object to be converted into a deep sea scenario. The corresponding underwater images were then simulated using different methods. Only the proposed method accounts for the impact of lighting geometry configuration. To ensure a fair comparison, in-air shading was initially added to the texture image for other simulators, as this in-air shading effect for a specific lighting configuration is not available in the other renderers.

The results shown in Figure 3.15 reveal key differences among the evaluated methods. The UUV Simulator can only render the attenuation effect based on the AF model, neglecting the impact of the light sources. Consequently, the backscatter pattern caused by lighting is entirely absent from their images. Furthermore, their attenuation effect only considers the path from scene points to the camera, leading to colors that do not conform to the deep sea reference image. Similar issues arise in the results from WaterGAN. Due to a lack of deep sea images with depth maps and ground truth in-air images, the GAN was trained using parameters provided in the official repository[1] on the *Port Royal, Jamaica* underwater dataset[2]. As a result, the color and backscatter

[1] https://github.com/kskin/WaterGAN

[2] https://github.com/kskin/data

pattern of the light source in WaterGAN's results are highly influenced by the training data, which does not align with the setup used for this evaluation. UW_IMG_SIM manages to present the backscatter pattern of the light source. However, this effect is simply added as bright spots into the image without a clear physical interpretation. Additionally, their direct signal component has no dependency on the light source, which is also unrealistic. In contrast, the proposed approach excels in capturing all the discussed effects present in real images. It not only renders colors much closer to real images but also simulates attenuated shading on the topography and backscatter caused by the artificial light sources. These aspects are missing in other fast approaches, showcasing the superiority of the proposed method in generating realistic deep sea images at interactive frame rate.

### 3.3.4 Deep Sea Image Simulation from Real Scenes

As is well-known, the intricate interplay of complex lighting and water effects poses a significant challenge when adapting current computer vision methods for deep-sea applications. Developing effective algorithms for such scenarios necessitates access to datasets with GT information to assess these methods accurately. However, capturing a pristine deep-sea scene without the interference of water or artificial lighting effects is practically impossible. This situation presents a substantial roadblock to progress in deep-sea vision research, where the availability of synthesized images with GT data could serve as a viable solution. Presently, most methods either rely on rendering from virtual 3D models or employ the Atmospheric Fog model to transform real-world scenes to resemble shallow-water environments illuminated by sunlight. Unfortunately, there is a noticeable scarcity of dedicated image datasets specifically tailored for evaluating deep-sea vision.

To the best of our knowledge, there is currently no existing literature that comprehensively addresses the challenges and methodologies involved in this unique endeavor. Based on the deep sea simulator, this thesis introduces a comprehensive pipeline for synthesizing deep-sea images from existing real-world RGB-D benchmarks, and exemplarily generates the deep sea twin datasets for the well known Middlebury stereo benchmarks. The simulated deep sea images generated from the real world in-air vision benchmarks can serve two purposes: firstly, to test underwater stereo matching methods, and secondly, to train and evaluate algorithms designed for underwater image processing.

#### Existing Vision Benchmarks

On land, the development and validation of vision methods have greatly benefited from the availability of good benchmarks and test data. Datasets with GT have played a pivotal role in continuously evaluating algorithm performance and refining results. Notable in-door scene 3D vision benchmarks include the Middlebury dataset, which encompasses Stereo (Scharstein et al. 2014), Multi-View Stereo(MVS) (Seitz et al. 2006), and Optical Flow (Baker et al. 2011) datasets. It provides the official web page for evaluating the submitted results, which has been widely used in 3D vision and photogrammetry research. The ETH3D dataset offers MVS (Schops et al. 2017) and SLAM benchmarks (Schops et al. 2019), while others like ISPRS/EuroSDR (Nex et al. 2015) and Hessigheim 3D (H3D) (Kölle et al. 2021) provide dense matching benchmarks for airborne photogrammetry. The KITTI (Geiger et al. 2012) dataset is prominent in autonomous driving, providing sensor measurements, stereo images, and GT trajectories for Visual Odometry and SLAM

evaluation. Additional datasets such as the Málaga Urban dataset (Blanco-Claraco et al. 2014) which contributes stereo images and light detection and ranging (LiDAR) measurements in urban scenario for SLAM in autonomous driving, the EuRoC micro aerial vehicle dataset (Burri et al. 2016) which provides the images sets with GT poses and a detailed 3D scan of the in-door environment for visual-inertial SLAM, and the TUM RGB-D SLAM dataset (Sturm et al. 2012) which captured RGB-D images through Microsoft Kinect, with given GT trajectories.

In contrast, there is a scarcity of published underwater vision datasets available for evaluation due to the complexities of conducting well-controlled experiments in underwater environments. However, some efforts have been made to address this gap in the field of underwater robotics and image processing. For instance, datasets like the one proposed by (Mallios et al. 2017) involve imagery collected by an AUV in underwater caves, another dataset AQUALOC (Ferrera et al. 2019), was captured by a ROV in real ocean conditions for Visual SLAM research. In the domain of underwater image processing, datasets like the underwater image enhancement benchmark (UIEB) (Li et al. 2019) have been assembled using images sourced from the internet. However, the enhanced reference images are manually selected by human inspection among 12 enhanced results. (Akkaynak et al. 2019; Berman et al. 2020) utilized an underwater stereo camera system and captured in total 57 stereo pairs for underwater restoration evaluation, the reference distances are computed via SfM. Since water blocks the use of GPS underwater, the real world robotic vision datasets share the same problem: that the GT trajectories and distances are not precise enough for evaluating underwater Visual SLAM. Also, enhanced images used as GT reference images are still not equal to the medium-free images. As the demand of high-accuracy underwater vision is increasing, the real-world underwater evaluation datasets are by far not sufficient, where synthesized datasets seem to offer a solution.

Synthesized images are widely applied in learning based underwater research, with many approaches relying on the AF model to convert RGB-D images into underwater scenarios as the training data (Li et al. 2020; Li et al. 2017b; Ueda et al. 2019). However, this model cannot accurately simulate complex artificial lighting effects in deep-sea scenarios, for which the J-M model is better suited. Efforts have been made to adapt and extend the J-M model for deep-sea image synthesis (Sedlazeck et al. 2011; Song et al. 2021a). Additionally, ray-tracing techniques, such as volumetric rendering, have been employed to synthesize light transport in underwater media (Bitterli et al. 2017; Crane et al. 2007; Novák et al. 2018). As an example, (Zwilgmeyer et al. 2021) applied Monte-Carlo path-tracing rendering from the rendering engine Blender (Blender Online Community 2021) to generate the underwater dataset VAROS.

#### Scene Preparation and Pre-processing

Because the GT disparity maps obtained from real-world scenes are inherently imperfect, the raw disparity data undergo a series of essential pre-processing steps to ensure compatibility with the deep sea image rendering procedure. Firstly, the disparities must be converted from pixel unit into real-world depth values to facilitate physical rendering. Secondly, the converted depth map necessitates additional refinement to address incomplete data regions. Thirdly, initial normals are estimated from the refined depth map, accompanied by the creation of a mask to identify areas of depth discontinuity where normals may be uncertain. Lastly, the masked normals undergo

smoothing through the application of a median filter. The complete pre-processing pipeline is illustrated in Figure 3.16.

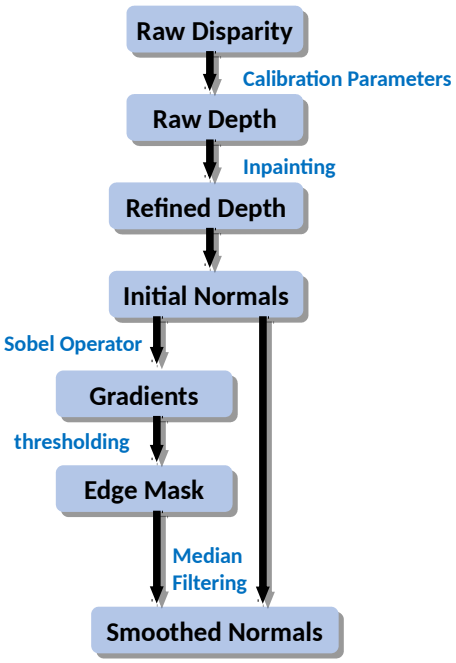



Figure 3.16: Workflow for pre-preprocessing real world depth data.

Stereo benchmarks provide disparity maps that serve as a means to directly evaluate dense image matching performance. Disparity values represent the pixel coordinate difference between corresponding points in a stereo image pair. However, for physical model-based simulation, it is essential to translate these disparity values into real-world depth information. The Middlebury dataset provides guidelines for this conversion, mapping the disparity value $disp$ [pixels] to the real depth $depth$ [meters] using camera calibration parameters:

$$depth = \frac{l_{base} \cdot f}{1000 \cdot (disp + d_{\text{offs}})}, \tag{3.13}$$

where $l_{base}$ represents the camera baseline [mm], $f$ is the pinhole's focal length [pixels] and $d_{\text{offs}}$ is the horizontal difference of the principal points.

Real-world raw depth maps often exhibit empty values where depth information is absent (See Figure 3.17 left). These maps are typically generated using external high-resolution devices, such as structured light systems (Scharstein et al. 2003). In these scenarios, the offset between the camera and the infrared emitter can cast stereo shadows, and specular surfaces can result in missing data. These incomplete depth maps cannot be directly utilized for deep sea rendering since the

rendering process relies on distance information for each pixel. Otherwise, significant artifacts would be introduced in areas where valid depth information is missing. To mitigate such issues, the incomplete depth maps need to be filled. It is important to note that these filled values are solely used for generating the appearance of underwater images and do not serve as GT data for evaluation purposes. Various methods have been proposed to address the challenge of filling in missing depth values. This thesis employs an inpainting technique based on the Navier-Stokes equations, as presented in (Bertalmio et al. 2001), to refine the depth maps.

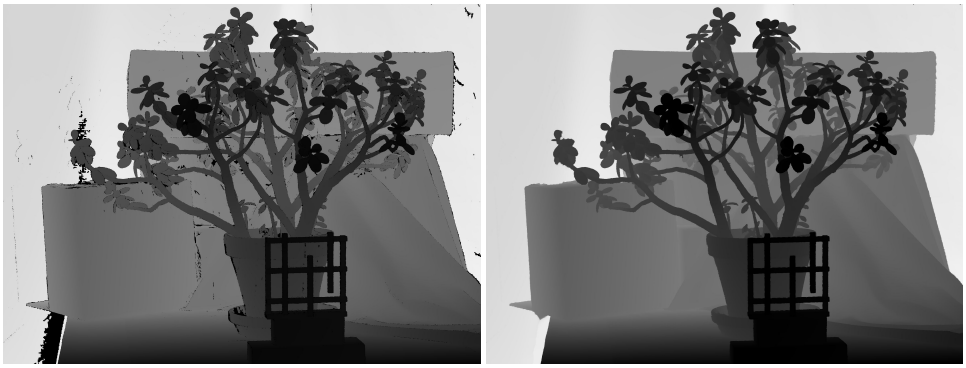


Figure 3.17: Left: The original raw depth map contains incomplete data. Right: Refined depth map using inpainting.

In addition to dealing with incomplete depth values, real-world benchmarks often feature complex scene geometry for evaluation purposes. This complexity introduces depth discontinuities that must be carefully managed when calculating surface normals from the depth map using local operators. Depth discontinuities have the potential to produce erroneous normal computation results. This can result in noticeable dark contours around objects in the image, especially under the standard Lambertian reflection model. To mitigate such artifacts, normals that face away from the camera are subjected to median filtering. The computation of normal maps from the imperfect depth map involves the following steps:

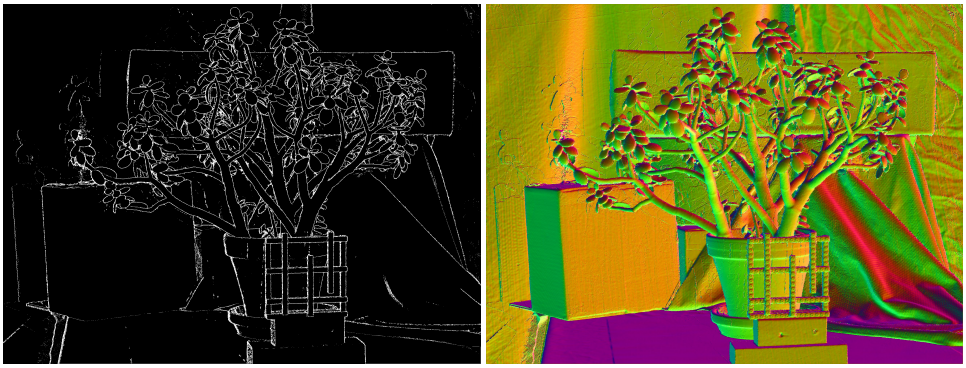


Figure 3.18: Left: Filtering mask for normal map edge smoothing. Right: Smoothed normal map.

1. **Compute initial normals**. The initial normal for each pixel is computed from the cross product of two vectors formed from the 3D difference of the neighboring pixels.

2. **Extract filtering mask**. Whenever a normal is facing away from the camera (z-component thresholding) and has a high local variation (thresholding of Sobel operator result) a potential depth discontinuity is detected and marked in a mask image (see Figure 3.18, left).

3. **Normal map smoothness**. A median filter is applied to the masked region in the initial normal map, which selects either the foreground or the background normal.

Moreover, given that the J-M model does not account for multi-scattering effects, this thesis integrates the Phong reflection model to enhance the results. This integration approximates multi-scattering effects through ambient light. The Phong reflection model (Phong 1998) is a widely-used model in computer graphics that describes reflected light as the sum of ambient, diffuse, and specular components. As underwater specular effects are rare, this thesis excludes the specular term from the original Phong reflection model and incorporates it into the J-M model. The direct signal $E_{direct,Phong}(\mathbf{x},\lambda)$ with Lambertian Phong reflection can be expressed as follows:

$$E_{direct,Phong}(\mathbf{x},\lambda) = I_0(\mathbf{x},\lambda)\cdot E(\theta,\lambda)\frac{e^{-c(\lambda)\cdot(d_1+d_2)}}{d_1^2}(\cos\tau + \mathrm{f}_{\mathrm{ambient}}). \tag{3.14}$$

where $\mathrm{f}_{\mathrm{ambient}}$ is the ambient factor in the range $(0,1)$, and the remaining parameters are defined in Equation 3.8.

To render the backscatter component, this section adopts the popular H-G phase function (see Figure 3.2) which is formulated as follows:

$$p_{HG}(\psi) = \frac{1}{4\pi}\frac{1-g^2}{(1+g^2-2g\cdot\cos\psi)^{3/2}}. \tag{3.15}$$

where $\psi$ = angle between the incident light ray and the camera viewing ray
$g$ = the asymmetry parameter in the range $(-1,1)$

### Parameter Settings

Determining accurate and physically meaningful rendering parameters is a complex endeavor. It necessitates expertise and specialized instruments to define these values correctly. Incorrectly configured parameters can drastically alter the visual output, as illustrated in Figure 3.19, resulting in synthetic images that deviate from realism.

In this rendering task, the camera intrinsics are obtained from the original Middlebury datasets. The attenuation parameters are referred to (Akkaynak et al. 2017) for Jerlov water type IB. For the remaining radiometric parameters, reference is made to (Song et al. 2021b) (also see Section 5.3.3), which provides an extensive elucidation of rendering settings based on real deep ocean images. A comprehensive list of parameter settings for rendering can be found in Table 3.1. Concerning the parameters introduced in the modified models detailed in Section 3.3.4, the default values of

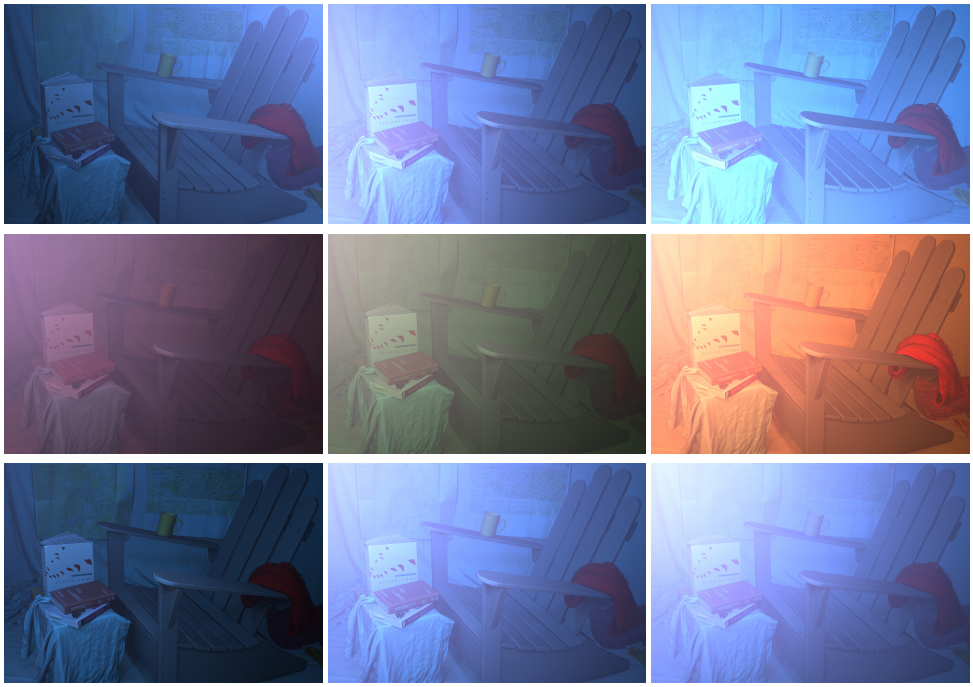


Figure 3.19: Simulated images vary significantly with different settings of lighting conditions (first row), attenuation parameters (second row), and $g$ in H-G phase function (third row).

ambient factors $f_{ambient}$ are set to 0.2 for all RGB channels, and the $g$ parameter of the H-G phase function is assigned a value of -0.4.

Two distinct lighting scenarios are defined for the synthesis of deep sea images. One scenario situates a spotlight 0.5 meters to the right of the origin within the local stereo camera coordinate system (Setup 1). The other scenario positions the light directly above the origin with a 0.5 meter separation (Setup 2). Both setups align the central axis of the spotlight in parallel with the camera's viewing direction.

### Rendering of the Dataset

The synthesized deep sea images using the Middlebury 2014 Stereo datasets is demonstrated in Figure 3.20. The rendering uses original in-air images paired with their respective raw disparity maps as inputs. Initially, the disparity maps are converted into depth maps based on the camera calibration parameters, and then undergo refinement to address any incompleteness in the depth data. The resulting deep sea stereo images, generated with both lighting Setup 1 and 2, are subsequently derived. For comparative purposes, images using the popular AF model are also rendered with the same input data.

Each synthetic stereo pair serves as a valuable resource for assessing corresponding underwater stereo matching methods. As the synthesized deep sea stereo images share the same GT disparities with the original data, they can be evaluated directly using the evaluation SDK provided by

| Parameter Name | Values |
|---|---|
| scale factor | 3.5 |
| scale factor bs | 1600.0 |
| volumetric max depth | 4.0 |
| num volumetric slabs | 10 |
| slab sampling method | EQUAL_DISTANCE |
| white balance | [2.498, 1.0, 1.448] |
| water attenuation RGB | [0.37, 0.044, 0.035] |
| light spectrum RGB | [0.25, 0.35, 0.4] |
| light RID type | 1 |
| auto iso | false |

Table 3.1: Parameter settings for the rendering.

the Middlebury official website. This SDK enables standardized evaluations against GT disparity maps. Additionally, the synthesized deep sea images, along with their original in-air counterparts, can be assembled into in-air/underwater pairs. These pairs are valuable for training and testing underwater image restoration techniques tailored to specific underwater image formation approximations.

### Evaluation of Rendering Results

As shown in Figure 3.20, scenes from the Middlebury 2014 datasets were synthesized to approximate certain aspects of deep sea scenarios using two different lighting setups. These synthesized images exhibit certain characteristics of deep sea images. Notably, the distinctive scattering pattern and the uneven illumination shading, induced by the artificial spotlight, allow observers to discern the lighting direction within the image. While the model employed in this synthesis is not exhaustive (accounting for only single scattering and featuring a limited phase function), it produces more visually realistic results compared to simulations using the AF model.

During the process of synthesizing deep sea images for the Middlebury datasets, few limitations of the proposed method have been observed. When the raw disparity data contains large areas with missing data, inpainting algorithms may struggle to accurately recover the complex scene. However, it is necessary to note that these areas lacking GT data do not contribute to the evaluation metrics. Additionally, it was observed that the original Middlebury ground truth disparities contain some artifacts, possibly due to areas within the shadow region of the structured light. This artifact is also reflected in the official GT samples' standard deviations. While these issues may not be immediately noticeable in the disparity maps, they become apparent when shading is applied with spotlights, as seen in Figure 3.21. This effect was retained in the synthetic results as it reflects the "true" geometric information of the GT depth, akin to "millimeter-depth scratches" on the object surface.

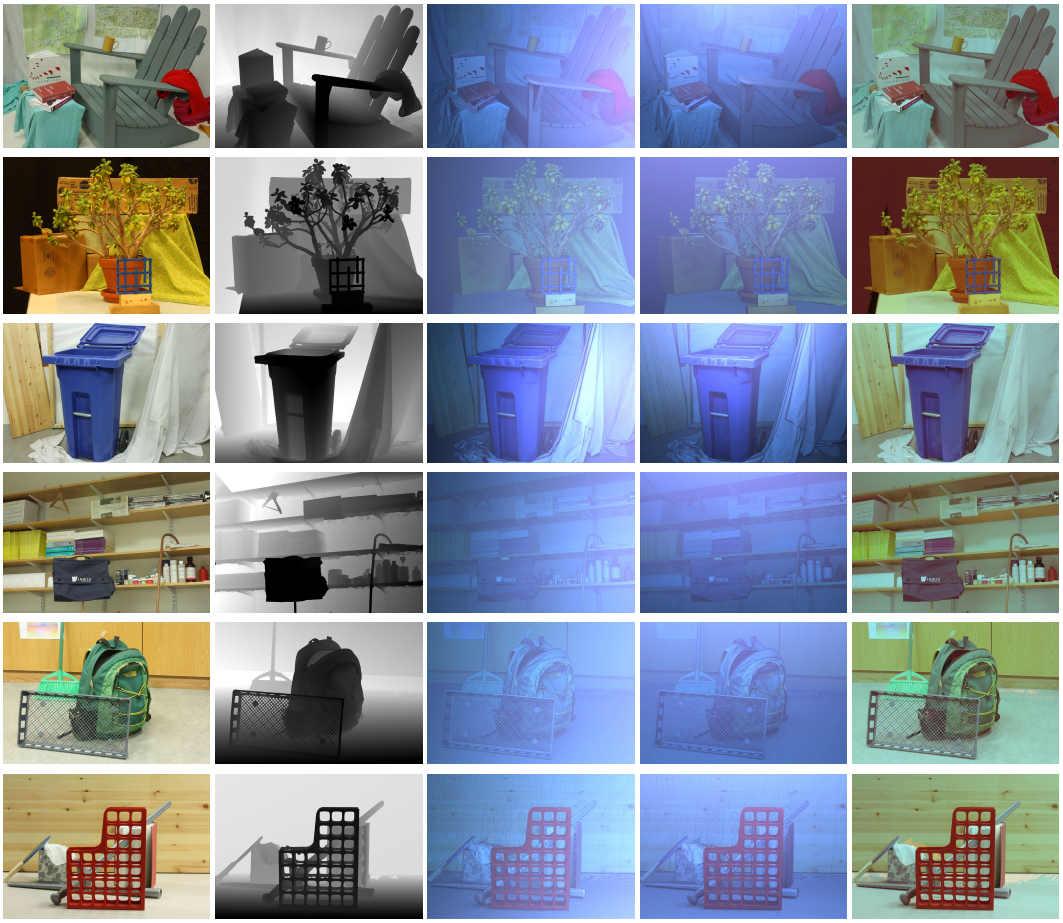


Figure 3.20: Examples of synthesized deep sea image twins for Middlebury 2014 Stereo datasets. From left to right: original in-air images from Middlebury dataset, Refined depth maps, Synthesized deep sea images with Setup 1 and 2. Synthesized underwater images using the atmospheric fog model (with the same attenuation parameters, background color was set to [110, 137, 212] for RGB). From top to bottom: left view image of Middlebury Adirondack, Jadeplant, Recycle, Shelves, Backpack, and Sword2 dataset.

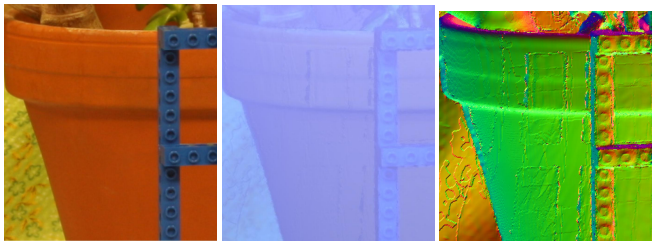


Figure 3.21: The slight noise present in the depth map becomes more noticeable when subjected to spotlight shading. From left to right: In-air color image, Deep sea synthesis, and Normal map.

# 4 Underwater Image Restoration

As discussed in the previous chapter, subsea images are affected by water and artificial lighting effects, requiring complex post-processing before creating mosaics (see Figure 4.1). For flat terrain, the application of image processing techniques to achieve high-quality texture becomes a higher priority than focusing on intricate geometric details in large-scale 3D visual mapping tasks. Such underwater image processing either utilizes a physical based image formation model or targets at qualitative criteria. The corresponding approaches are named restoration or enhancement, respectively. Restoration techniques should incorporate real-world distances and optical properties in order to recover the "true" color as perceived in air. Ideally, subsea visual mapping should deliver correct spectral information of the seafloor that enables later scientific usage (e.g. inferring material properties, identification of fauna etc.), which demands a "real" restoration during the image processing and not only an image that looks plausible. Unfortunately, the extra information that is required to achieve this is often not available, in particular in single image restoration methods. These methods thus often utilize prior knowledge or assumptions (e.g. gray world) to infer depth variations and combine the depth proxy with a physical model to restore images.

On the other hand, image enhancement is a very useful technique for many applications, but it is challenging to quantitatively evaluate. Often, the suggested method for qualitative evaluation is based on human visual inspection, assessing how many humans prefer the enhanced image (Mangeruga et al. 2018). Consequently, this thesis primarily focuses on image restoration techniques rather than image enhancement.

## 4.1 Methods for Underwater Image Restoration

The low level physics of light transport in water are well understood (Mobley 1994; Preisendorfer 1964) when looking at infinitesimally small volumes. Using statistical or physical models, the amount of light leaving a small volume into a particular direction can be predicted from the water parameters and the distribution of the incoming light over all directions. Considering the interactions of all the (infinitesimally) small volumes of an underwater scene at the same time in order to obtain a closed-form solution for image restoration is a challenging, if not impossible, endeavor. Consequently, several approximations to the low-level physical model have been proposed in the literature (see Section 3.2.2), including assuming a macroscopic atmosphere-like fog model for shallow water, a single scatter-model for artificial light sources and numerical/discretized simulation of the full problem using Monte-Carlo-based methods. This thesis outlines the underwater image restoration methods according to the image formation approximations they used.

Early works modeled underwater effects (mainly scattering) using a point spread function (PSF) (Mertens et al. 1977). Based on the PSF, a group of methods (Barros et al. 2018; Chen et al. 2019; Han et al. 2017; Hou et al. 2007; Liu et al. 2001) synthesize underwater images by generating in-

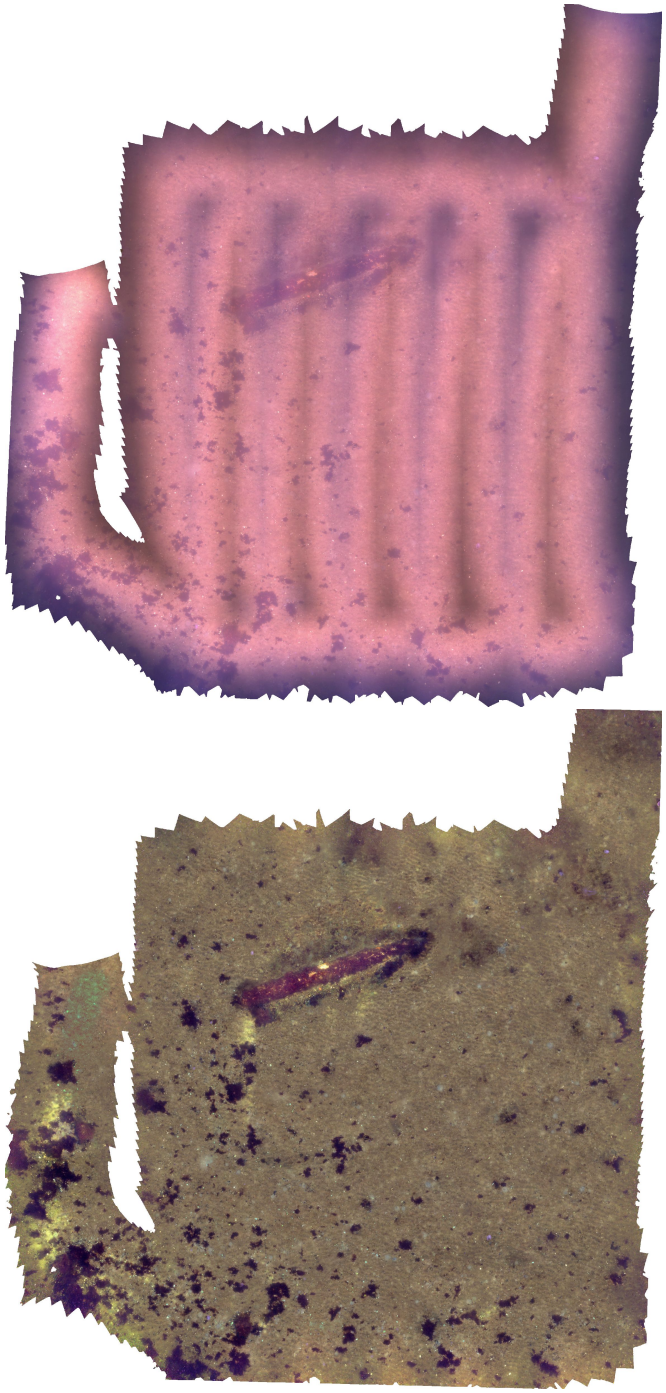


Figure 4.1: Around 20m×20m area of orthomosaic constructed from 106 images taken by an AUV with artificial illumination in Kiel Fjord, Germany. Top: orthomosaic directly generated from raw images (During capturing, the camera red channel gain was set to a higher number in order to acquire more contrast). Bottom: sample map reproduced from (Köser et al. 2021) to illustrate the importance of image restoration for large mosaics. Image restoration is vital for producing high quality subsea mosaics as it restores the correct spectrum information, improves the contrast and removes the uneven lighting, which will benefit the later biological, geochemical, geological and mapping applications.

air images of scenes, convolving them by the imaging system's response at the particular distance and applying the water effects of attenuation and backscatter. The underwater light transmission can then be simplified as a linear system and the restoration is basically a denoised deconvolution on images.

### 4.1.1 Atmospheric Fog Approximation based Methods

The most commonly adopted underwater image formation approximation for shallow water is derived from the AF model (Equation 3.6), which describes the underwater image as a linear combination of the attenuated signal and the backscatter. These methods assume the scene is illuminated by sunlight. But rather than explicitly modeling the sun, it is assumed that the object is illuminated uniformly and that the intensity received at the camera is a blend of attenuated object color and backscatter. The backscatter is often represented by a uniform background color and the attenuation between object and camera is induced by a transmittance map that depends on the distance to each scene point.

Many variations have been developed starting from this formulation for underwater applications. The atmospheric model was initially designed for in-air dehazing applications and it assumes that the scene is seen under the homogenous illumination, ignoring particular water properties. For non-homogeneous illumination cases, $\boldsymbol{T}(\mathbf{x})$ often multiplies with an extra illumination term which approximates the light propagation by Koschmieder's model (Koschmieder 1924). The basic atmospheric model applies the same coefficient in the transmission and the backscatter term which does not represent the underwater conditions well (Akkaynak et al. 2018; Song et al. 2021a). According to the definition from (Mobley 1994), the attenuation in the transmission is composed of absorption and total scattering. (Blasinski et al. 2014) simplified the backscatter term and extended the total attenuation by the summation of pure water and three other particle absorption coefficients. (Akkaynak et al. 2018) revised the model by applying different attenuation coefficients associated with the direct signal and the backscatter.

Most of these methods are proposed for single image restoration which is ill-posed: They require additional distance measurements (e.g. a depth map) or have to guess a proxy depth map derived from priors for the restoration. Generally, these methods can be concluded to three basic steps: scene distance (or equivalent representations) estimation, backscatter removal and transmission map estimation.

#### Scene Distance Estimation

Scene distance information, often represented as depth or distance maps, plays a crucial role in physical model-based restoration approaches. It serves as a fundamental requirement for tasks such as estimating the transmission image, correcting for attenuation effects, and facilitating the removal of backscatter in accordance with the image formation model. Depth information can be directly acquired using external devices e.g. a Lidar (He et al. 2004) or acoustic sensors (Kaeli et al. 2011), estimated from images pairs via stereo matching (Geiger et al. 2010; Shortis et al. 2009) or images with structured light (Bodenmann et al. 2017; Narasimhan et al. 2005a; Narasimhan et al. 2005b; Sarafraz et al. 2016) or images captured by light-field cameras (Tao et al. 2013; Wang et al. 2015). Depth information can also be estimated from redundant measurements: A group of methods (Hu et al. 2018; Schechner et al. 2004; Schechner et al. 2001; Treibitz et al. 2008) use

polarization filters and acquire multiple images with varying polarizer orientations to infer depth information from the estimated backscatter. (Nayar et al. 1999) estimates the structure of a static scene from multiple images with different illumination conditions. SfM has also been applied to generate the depth map (Akkaynak et al. 2019; Sedlazeck et al. 2009), the scale information can be obtained e.g. from a stereo system, from reference targets in the scene with known sizes or from navigation data. These methods are particularly suitable when multiple images with sufficient overlap and baseline are available, making them well-suited for visual mapping and image mosaicing tasks.

When direct depth information is unavailable, it can be inferred or approximated as often done in single underwater image restoration approaches. A popular idea is related to using the dark channel prior (DCP) (He et al. 2010), which was applied successfully for single image dehazing in white or bright grey fog or smoke. It assumes that in a haze-free image most of the local patches should contain at least one color channel with a very low intensity, but in a real image in fog, the more fog is in between the observer and the scene, the more the "dark" channels appear brighter. DCP inspired the development of single image dehazing methods and later this scene-depth derivation method has also been intensively applied in single underwater image enhancement (Chao et al. 2010; Chiang et al. 2011; Li et al. 2016a; Li et al. 2016b; Mathias et al. 2019; Yang et al. 2011; Zhao et al. 2015). Nevertheless, due to the severe attenuation of red light in underwater images, the standard DCP result does not fit for underwater scenarios and requires some modifications: (Carlevaris-Bianco et al. 2010) computes the intensity difference between the red channel and the maximum of the green and blue channels per-patch which terms maximum intensity prior (MIP). (Drews et al. 2013) proposes Underwater DCP (UDCP) which omits the red channel and apply DCP only in the green and blue channels. Later (Galdran et al. 2015) extends the UDCP with the inverted red channel, namely the Red Channel Prior (RCP). (Lu et al. 2015) discovered that the lowest pixel value in a turbid underwater images is not always the red channel but is occasionally the blue channel, it uses these two channels through a median operator to define the underwater median DCP (UMDCP). (Łuczyński et al. 2017a) inverts red and green channel to calculate the DCP by shifting the RGB coordinate system of underwater images from blue to white. (Peng et al. 2018) suggests a generalized DCP (GDCP) based on the depth-dependent color change, via calculating the difference between the ambient light and the raw intensity.

Besides DCP and its derivatives, some other priors are also proposed as a proxy to indicate depth variation in the image. (Peng et al. 2015) leverages the image blurriness which is increasing with distance and suggests the blurriness prior, later (Peng et al. 2017) combines it with the MIP and proposes the image blurring and light absorption (IBLA) prior. (Fattal 2014) discovers that pixels in a small image patch distribute along a straight line in RGB color space, known as the Color-Lines Prior (CLP). The Color Attenuation Prior (CAP) (Zhu et al. 2015) creates a linear model for depth estimation according to the brightness and the saturation of the image. (Berman et al. 2016) introduces a non-local prior, the Haze-Lines Prior (HLP), which suggests that pixels in a image can be clustered into few clusters. Pixels which belong to the same cluster in a hazy image are distributing along a line in RGB color space and all these lines pass through the background light. (Bui et al. 2017) proposes the Color Ellipsoid Prior (CEP) based on the observation that the vectors in the RGB color space of a small patch from hazy images are clustering in a ellipsoid. In the underwater scenario, image degradation is influenced not only by the object distance

but also by the wavelength dependent attenuation, which the standard HLP does not consider. (Wang et al. 2017b) claims that the pixels in the same color cluster will no longer form a straight line but a power function curve in RGB space, which is named Attenuation-Curve Prior (ACP). Afterwards, (Wang et al. 2017a) improves the ACP to the adaptive ACP (AACP), which is more general for different kinds of imaging environments.

Finally, learning based depth estimation approaches also been intensively studied on land (Eigen et al. 2014; Godard et al. 2019; Li et al. 2018b; Pillai et al. 2019) and have later been transferred to the underwater field (Gupta et al. 2019; Zhou et al. 2021b). However, similar to other prior based methods, learning-based approaches provide plausible depth information, but not necessarily physically meaningful and depends on the training data, lacking a direct correspondence to real-world distances.

#### Backscatter Removal

As an additive effect, backscatter introduces a loss of contrast or a foggy appearance that increases with distance. The total backscatter that the camera perceives accumulates all the scattered light along a viewing direction through the medium between the camera and the object. Subtracting the known backscatter component, if available, can effectively enhance image contrast. The challenge of backscatter removal has been studied extensively in image de-hazing, and many current underwater methods draw from these techniques. Physical model based de-hazing mechanisms require the knowledge of the scene depth, therefore de-hazing is highly correlated to the depth estimation and, vice versa, scene depth can be achieved as a by-product once de-hazing is solved.

This thesis classifies image de-hazing solutions into four main categories: Hardware-based, multiple-image based, prior-based approaches and learning-based.

(1) Hardware-based approaches use additional devices for image acquisition, for instance, directly blocking the backscattered signal through range gated imaging (Li et al. 2009; Tan et al. 2005; Tan et al. 2006), taking at least two static scene images with different orientations of a polarization filter in front of the camera (Schechner et al. 2007; Schechner et al. 2004; Schechner et al. 2005; Schechner et al. 2001; Schechner et al. 2003; Shwartz et al. 2006) or the light source (Dubreuil et al. 2013; Hu et al. 2018; Huang et al. 2016; Treibitz et al. 2006; Treibitz et al. 2008), capture images by a light field camera system (Skinner et al. 2017) or a stereo imaging system (Roser et al. 2014).

(2) Multiple-image approaches have first been proposed for in-air applications which take multiple images under varying visibility conditions and scene depth, and backscatter is estimated simultaneously during the optimization (Liu et al. 2018a; Narasimhan et al. 2002; Narasimhan et al. 2003a; Tarel et al. 2009), similar underwater approaches are also introduced in Section 4.1.3. These methods are developed for webcam like stationary settings. They not only require a static camera, but also demand significant changes between different conditions. When illumination configurations are relatively fixed, the non object image contains the complete backscatter information. (Fujimura et al. 2018; Tsiotsios et al. 2014) assume that images share the same backscatter component and subtract the non-object image from the underwater images to remove the backscatter. In shallow water, this solution is difficult to apply since the amount of scatter observed depends on the camera orientation relative to the sun as well as the water depth through which the sunlight has passed. In deep sea mapping scenarios, many UVs are equipped with fixed

artificial lighting systems and often operate at a relatively high altitude. As a result, the backscatter pattern remains stable across images in such scenarios. Additionally, it can take hours for UVs to dive down to the seafloor, during this period of time, large amount of pure water images with only backscatterred lighting patterns are acquired, which is ideal for effective backscatter removal (Bodenmann et al. 2017; Köser et al. 2021).

(3) Prior-based approaches are commonly used for single image approaches in shallow water with sunlight. When the scene geometry and distance are precisely known (Hautière et al. 2007; Kopf et al. 2008; Narasimhan et al. 2003b), backscatter can be directly fitted by an analytical model, or separated from a raw image by Independent Component Analysis (ICA) (Fattal 2008). In situations where scene geometry is not measured, prior knowledge can be used to approximate the depth map. According to the AF model in Equation 3.6, the backscatter component of an image can be expressed as $(1 - \boldsymbol{T}(\mathbf{x})) \cdot \boldsymbol{B}_{\infty}$. Consequently, the background light (BL) $\boldsymbol{B}_{\infty}$, which is also named background color, veiling light, ambient light or water color in the literature, is needed for computing the backscatter component. Typically, the value of the pixel that does not observe an object (with maximum depth) is picked as the BL (Kratz et al. 2009). Most of the priors were initially proposed for in-air de-hazing, such as DCP, CLP and HLP (see Section 4.1.1), they often take the uniform BL assumption over the entire field of view. DCP based in-air de-hazing approaches select the brightest pixel (in the image or dark channel) from a far scene as the BL (He et al. 2010; Tan 2008). However, this approach can yield erroneous results in the presence of bright objects in the scene. Several adaptations were developed for more accurate BL selection, such as using hierarchical quadtree ranking (Emberton et al. 2015; Kim et al. 2013; Park et al. 2014a; Peng et al. 2017; Wu et al. 2017), patch-based selection (Chiang et al. 2011; Serikawa et al. 2014), estimated from different priors or using extended models (Akkaynak et al. 2019; Carlevaris-Bianco et al. 2010; Henke et al. 2013) and additional selection according to some other rules (Ancuti et al. 2010; Li et al. 2017a; Wang et al. 2014; Zhao et al. 2015). Besides that, the BL can also be detected from the smoothest spot on the background for in-air de-hazing (Berman et al. 2016; Fattal 2014) and underwater backscatter removal (Berman et al. 2020; Berman et al. 2017; Li et al. 2018a; Lu et al. 2015; Peng et al. 2017; Peng et al. 2015; Wang et al. 2017a). It's important to note that assuming a single, uniform BL value may not hold in deep-sea scenarios, where backscatter depends on lighting configurations (Song et al. 2021a) and varies significantly with image position. Using a local estimator to provide a more accurate backscatter map is desired for precise artificial lighting backscatter removal (Ancuti et al. 2016; Li et al. 2018a; Tarel et al. 2009; Treibitz et al. 2008; Yang et al. 2019).

(4) Learning-based image dehazing has become very popular in recent years, with various approaches such as (Cai et al. 2016; Fu et al. 2017; Liu et al. 2019; Liu et al. 2018b; Ren et al. 2018; Zhang et al. 2017a). However, these methods often face a common challenge where the quality of dehazing heavily relies on the training data, making it challenging to predict how well they will generalize to different scenes.

Backscatter is actually a macroscopic effect that results from the volume scattering function, or the phase function, of the medium (Mobley 1994). These functions characterize in which directions an incoming photon is scattered when it interacts with the medium. In ocean water, this function has a peak in the backwards direction, therefore backscatter is an important effect. But photons are also redirected into other directions. In particular also small optical density variations (due to temperature, pressure or salinity fluctuations) of the medium lead to tiny direction

changes of photons. On a macroscopic level, all these effects are summarized as forward scattering, leading to distance-dependent unsharpness of the image, since photons deviate slightly from the direct line of sight. In simulation, forward scattering is often modeled by analytical filtering (Fujimura et al. 2018; Murez et al. 2015; Negahdaripour et al. 2002), that incorporates the underwater optical properties and convolves the image with the appropriate blur kernel. When removing forward scattering, PSF (or its frequency domain form modulation transfer function (MTF)) is often estimated (Barros et al. 2018; Chen et al. 2019; Cheng et al. 2015; Han et al. 2017; Hou et al. 2007; Liu et al. 2001), and one tries to reverse the effects by deconvolution. Other filters such as joint trilateral filter (JTF) (Serikawa et al. 2014; Xiao et al. 2012), self-tuning filter (Trucco et al. 2006), trigonometric bilateral filter (Lu et al. 2013) and Wiener filter (Wang et al. 2011) have also been used to describe the forward scattering effect. However, these methods are essentially spatially varying image sharpening operators that can introduce artifacts. Hence, many image restoration methods simply ignore forward scattering.

#### Transmission Estimation

From Equation 3.6, after removing the additive backscatter from the image, the transmission term $T(\mathbf{x})$ is estimated to restore the scene radiance from the direct signal. Similar to the Retinex model for artificial lighting compensation introduced in Section 4.1.2, the direct signal in underwater image formation is represented by the product of the transmission and the object reflectance. Transmission is reciprocal to the attenuation (Mobley 1994; Preisendorfer 1964), which has to be integrated along the line of sight, leading to an exponential expression based on the Beer-Lambert law and depends to the scene depth and water attenuation coefficient. Therefore, transmission is closely related to the scene distance. Once the attenuation coefficient is known, the transmission can be computed to recover the scene radiance (Akkaynak et al. 2019; Schechner et al. 2004).

The attenuation coefficients can either be directly measured by optical instruments like transmissiometers (Bongiorno et al. 2013), or be estimated from images (Akkaynak et al. 2019; Schechner et al. 2004). Jerlov's classification of global ocean waters (Jerlov 1968) provides a foundation for measuring their attenuation properties. Based on his work, the attenuation parameters can be directly inferred for specific water types (Akkaynak et al. 2017; Solonenko et al. 2015). However, once taken transmissiometer or spectrometer measurements might not be universally applicable to all captured images. Even within the same water type, attenuation can vary with factors such as season, depth, and wavelength. Additionally, the color of the image also depends on the spectral sensitivity of the camera, which is often not known. In such cases, attenuation coefficients can be estimated from in-situ images by photographing a reference target with a known spectrum at different known distances (Blasinski et al. 2014; Winters et al. 2009).

If neither scene distances, nor the reference target are available, an approximate scene layout can be derived from priors to estimate the transmission. For example transmission estimation make use of DCP (Chao et al. 2010; Chiang et al. 2011; Serikawa et al. 2014; Yang et al. 2011; Zhao et al. 2015), MIP (Carlevaris-Bianco et al. 2010; Li et al. 2016a), UDCP (Drews et al. 2013; Emberton et al. 2015; Lu et al. 2015), RCP (Wen et al. 2013), CLP (Zhou et al. 2018), HLP (Berman et al. 2020; Berman et al. 2017) and ACP (Wang et al. 2017a; Wang et al. 2017b). The Red channel is the most degraded channel in an underwater image, thus it has also been used to estimate the transmission map (Li et al. 2016a).

Per-pixel transmittance estimation is sensitive to the image noise. In order to achieve a dense and accurate transmittance map, post refinement is often needed to improve the transmittance estimation quality. A popular refinement technique is guided image filtering (He et al. 2012), this edge-preserving smoothing operator has been widely applied in transmission map refinement (Berman et al. 2020; Drews et al. 2015; Li et al. 2016a; Wen et al. 2013; Zhou et al. 2021a). Other refinement techniques are e.g. median filter (Tarel et al. 2009), fuzzy segmentation (Bui et al. 2017), Markov random field (Fattal 2008; Fattal 2014; Tan 2008), weighted least squares (WLS) filter (Emberton et al. 2015) and image matting (Chiang et al. 2011; Drews et al. 2013).

### Exemplary Systems

This section gives a detailed survey on the prominent underwater image restoration pipelines. Their corresponding approaches for estimating depth, backscatter (including BL) and transmission (with refinement) are introduced and summarized in Table 4.1.

(Schechner et al. 2004) images the scene through a polarizer at different orientations, the backscatter component is derived from the extreme intensity measurements. Global parameter BL is estimated by measuring pixels corresponding to non object regions, which is later used to derive the transmission map. It is the pioneer work which utilizes the atmospheric model for underwater image restoration.

(Trucco et al. 2006) assumes uniform illumination and low-backscatter conditions, and considers only the forward scattering component. They present a self-tuning restoration filter based on a simplified J-M model. The Tenengrad criterion (average squared gradient magnitude) is measured as the optimization target to determine the filter parameters by a Nelder–Mead simplex search. Image restoration is performed by inverting the filter in frequency domain on the raw image.

(Hou et al. 2007) models image formation as the original signal convolved by the imaging system's response and extends the PSF by incorporating underwater effects. The actual image restoration is then implemented by a denoised deconvolution.

(Sedlazeck et al. 2009) first utilizes SfM and dense image matching to generate depth maps for color correction. The BL is defined from the background patch in the image. Based on the atmospheric model, the backscatter and transmission (one attenuation coefficient) are estimated from a set of known white objects seen from different distances.

(Chao et al. 2010) first introduces DCP from (He et al. 2010) to underwater image de-scattering. The pixels with highest intensity among the the brightest pixels in the dark channel is picked as the BL. The dark channel of the normalized image is used to estimate the transmission. It removes the scattering effect in the image but the absorption issue still remains unsolved.

Inspired by DCP, (Drews et al. 2013) proposed UDCP which considers the blue and green channels are underwater informative and ignores red channel. It provides a rough initial estimate of the medium transmission which is later refined by image matting. Similar to DCP, the BL is estimated by finding the brightest pixel in the underwater dark channel.

(Galdran et al. 2015) inverts the red channel and proposes the RCP for BL and transmission estimation. The BL is picked from the brightest 10% of pixels the one that has lowest red intensity. The transmission map is later refined by using the guided filter.

(Emberton et al. 2015) adopts a hierarchical rank-based estimator for backscatter removal. The method exams over three features in the image, UDCP, the standard deviation of each color chan-

nel and magnitude of the gradient, to estimate the BL. The transmission map is generated from the UDCP and refined with the WLS filter (Farbman et al. 2008).

(Ancuti et al. 2016) uses the DCP over both small and large patches to locally estimate the backscatter, later fuse them together with the Laplacian of the original image to improve the underwater image visibility. These three derived inputs are seamlessly blended via a multi-scale fusion approach, using saliency, contrast, and saturation metrics to weight each input.

(Peng et al. 2017) computes the blurriness prior according to their previous work (Peng et al. 2015). The BL is also determined from the candidates estimated from blurry regions. Afterwards, the scene depth is estimated based on light absorption and image blurriness and refined by image matting or guided filter. The transmission map then is calculated for scene radiance recovery.

(Wang et al. 2017a) omits the depth estimation and acquires relative transmission based on ACP. It first filters the smooth patches with low total variation (TV), then the homogeneous BL is located where the pixel has considerable differences in R-G and R-B channel; Pixels are classified into attenuation-curves in RGB space and turned into lines using logarithm, transmission of the red channel is estimated from each line, and refined by a WLS filter similar to (Berman et al. 2016). The attenuation factor is then estimated to compute B,G transmissions.

Inspired by the illumination estimation method from (Rahman et al. 2004), (Yang et al. 2019) decomposed the dark channel and extracted the transmission based on the Retinex model. The backscatter light is obtained locally by using Gaussian lowpass filtering of the observed image. Afterwards, a statistical colorless slant correction and contrast stretch is adopted to correct the color.

(Akkaynak et al. 2019) applies a revised image formation model (Akkaynak et al. 2018) which formalizes the direct signal and the backscatter components with distinct attenuation coefficients. It first generates the scene depth using SfM. Estimation of the backscatter (BL and backscatter attenuation coefficient) is inspired by DCP, but is based on the darkest RGB triplet and utilizes a known range map. The transmission (direct signal attenuation coefficient) is estimated using an illumination map obtained using local space average color as input.

(Bekerman et al. 2020) provides a method for robustly estimating attenuation ratios and BL directly from the image. The initial BL is searched in a textureless background area and is later fine-tuned through an iterative curve fitting minimization. In each iteration the attenuation ratios are calculated accordingly. In the end, the transmission is estimated based on the HLP from (Berman et al. 2016) and regularized by a constrained WLS for scene radiance restoration.

### 4.1.2 Artificial Lighting Pattern Compensation

Artificial light patterns play a significant role in affecting the global homogeneity of underwater mosaics. Therefore, compensating for these lighting patterns is crucial for the performance and results of subsequent visual mapping processes. When dealing with small brightness differences, especially in cases where cameras have a very narrow field of view and almost uniform illumination, techniques similar to those used for image vignetting correction in air, such as multi-band blending strategies (e.g., as seen in (Brown et al. 2007)), can be applied during image stitching. This helps to make the patterns less conspicuous. However, for wide-angle lenses, which are often used in deep sea mapping, achieving uniform illumination becomes more challenging and may even be impossible. Unfortunately, most of the restoration methods mentioned earlier in the

Table 4.1: Underwater image restoration methods with their processing details. Abbreviations of estimated parameter (Est.), Gaussian lowpass filtering (GLF), image matting (IM), guided filtering (GF), Gray World Hypothesis(GWH)

| Method | Depth | Backscatter [BL] | Transmission [Refinement] | Features |
|---|---|---|---|---|
| Schechner et al. (Schechner et al. 2004) | (by product) | polarization [Est.] | Est. [n.a.] | First used atmospheric model underwater. |
| Trucco et al. (Trucco et al. 2006) | Est. | n.a. | filtering [Nelder–Mead] | Self-tuning filtering, only forward scattering |
| Hou et al. (Hou et al. 2007) | n.a. | PSF [n.a.] | PSF [n.a.] | Modeled all effects in one convolution. |
| Sedlazeck et al. (Sedlazeck et al. 2009) | SfM | Est. [pure water patch] | Est. [known object color] | First used SfM / dense scene reconstruction. |
| Chao et al. (Chao et al. 2010) | n.a. | DCP [n.a.] | DCP [n.a.] | First used DCP underwater. |
| Drews et al. (Drews et al. 2013) | n.a. | UDCP [UDCP] | UDCP [IM] | Proposes UDCP. |
| Galdran et al. (Galdran et al. 2015) | n.a. | RCP [RCP] | RCP [GF] | Proposes RCP. |
| Emberton et al. (Emberton et al. 2015) | n.a. | UDCP [UDCP+clustering] | UDCP [WLS] | Hierarchical rank-based BL estimation. |
| Ancuti et al. (Ancuti et al. 2016) | n.a. | multi-scale DCP [DCP] | DCP [weighted fusion] | Multi-scale descattering. |
| Peng et al. (Peng et al. 2017) | blurriness | blurriness [blurriness] | Est. [IM or GF] | Propose blurriness prior. |
| Wang et al. (Wang et al. 2017a) | n.a. | ACP (TV+MIP) | ACP [WLS] | ACP based. |
| Yang et al. (Yang et al. 2019) | n.a. | DCP (GLF) | DCP [n.a.] | Gaussian filter for BL estimation. |
| Akkaynak et al. (Akkaynak et al. 2019) | SfM | DCP (Est.) | Est. [GWH] | Use revised model and real depth. |
| Bekerman et al. (Bekerman et al. 2020) | n.a. | Est. (Est.) | HLP [WLS] | Estimate parameters + HLP. |

literature barely address artificial lighting effects. Since the exact illumination conditions are often unknown or difficult to measure, this problem has typically been approached subjectively, relying on qualitative criteria. This thesis individually explores this issue in more detail here to emphasize the importance of considering lighting compensation in deep sea visual mapping.

Methods for addressing lighting dispersion issues in underwater image processing can be categorized into two main groups. The first group comprises histogram-based methods, which rely on analyzing pixel intensity distributions and applying techniques like histogram equalization and stretching to achieve a more uniform illumination appearance. The second group is based on the Retinex theory (Land 1977; Land et al. 1971), which assumes that an image results from the product of an illumination and a reflectance signal. In this approach, the illumination signal is modeled and and exploited to recover the reflectance image. The Retinex theory, which has been adopted to estimate the local illuminant (Beigpour et al. 2013; Bleier et al. 2011; Finlayson et al. 1995; Kimmel et al. 2003) in image processing has later also been utilized in underwater cases (Fu et al. 2014; Zhang et al. 2017b). The work by (Garcia et al. 2002) provides a comprehensive overview of addressing lighting dispersion issues in image processing and categorizes solutions into four strategies. This thesis adopts their definitions and summarize the relevant research into the following three categories:

(1) *Exploitation of the illumination-reflectance model*, it considers the image as a product of the illumination and reflectance, the illumination-reflectance model is estimated by a smooth function. The uneven lighting effect is then eliminated by removing the illumination pattern. Several methods have been proposed: (Pizarro et al. 2003) averages frames to estimate an illumination image in log space. (Arnaubec et al. 2015) employs a mean or median filter to extract the illumination pattern and describes this spot pattern as a third order polynomial. (Köser et al. 2021) robustly estimates all multiplicative effects including the light pattern, also using a sliding window median. (Bodenmann et al. 2017) also approximates the lighting and water effects as an multiplicative factor. It is estimated empirically from a series of images taken at different distances on known seafloor objects. (Borgetto et al. 2003) uses natural halo images to model the lighting pattern. (Johnson-Roberson et al. 2010) assumes a single unimodal Gaussian distribution to correct illumination variations and later (Johnson-Roberson et al. 2017) proposes a two-level clustering process to improve the performance. (Rzhanov et al. 2000) de-trends the illumination field through a polynomial spline adjustment.

(2) *Histogram equalization* is a method that adjusts the intensity histogram of an image to a desired shape, which effectively enhances image contrast by flattening its histogram. However, it tends to perform inadequately in scenarios with non-uniform illumination, such as deep sea images. To address this issue, Adaptive histogram equalization (AHE) (Pizer et al. 1987) was applied in (Eustice et al. 2000), to enhance the mosaicing images by equalizing the histogram in the local window through the entire image. In (Eustice et al. 2002), a variant of AHE called contrast limited adaptive histogram equalization (CLAHE) (Zuiderveld 1994) is utilized. This method applies histogram equalization within individual blocks of the image and employs interpolation techniques between neighboring blocks to eliminate boundary artifacts. Additionally, (Lu et al. 2013; Lu et al. 2015) expand the histogram in different color spaces based on pixel intensity redistribution.

(3) *Homomorphic filtering*: Due to the multiplicative nature of illumination effects, they translate into additive components when working within a logarithmic space. In this context, modeling

the illumination component becomes feasible through techniques like low-pass filtering or parametric surface fitting, particularly because the illumination-reflectance model maintains linearity (Bazeille et al. 2006; Guillemaud 1998; Singh et al. 1998; Singh et al. 2007).

In addition to the approaches mentioned above that specifically address artificial lighting compensation, several underwater image enhancement methods also take this issue into consideration (Chiang et al. 2011; Peng et al. 2017) during the depth or transmission estimation, or fused with several processing steps (e.g. Gamma correction (Cao et al. 2014), white balancing) to enhance the image contrast (Ancuti et al. 2017a; Ancuti et al. 2017b; Ancuti et al. 2016; Ancuti et al. 2012; Bazeille et al. 2006).

In the early stages of research in this field, the focus was primarily on monochromatic images. At that time, underwater image processing methods aimed to enhance image contrast (by removing scattering) and compensate for light patterns, particularly for mosaicing purposes. However, these early lighting pattern compensation approaches were not based on physical principles, and they were mainly utilized in image enhancement applications. Consequently, these approaches often produced mosaic patterns that exhibited quantitative properties closely tied to image content. As a result, any changes in relative geometry between the camera, light sources, and the scene could lead to abrupt patterns in the mosaic.

### 4.1.3 J-M Approximation based Methods

The J-M model, which considers the propagation of artificial light sources, is better suited for deep sea scenarios. It assumes single scattering and approximates forward scattering and backscatter using PSF and VSF, respectively. Based on the J-M model, if any one of the property among scene depth, water parameters and lighting configuration is known, the remaining unknown properties can be derived from variations in appearance between corresponding image pairs from multiple images. In most of the cases, the water properties are part of the unknown parameters to be estimated. These water properties typically consist of two groups of parameters: attenuation and VSF parameters, with the number of VSF parameters depending on the phase function model used. Some researchers assume that the proportion of scattered light has a uniform directional distribution, such that the corresponding VSF becomes constant (Bryson et al. 2016) and might be negligible during the restoration. Only a few works actually have attempted to also estimate the VSF parameters from images. Different phase function models have been used, such as (Narasimhan et al. 2005a; Narasimhan et al. 2005b; Tsiotsios et al. 2014) use the phase function model from (Chandrasekhar 2013), (Murez et al. 2015; Nakath et al. 2021; Narasimhan et al. 2006; Spier et al. 2017; Tian et al. 2017) utilize the HG phase function and (Pegoraro et al. 2010) models a general phase function model by using Legendre polynomial basis or Taylor series.

Similar to the depth cue estimation in hazy images, this group of methods requires multiple correspondences with variations to solve the final optimization. When capturing multiple images under different known lighting configurations, this problem becomes a typical underwater photometric stereo problem (Fujimura et al. 2018; Murez et al. 2015; Narasimhan et al. 2005a; Negahdaripour et al. 2002; Queiroz-Neto et al. 2004; Tian et al. 2017; Tsiotsios et al. 2014). (Spier et al. 2017) shows that the water properties can be derived even from empty scene backscatter images with a controlled light source movement. If the scene depth information is given, it becomes

a light source calibration problem using a known lambertian surface (Park et al. 2014b; Weber et al. 2001) where, however, additional water effects have to be considered.

Estimation of the unknown parameters requires the observations to be in a good configuration (e.g. significant differences). As directly solving the equations can be very complex or intractable, often iterative methods are employed that minimize some error function in a gradient descent manner. Those schemes need to start from good initial values, otherwise parameter estimation can be trapped in local minima or degenerate cases. Additional constraints with respect to the lighting configurations together with scene depth information can further strengthen the robustness of water parameters estimation (Bryson et al. 2016).

### 4.1.4 Monte Carlo based Methods

The J-M approximation only accounts for single scattering in the model, which represents a simplification of the underwater radiative transfer. (Mobley 1994) introduced Monte Carlo techniques to solve the underwater Radiative Transfer Equation (RTE) and discussed ray-tracing methods for simulating the propagation of light rays underwater. Powered by advances in GPU technology and physics-based simulation, modern graphic engines can efficiently synthesize complex underwater effects using ray-tracing (Zwilgmeyer et al. 2021). Recent approaches have even employed Monte Carlo-based differentiable ray-tracing to replace an explicit image formation model in image restoration. This involves characterizing the water with differentiable properties and optimizing the process (Nakath et al. 2021). Such an approach can implicitly handle multi-scattering, shadows, and various phase functions.

### 4.1.5 Learning based Methods

Many learning-based underwater image restoration methods have emerged over the last decade, including works like (Fabbri et al. 2018; Lu et al. 2021; Torres-Méndez et al. 2005; Yu et al. 2018). However, (Akkaynak et al. 2019; Bekerman et al. 2020) have addressed the shortcomings of these methods, such as their strong dependence on training data, and there is still large uncertainty in what scenarios they can reliably be applied e.g. when a robot is diving to a previously unseen ocean region and for other open applications. One of the main challenges is the lack of underwater image datasets with GT in terms of the in-air appearance, which is essential for training. In particular, it is very difficult to know how a particular seafloor spot in the deep sea would really look without water, which is however what would be naturally needed for training. Current learning based methods either use synthetic images or restoration results from other methods as the training data, which make their training problematic. Meanwhile, deep sea images' appearances strongly depend on the camera-lighting-scene configurations and water properties, which is even more challenging for learning methods to restore such images with general training sets. Therefore, this thesis did not further investigate learning based methods in it as they are currently not applicable for deep ocean mapping.

## 4.2 Challenges in Deep Sea Image Restoration

The preceding sections have provided an overview of image restoration techniques. These algorithms are designed to mitigate the degradation caused by scattering, attenuation effects, and artificial light cones. However, it is evident that there exist certain gaps and missing components when aligning the practices of deep ocean visual mapping with current image processing approaches:

1. The majority of underwater image restoration methods have their foundations in AF models or related derivatives. However, these models are only suitable for shallow water scenarios where homogeneous sunlight serves as the primary source of illumination. Single image restoration, although extensively studied, grapples with the inherent challenge of being an ill-posed problem, necessitating supplementary observations and often neglecting the need for mapping consistency—something crucial for practical applications. Approaches like the DCP offer a means to restore single images without the requirement for additional measurements, which has gained widespread adoption. Nonetheless, akin to enhancement methods, most single-view restoration techniques do not incorporate true distance information, potentially leading to consistency issues when applied to image sequences. Furthermore, the presence of artificial lighting can significantly impact the accuracy of prior estimations.

2. Removing artificial illumination patterns (also light cones) has the most significant impact on underwater mosaicing, but so far it did not draw much attention within the underwater image restoration community. Current lighting compensation methods either analyze quantitative properties in single images, which may perform inconsistently over image sequences, especially when the scene contents change significantly, and are not able to handle complex lighting conditions; or subtract some sort of "mean" pattern of an image sequence, which has strict assumptions on flatness and uniformity of the scene and the relative poses between the camera, the light sources and the scene have to be stable.

3. The J-M approximation-based approaches hold promise for addressing the challenges of deep sea image restoration with artificial illumination, as they consider the propagation of point light sources—a favorable characteristic for this particular problem. Since the scene depth estimation and image restoration is a chicken-egg dilemma, current methods necessitate multiple observations of the same 3D point, enabling them to estimate both the water properties and the scene depth. Many of these approaches have been primarily demonstrated in turbid media within controlled laboratory settings, where significant variations can be observed over short distances, ensuring robust parameter estimation. At the same time, the J-M methods model each light source individually, which becomes tricky for complex lighting conditions. Recent imaging platforms tend to utilize many LEDs in complex configurations, such that it becomes more difficult and impractical to execute calibration for each light source separately.

4. The J-M approximation only considers single scattering, which is a simplification for the complex underwater radiative transfer. The upcoming GPU-enabled Monte Carlo based ray tracing simulates the light propagation in the micro scale physics, and is able to solve

more challenging restoration problems with multi-scattering and shadows. However, it still has a similar problem as the J-M based approaches that restoration and reconstruction depend on one another, and multi light sources increase the computational complexity.

5. Learning based approaches can suffer from the consistency problem. The difficulties of acquiring ground truth for underwater (and more so: deep sea) images becomes the bottle neck of developing training based restoration approaches.

6. In situ calibration in the deep sea is still challenging. To our best knowledge, there is no real implementation yet for calibrating radiometric, light pattern and water properties in the deep ocean.

## 4.3 Advanced Underwater Image Restoration in Complex Illumination Conditions

In the process of restoring underwater images, it's essential to estimate the unknown parameters, such as water properties and lighting configurations. When a particular underwater image formation model or rendering pipeline is predefined, these parameters can be estimated through redundant observations. This can involve utilizing information from multi-view images, or different areas within a single image with varying depths. Ultimately, the image restoration process can be viewed as the inverse rendering procedure applied to the underwater images, leveraging the estimated parameters to recover the scene's appearance.

In the context of deep sea visual mapping, existing methods encounter challenges in achieving high-quality images devoid of water-induced distortions and complex artificial lighting effects, ultimately failing to produce images that closely resemble their appearance in air. To overcome these limitations, this thesis introduces a novel and versatile solution for underwater image restoration. This approach effectively addresses the drawbacks associated with current methods, including the limitations of the AF and J-M models, as well as models based on qualitative criteria. Notably, this novel method does not rely on additional knowledge of underwater environments, such as lighting conditions and water properties. It aims at restoring the true colors of underwater images while eliminating uneven lighting artifacts resulting from artificial light sources, capable of handling illumination conditions ranging from simple to complex. Furthermore, it maintains uniform brightness and true colors across image sequences, a critical aspect for subsequent 3D reconstruction and photomosaicing processes.

### 4.3.1 The General Image Formation Model

Section 3.2.1 classifies various underwater image formations into four fundamental types based on their illumination conditions. Different image formation models are employed to represent these distinct types if images, each of them has unique advantages and drawbacks.

As discussed in Section 4.2, the AF Model and its variations are widely adopted due to their simplicity. These models assume homogeneous scene illumination, where the total backscatter is determined by single BL, which is dependent on water properties. The key advantage of the AF Model lies in its minimal parameterization (involving no integrals) and the ability to estimate

these unknown parameters with information from a single image. However, a notable limitation is its inability to accurately describe complex total backscatter patterns with only one parameter, particularly in the case of Type III and IV images.

The J-M Model offers a more intricate representation of underwater image formation, addressing the manifold scattering pattern cased by artificial point light sources. It considers the integration of scattered light along the viewing ray from all light sources, while also considering attenuation across the entire transmission path—from the light sources to the object and back to the camera. Estimating water parameters using the J-M Model often necessitates multiple-view correspondences. Nevertheless, the J-M Model, along with Monte Carlo-based methods, demands detailed knowledge of each individual light source, restricting its applicability in scenarios with complex lighting configurations.

Upon analyzing the AF and J-M models, it becomes apparent that both can be expressed in a more generalized form. Assuming object shading has been effectively compensated, the AF and J-M models can both be summarized as a combination of a multiplicative term, representing the direct signal, and an additive term, representing the backscattered signal:

$$I = \alpha \cdot I_0 + \beta. \qquad (\alpha, \beta > 0) \tag{4.1}$$

In underwater images, pixel intensity for each channel $I$ is expressed as the product of the object albedo $I_0$ and the transmission factor $\alpha$, added by the backscatter component $\beta$. It is important to note that the intensity observation refers to the intensity after shading compensation. Assuming that the object surface is Lambertian, shading compensation can be performed by dividing the original pixel intensity by $\cos\tau$, where $\tau$ is the angle between object surface normal and incoming light. Assuming the light originates from the camera position, and the surface normal can be calculated from the corresponding depth map. Underwater image restoration can be considered as an inverse processing that aims to recover the object albedo from the underwater observations $I$. It is achieved by subtracting $\beta$ from the observed image and dividing the result by $\alpha$:

$$I_0 = \frac{I - \beta}{\alpha}. \tag{4.2}$$

Having established a unified representation for underwater image formation, the next step is to explore its application in handling complex illumination scenarios. As discussed in Section 3.3.2, the backscatter pattern often remains relatively stable within the viewing frustum in front of the camera, which is a common setting in the AUV deep sea mapping. To accelerate the rendering process, a 3D lookup table was employed to store pre-rendered backscatter patterns. It involves dividing the 3D view frustum of the camera into multiple slabs (McGlamery 1975), with each voxel in a slab storing pre-computed backscatter values for each RGB channel. Based on the 3D lookup table structure, the backscatter component can be pre-rendered, facilitating fast conversion from in-air images to underwater images. Conversely, when the lookup table is available, it can also be applied to eliminate water effects, enabling the transformation of underwater images into a representation akin to in-air scenes.

Building upon this concept, this thesis adopts a similar structure for parameter storage. Each voxel in this container holds two key factors: $\alpha$ (a multiplicative factor) and $\beta$ (an additive backscatter factor) for each color channel. This configuration forms a lookup table, as illustrated in Fig-

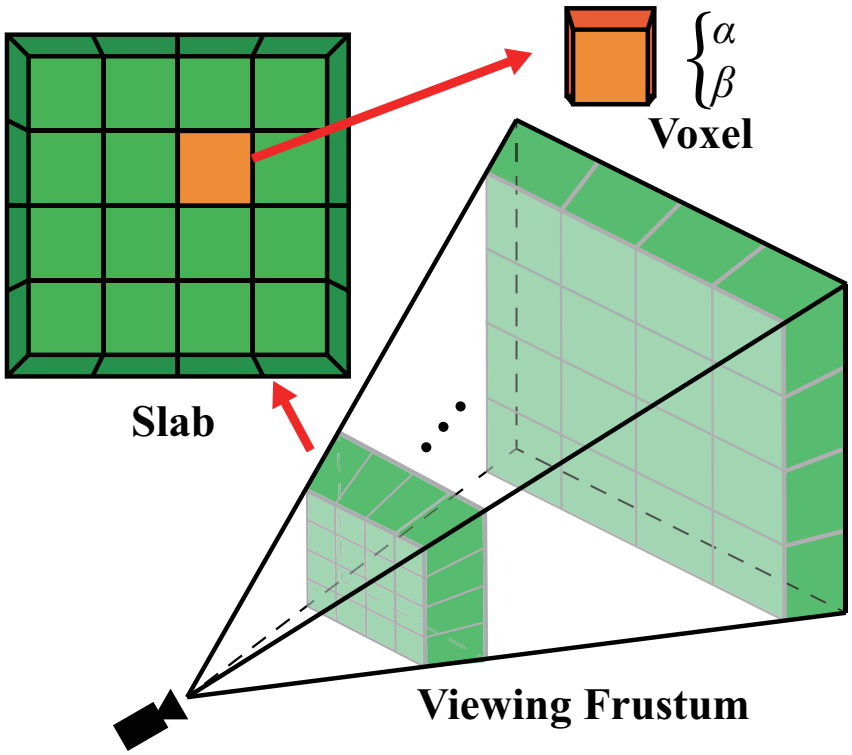



Figure 4.2: Proposed 3D lookup table structure. The camera viewing frustum is sliced into several slabs and each slab is constructed by a plane of voxels. Each voxel with in a slab stores two parameters: a multiplicative factor $\alpha$ and a additive factor $\beta$, for each color channel. These parameters represent the combined effect of lighting and water at that particular 3D position. Giving the stable lighting and water conditions during a single mission, either under homogeneous illumination in shallow water or co-moving artificial light source in deep water, the parameters in the lookup table are relatively fixed, enabling rapid batch restoration of entire image sequences.

ure 4.2. Since underwater image formation primarily concerns water and lighting effects within the camera's viewing frustum, this generalized underwater image formation representation, combined with the 3D lookup table structure, offers a means to model complex artificial lighting effects without requiring detailed information about each individual light source.

This novel underwater image formation model, though presented with a straightforward formulation and structure, proves highly effective in addressing various types of underwater images. It is not only suitable for underwater image applications, but can also be extended to in air cases such as in fog or with active illumination.

### 4.3.2 Observations and Constraints

Once the underwater image formation model is established, the subsequent step in image restoration involves the estimation of the lookup table parameters. Estimation of the parameters in the lookup table can be accomplished through a variety of constraints derived from underwater im-

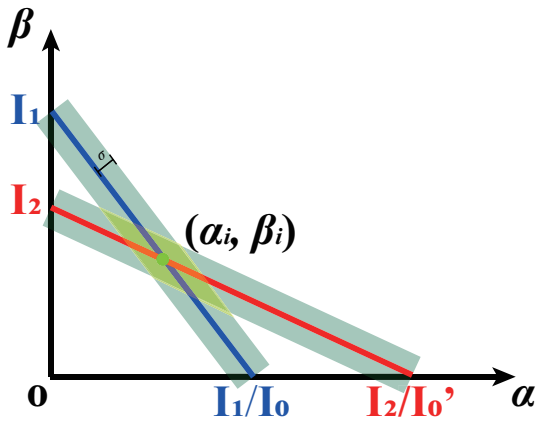



Figure 4.3: One observed color ($I$) with a known color ($I_0$) can only provide a constraint on $\alpha$ and $\beta$ along a line in the $\alpha$-$\beta$ plane. To obtain a unique solution for each voxel, at least two observations with different known colors are required. As shown in the figure, the blue line is the constraint from one observed underwater color $I_1$ at voxel $V_i$ with known color $I_0$, while the red line refers to the constraint from another underwater color observation $I_2$ at the same voxel with second known color $I_0'$. The intersection point of the two lines (in green) provides the unique solution $(\alpha_i, \beta_i)$ for voxel $V_i$. Due to the uncertainty $\sigma$ in the observations, each line is only constrained in the green interval and the ambiguity of the solution is defined by the intersection of the two constraint regions (in yellow). To minimize this ambiguity and reduce the uncertainty of the solution, slopes of two lines ($-I_0$ and $-I_0'$, respectively) should be perpendicular to each other in order to achieve minimum intersection of intervals. Therefore, two known colors with widely disparate values should be used for the observations.

ages. This thesis introduces several physical constraints that can be leveraged, including "Known Color Constraints", "Correspondence Constraints", "Smoothness Constraints" and "Pure Water Constraints". These constraints are grounded in real-world physics and provide effective means for accurately estimating the lookup table parameters.

### Known Color Constraints

When filming an object with known color (albedo), Equation 4.1 can be used directly to form the known color constraint, which becomes an equation of a simple line on the $\alpha$-$\beta$ plane. However, as shown in Figure 4.3, a single known color constrain is insufficient to solve for the two unknown parameters in each voxel. At least two observations ($I_1$ and $I_2$) with different known color objects ($I_0$ and $I_0'$, respectively) on the same voxel $V_i$ are required to obtain the unique solution for corresponding $\alpha_i$ and $\beta_i$ (see Equation 4.3). Moreover, due to the errors in measurement, each known color constraint provides an interval of solutions rather than a single line. To minimize the intersection of intervals and reduce the uncertainty of the solution, the two known colors are supposed be widely disparate.

$$\begin{cases} I_1 = \alpha_i \cdot I_0 + \beta_i. \\ I_2 = \alpha_i \cdot I_0' + \beta_i. \end{cases} \tag{4.3}$$

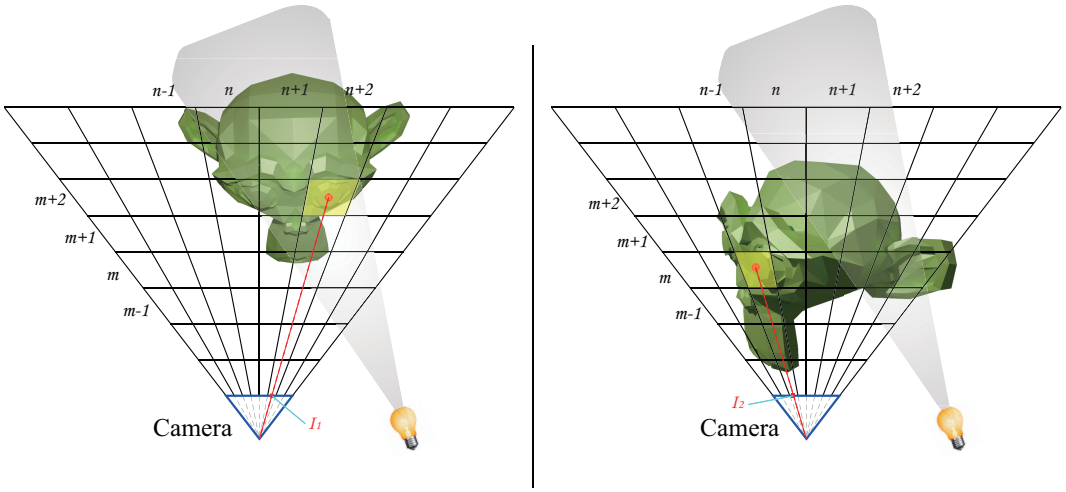



Figure 4.4: Correspondence constraint can be constructed from image correspondence in the scene. As it is shown, a point is filmed by two images with pixel intensities $I_1$ and $I_2$, correspond to voxels $V(n+2,m+2)$ and $V(n-1,m)$, respectively. Parameters for these two voxels are integrated to form one correspondence constraint: $\alpha_{(n-1,m)} \cdot I_1 - \alpha_{(n-1,m)}\beta_{(n+2,m+2)} - \alpha_{(n+2,m+2)} \cdot I_2 + \alpha_{(n+2,m+2)}\beta_{(n-1,m)} = 0$.

In principle, an ideal diffuse object that reflects all visible light wavelengths equally and a perfect black body that absorbs all incoming light will minimize the uncertainty of the solutions. In this case, the backscatter factor ($\beta$) in the lookup table can be directly measured by filming the black body in the medium at different distances. Once all $\beta$ values are fixed, $\alpha$ values can be computed directly by subtracting the corresponding $\beta$ from images of the ideal diffuse object ($\alpha = (I - \beta)/I_0$, where $I_0 = 1$).

### Correspondence Constraints

Similar to the feature matching problem in SfM, pixel color correspondents between images can be established in order to estimate the parameters in the 3D lookup table (see Figure 4.4). When the same object is filmed by two images w.r.t different voxels in the lookup table, two equations can be generated according to Equation 4.1:

$$\begin{cases} I_1 = \alpha_1 \cdot I_c + \beta_1. \\ I_2 = \alpha_2 \cdot I_c + \beta_2. \end{cases} \tag{4.4}$$

Where $I_1$ and $I_2$ are the two different observed color of the correspondents which share the same unknown object albedo $I_c$. This type of constraint is not sufficient to directly estimate the lookup table parameters, as each pair of image correspondences contains four unknowns. Equation 4.4 can be extended to include multiple observations of the same point in different images, but this does not help in solving the problem as more unknowns are added to the equation system.

At least four pairs of images observing four differently colored objects at the same position in the local camera coordinate system, it is possible to achieve a unique solution for that voxel. Unfor-

tunately, it is difficult to obtain such complex constraints in practice. Therefore, from Equation 4.4 this thesis constructs one constraint for each pair of correspondences, which can be derived as:

$$\alpha_2 \cdot I_1 - \alpha_2\beta_1 - \alpha_1 \cdot I_2 + \alpha_1\beta_2 = 0. \tag{4.5}$$

Extracting reliable color correspondences between images is a critical task. Traditional image matching is achieved by using key points (e.g. SIFT (Lowe 2004) and SURF (Bay et al. 2006) features), often based on gradient features or located around image corners or edges with significant changes in pixel intensities. These areas usually have unreliable and inaccurate color information due to the rapid changes in intensity. Color correspondents require to be extracted from homogeneous regions. This thesis utilizes super-pixel (Achanta et al. 2012) to segment the image into patches, where each patch exhibits a relatively homogeneous color. Specifically, the color information for each patch is extracted from its center, which is then used to estimate the lookup table parameters.

#### Smoothness Constraints

If not every voxel in the lookup table is constrained by enough observations from images, additional constraints are required to impose smoothness on the estimated parameters. The smoothness constraint can be expressed in a simple form as follows:

$$w_{s,\alpha} \cdot (\alpha(x,y,z) - \alpha(x \pm 1, y \pm 1, z \pm 1)) = 0. \tag{4.6}$$

Here the lookup table parameter $\alpha$ at gird position $(x,y,z)$ is smoothed with its six neighbors. Similar constraint can be applied to $\beta$. The choice of weight $w_s$ in the smoothness constraint is crucial as it is intended to balance neighbouring parameters while preserving the complex light pattern. Typically, voxels located further away from the light sources have smoother illumination, so they are supposed to have stronger weights in the smoothness constraint compared to the closer ones. More details regarding to weighting of the constraints are discussed in Section 4.3.4.

In addition, it needs to be noted that each observation from images each observation from the images may not exactly correspond to the center of a voxel. To prevent the resulting estimations in the lookup table from being pixelated, each observation is assigned to interpolated parameters based on its neighboring voxels, rather than the parameters at its nearest neighbor. This results in increased smoothness through the estimated parameters. In this paper, trilinear interpolation with eight neighbors is used to interpolate the lookup table parameters for each observation. However, having a unique solution on one point is not sufficient to assign unique values to its neighboring voxels. Therefore, it is necessary to ensure that at least eight points with unique solutions are presented in each group of eight neighboring voxels.

#### Pure Water (Complete Backscatter) Constraints

During deep ocean missions, underwater vehicles take hours to dive down to the sea floor. During this period, camera records numerous images of pure water, containing only illuminated water in the scene. These images are usually considered as useless data for the mission. However, they contain the maximum illumination backscatter information, which can also be used to set up

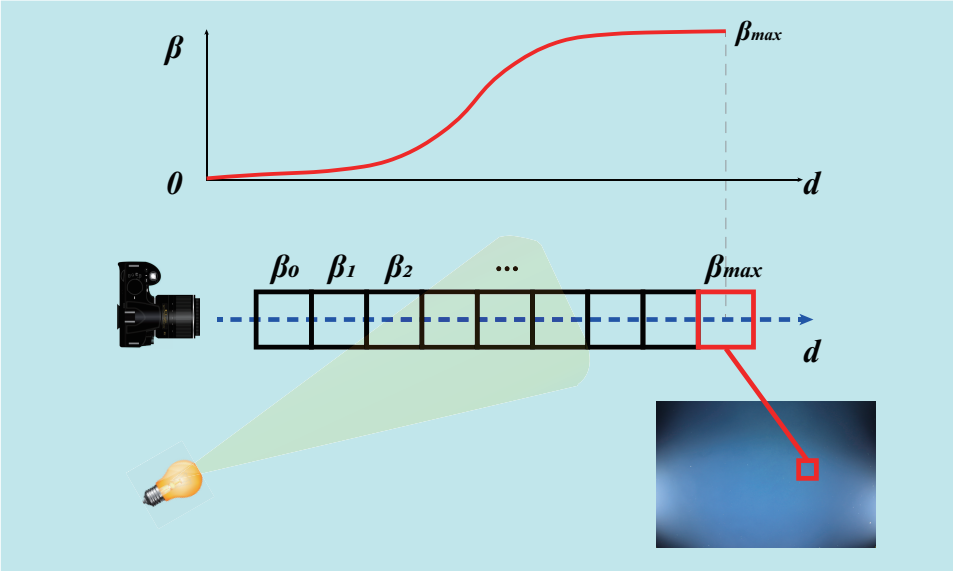



Figure 4.5: The backscatter factor $\beta$ is monotonically increasing along each viewing ray. Pure water image records the full backscatter information present in the scene, which corresponds to the maximum $\beta$ value for each viewing ray.

constraints for lookup table estimation (see Figure 4.5). Each pixel in a pure water image, denoted as $I_{pw}$, can contribute a direct constraint to all the $\beta$ terms at each slab $N$ along the same viewing ray:

$$\beta_N \leqslant I_{pw}. \tag{4.7}$$

This constraint establishes the upper bounds for the $\beta$ values. When underwater imaging platforms are operating at high attitude, pure water images can be directly used to subtract the backscatter component from the underwater images (Köser et al. 2021), which will be discussed in Section 4.3.6.

### 4.3.3 Hierarchical Parameter Estimation Strategy

Estimating parameters for the entire lookup table poses a challenge as it requires sufficient observations for each voxel to achieve a unique solution. To address this problem, this thesis proposes a hierarchical strategy for parameter estimation that proceeds from coarse to fine resolution. The optimization solver starts to estimate the lookup table at very low resolution, and the estimated parameters as used as the initial values for the next iteration with higher resolution until the final target resolution is reached. This approach allows for a more efficient and accurate estimation of the parameters and enables us to fill the entire lookup table, even in areas where there are no observations available.

### 4.3.4 Weights and Accuracy

In the process of estimating the lookup table parameters, proper weighting of the constraints is crucial, as pixel observations may have varying degrees of uncertainty due to different distances and illumination conditions. To achieve this, a lookup table is pre-rendered under single point light illumination, using predefined water parameters, and is used to define the weights for the three types of constraints.

To compute the weights for the smoothness constraints, the mean gradient magnitudes within and between the slabs of the lookup table are used. As a general trend, the illumination becomes weaker and smoother as the distance from the light source increases, resulting in parameter values that are closer in proximity. The mean gradient magnitudes $\overline{G}$ of $\alpha$ and $\beta$ are calculated for each slab to quantify the smoothness of individual slabs. These values are then utilized to calculate the weights $w_S$ for the corresponding smoothness constraints on slab $N$ and between neighboring slabs $(N, N+1)$. This is achieved through the following equations:

$$\begin{aligned} w_{s,\alpha}(N) &= 0.01 \times 0.7/\overline{G}_N \\ w_{s,\alpha}(N, N+1) &= 0.01 \times 0.7/\overline{G}_{N,N+1} \\ w_{s,\beta}(N) &= 0.01/\overline{G}_N \\ w_{s,\beta}(N, N+1) &= 0.01/\overline{G}_{N,N+1} \end{aligned} \tag{4.8}$$

It is important to note that $\alpha$ and $\beta$ are in different value scales, and hence a factor of 0.7 which represents the expected average intensity of the scene, is included in $w_{s,\alpha}$ to bring them to the same scale. Additionally, an empirical value of 0.01 is used in all smoothness weights to balance their impact compared to other constraint types.

Weights of observed pixel intensities for each color channel are determined by their signal-to-noise ratio (SNR). The digital camera noise is usually categorized into three main sources: shot noise, dark current noise, and read noise. In underwater robotic mapping missions, fixed exposure time and a small aperture are often used to prevent motion blur and maintain a large depth of field. In such scenarios, the dark current noise portion in the image can be considered a constant term, and read noise is also constant as the entire image sequence is captured by the same camera and dynamic range. Pixel values have different uncertainties based on the scene depth and illumination conditions. Objects at further distances are usually under weaker illumination, leading to lower SNR and larger uncertainty due to fewer photons reaching the pixel. Additionally, the forward scattering effect becomes more significant as the distance increases, which further degrades pixel observation quality. This effect can be modeled using a distance-dependent Gaussian PSF (Jaffe 1990). This thesis integrates the SNR and forward scattering models, along with the inverse distance weight, to calculate weights for pixel observations (known color and corresponding constraints) in the lookup table parameter estimation. The weight of known color constraints $w_{kc}$ is modeled as follows:

$$
\begin{aligned}
w_{kc} &= \frac{1}{d_v} \cdot \frac{snr}{(e^{0.5*d})^2} \\
\text{where } snr &= I/(n_{shot} + n_{const}) \\
n_{shot} &= 0.01 \cdot \sqrt{\overline{I}_{sN}} \\
n_{const} &= \overline{I}_{sN}/(snr_{0,g} * \overline{I}_{s0,g})
\end{aligned}
\tag{4.9}
$$

Here, $\frac{1}{d_v}$ represents the inverse distance weight, and $d_v$ is the observed point's distance from the corresponding voxel center. The PSF is approximated by $\frac{1}{(e^{0.5*d})^2}$, where $d$ is the camera distance to the observed point. The approximation of expected shot noise $n_{shot}$ is computed from the mean intensity $\overline{I}_{sN}$ of slab $N$, which can be approximated under the gray world assumption (with intensity 0.7) by $\overline{I}_s = \overline{\alpha}_N \cdot 0.7 + \overline{\beta}_N$, where $\overline{\alpha}_N, \overline{\beta}_N$ are the mean values on slab N. The green channel of the first pre-rendered slab $snr_{0,g}$ is used as the reference value, which assumes a 20 db SNR. The constant noise for RGB channels $n_{const}$ can be computed by referring to the first slab green SNR.

Similarly, the correspondence constraint weight $w_{c_{1,2}}$ can be computed from two corresponding known color weights according to the Pythagorean theorem:

$$
w_{c_{1,2}} = \frac{w_{kc_1} \cdot w_{kc_2}}{\sqrt{w_{kc_1}^2 + w_{kc_2}^2}}. \tag{4.10}
$$

### 4.3.5 Experiments and Results

This section presents the methods to estimate the lookup table parameters for image restoration under complex illumination conditions by utilizing the combination of the constraints mentioned above. It involves using known color calibration objects as references to estimate the lookup table parameters. Section 4.3.2 has discussed that the basic model (Equation 4.1) contains two unknown parameters ($\alpha$ and $\beta$) for each color channel in each voxel of the 3D lookup table, at least two known colors on the calibration objects are necessary to estimate the parameters in each voxel. Additionally, the two known colors should be widely separated in order to obtain robust parameter estimations.

To validate the effectiveness of the method, several experiments were conducted. The initial experiment involved the use of simulated in-air data with an artificial point light source to prove the concept of calibrating the lookup table using multi-view images and demonstrate its applicability in in-air applications. Subsequently, a real in-air lab experiment was conducted. The third experiment utilized simulated deep clear underwater datasets with two arbitrary color boards, followed by a simulated dataset with a turbid water setting, to test the effectiveness of our method. These experiments showcased that our approach is not limited to widely separated known colors and that the quality of restoration is closely related to the SNR of the input images. Furthermore, a real-world lab experiment was performed, employing a single chessboard with two color patches to demonstrate the possibility of simultaneous geometric calibration and lookup table estimation, which provides a practical solution for real-world applications. Furthermore, this thesis explores the feasibility and prerequisites for solving the lookup table parameters estimation without known

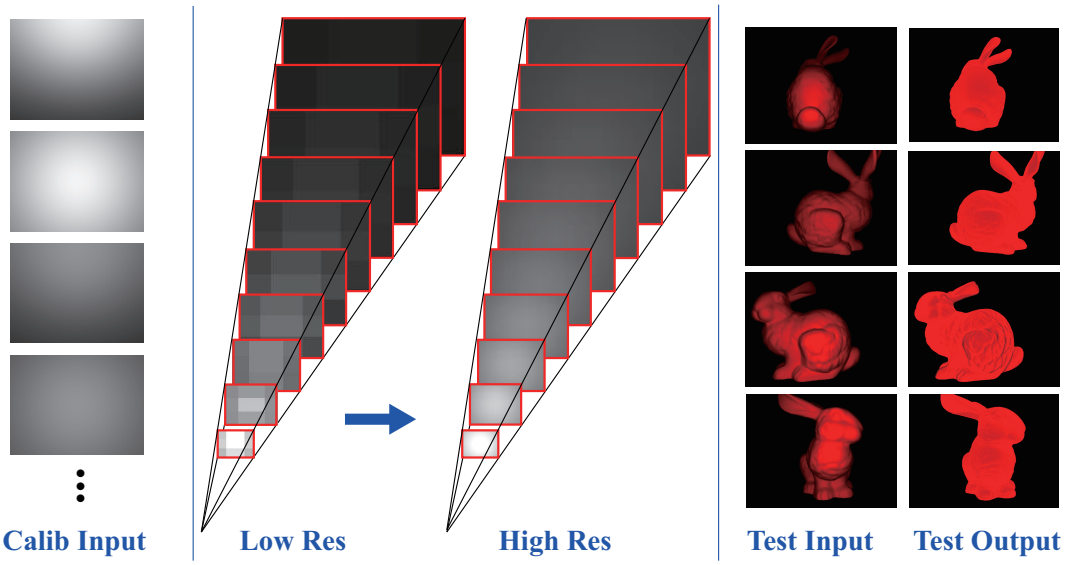



Figure 4.6: Experiment results on the synthetic in-air dataset. Left: simulated multi-view whiteboard images that were used as input to calibrate the lookup table. Middle: the coarse-to-fine estimation of the lookup table. Right: the test images under the same lighting configuration, along with their corresponding restored albedo images after applying the estimated lookup table.

the color of calibration objects. It presents the restoration results obtained from simulated in-air data with artificial illumination and outlines the challenges arising when applying this strategy to underwater scenarios.

### In-air Calibration by Using White Calibration Boards

***Validation on simulated data***: As mentioned previously, the proposed method can also be applied to correct artificial light patterns in images captured in-air. In this case, backscatter can be ignored (i.e., all $\beta$ values are set to 0), and only the transmission factor $\alpha$ in each voxel needs to be estimated. Therefore, one known color object is sufficient to calibrate the lookup table. Thirty in-air images of a simple white board with corresponding depth maps were simulated from different distances using Mitsuba3 (Jakob et al. 2022), where a point light source was placed at the same position and moved along with the camera. $40\times30$ sample points were extracted from each image to calibrate the parameters in the lookup table. Each sample point provided a known color constraint, together with the general smoothness constraint, allowed us to estimate the parameters using the Levenberg-Marquardt algorithm based on Ceres Solver (*Ceres Solver*).

As shown in Figure 4.6, the coarse-to-fine strategy first estimated a low-resolution ($4\times3\times10$) lookup table, which was then used as the initial values for the later high-resolution ($40\times30\times10$) lookup table parameter estimation. Once the lookup table was estimated, it was tested on images of a uniform red color textured Stanford Bunny, which were simulated under the same lighting configuration. As can be seen, the proposed method effectively removes the uneven light pattern, resulting in properly recovered albedo of the model.

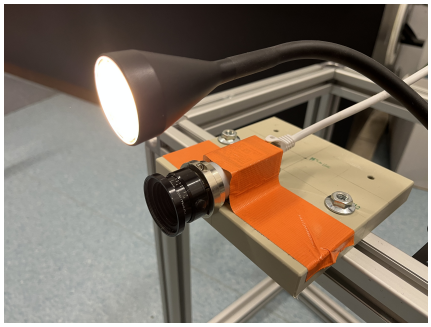
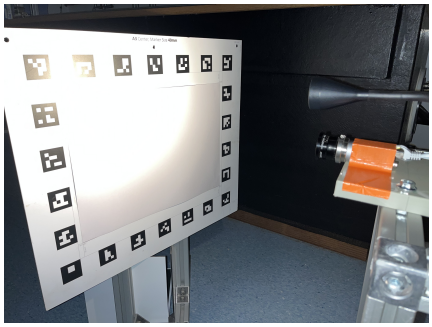



Figure 4.7: A rigidly-coupled camera-light system which was used in our laboratory experiment to capture several images of the self-designed calibration white board from different distances. Pixels in the center area of the board were used to calibrate the lookup table for the imaging system.

***Validation on real experimental images***: A similar experiment was conducted on real captured images using a camera-light system (consisting of a Basler acA1920-50gm camera with a Schneider Apo-Xenoplan 2.0/20 lens and a reading lamp) that is rigidly-coupled (see Figure 4.7). A self-designed calibration white board was used as the calibration object and multiple images of the board were captured from different distances to estimate the lookup table of the camera-light system. To ensure accurate calibration, we assumed that the camera was already geometrically calibrated and that all captured images were undistorted accordingly. Additionally, the camera's radiometric response was assumed to be linear. The area of interest (AOI) was the center of the board covered with white Lambertian material. Sample points were selected from this area in the images to calibrate the lookup table. To provide depth information for the sampled points, AruCo markers on the board edges are detected and the relative poses between the camera and the board were estimated.

Figure 4.8 shows the results of the real in-air lab experiment. For estimating the lookup table parameters, sample points with computed depth were extracted in the AOI from thirty-five images of the calibration board. The coarse-to-fine approach (from $8\times5\times10$ to $40\times25\times10$) was used for calibration, and the final obtained high resolution lookup table was used to restore the test tilted board images captured under the same system. As shown in the figure, the correction process successfully removed the uneven light pattern. Moreover, the plotted intensity distributions along the lines in test images before and after the correction demonstrated that the recovered albedo over the entire AOI is relatively constant.

#### Underwater Calibration by Using Two Different Color Boards

In underwater cases, two unknown parameters ($\alpha$ and $\beta$) need to be estimated in each channel in each voxel, at least two known color objects are required to calibrate the lookup table.

To validate the effectiveness of the proposed method, two underwater datasets with significantly different water types were simulated: clear deep water (Jerlov water type IA) and turbid coast water (Jerlov water type IC), using the state-of-the-art Monte Carlo ray-tracing technique based on Mitsuba3. Both datasets were rendered under the same camera-lighting setup, with a camera having 90 degree field of view and two rigidly co-moving point lights placed 40 cm to the

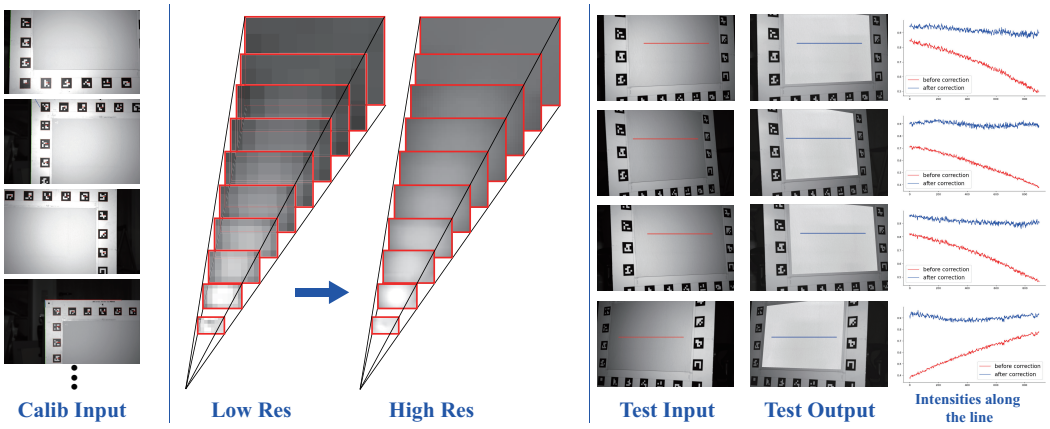



Figure 4.8: Experiment results on the real captured in-air dataset. Left: Multiple images of the self-designed calibration white board are used to calibrate the lookup table. AruCo markers on the board are detected to estimate the poses of the board, providing depth information for the AOI. Middle: Initial low-resolution lookup table estimation is refined to produce the final lookup table. Right: Test images of a tilted board captured under the same lighting configuration, with corresponding restored images. Pixel Intensities were sampled at the same position along a line from both test input and output images. Their plots indicate that the uneven lighting effect is properly removed after correction, resulting in a consistent albedo for the AOI.

left and right of the camera. To test the robustness of our method, additional challenges were deliberately introduced to the simulated data. These challenges included limiting the number of samples per pixel (spp) to 512 during the Monte Carlo ray-tracing procedure, which resulted in an approximate 10% noise level, and saving the simulated data as 8-bit RGB images rather than high dynamic range float images. This decreased the accuracy and the SNR of the calibration data. Furthermore, two arbitrary color boards are used for simulating the calibration dataset (specifically, boards with RGB colors of [181, 110, 30] and [80, 160, 90]), instead of using widely separated colors like black and white. To account for the different visibility conditions in the two types of water, viewing frustums in different ranges were defined for each dataset. In the case of the clear deep water dataset, the lookup table was defined for depths ranging from 0.5m to 2.5m. For the turbid coast water dataset, the lookup table was defined for depths ranging from 0.5m to 1.5m, as beyond this point the object was no longer visible.

To calibrate the lookup table under deep water settings, thirty-one color board images with depth maps were simulated from different distances ranging from 0.5m to 2.5m. During the coarse-to-fine optimization, the unknown parameters $\alpha$ and $\beta$ in each voxel were estimated simultaneously. Figure 4.9 illustrates the final obtained lookup table, which was used to restore the test images. The test images were generated from a virtual color checker that under the same environment settings as the calibration images.

Similarly, in the simulated turbid coast water experiment, ten images for each color board at distances ranging from 0.5m to 1.5m were rendered to calibrate the lookup table. Once the lookup table was estimated, images of a virtual color checker under the same turbid water conditions were rendered to test the restoration method, as shown in Figure 4.10.

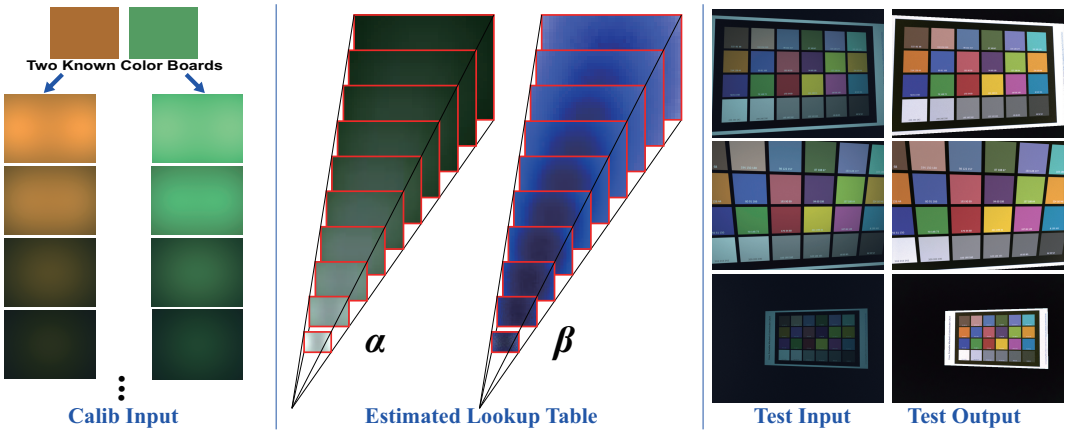



Figure 4.9: Experiment results on the synthetic clear deep water dataset. Left: Input images of two known color boards used to calibrate the lookup table. Middle: The final estimated lookup table visualizing the values of transmission ($\alpha$) and backscatter ($\beta$) parameters in the viewing frustum. The color mapping in the figure is scaled for better visualization. Right: Test images of a color checker rendered under the same lighting and deep water conditions, along with the corresponding restored images obtained using the calibrated lookup table.

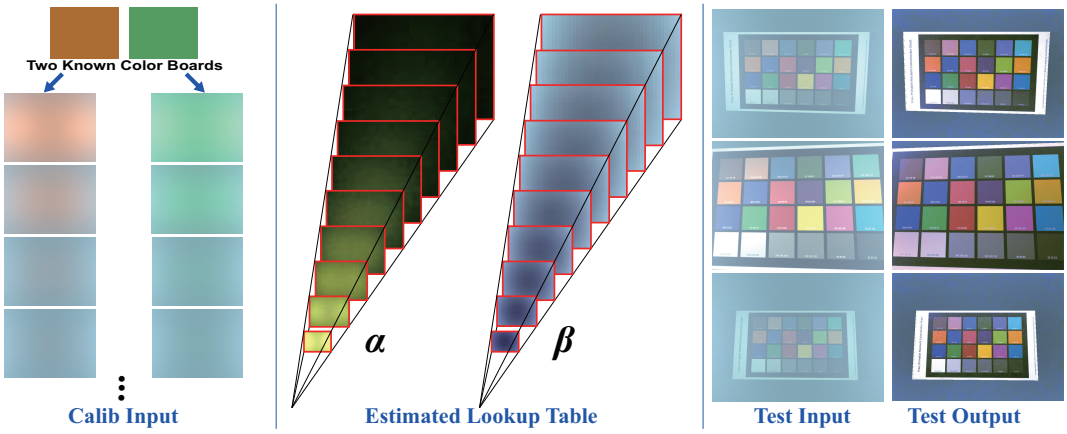



Figure 4.10: Experiment results on the synthetic turbid coast water dataset. Left: Input images of two known color boards used to calibrate the lookup table. The images demonstrate the strong scattering effects present in the turbid coast water environment, resulting in poor visibility of objects. Middle: The final estimated lookup table showing the values of transmission ($\alpha$) and backscatter ($\beta$) parameters in the viewing frustum. The color mapping in the figure is scaled for better visualization. Right: Test images of a color checker rendered under the same lighting and turbid water settings, along with the corresponding restored images acquired using the estimated lookup table.

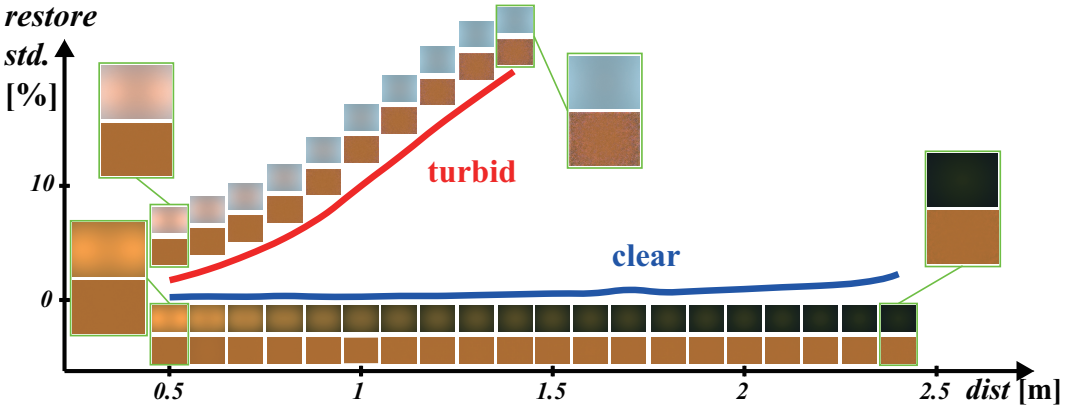



Figure 4.11: Standard deviation of restored images at different distances relative to the camera in the clear and turbid water datasets. The images displayed along each line represent the simulated underwater images, while the images below them depict the corresponding restored images used for computing the std. It is clear from the visualization that the SNR of turbid water images decreases much faster than that of clear water images, leading to higher std values in the restored images.

As depicted in Figure 4.9 and 4.10, the proposed method effectively eliminates water and lighting effects while accurately restoring object albedo. The restoration quality is directly influenced by the SNR of the input images. In Section 4.3.4, it has been discussed how images captured under stronger water effects and greater scene distances tend to exhibit lower SNR. In underwater imaging, as the scene distance increases, more light is absorbed by the water, leading to greater color attenuation, stronger forward-scattering effects and increased backscatter. Multiple images of a known color board at various distances relative to the camera in both clear and turbid water environments were rendered. These images were then restored using the corresponding estimated lookup table. The line plots shown in Figure 4.11 illustrate the standard deviation (std) of the restored images at different distances for both water conditions. As expected, the SNR of the images decreases with increasing distance, resulting in an increase in the std values of the restored images along the distance axis. In turbid water, the SNR decreases at a much faster rate compared to clear water images. This difference in SNR reduction leads to higher and more rapidly increasing std values in the restored images of turbid water conditions.

Table 4.2 presents the pairwise error of each color checker patch, computed as the absolute differences between the restored image and the ground truth color of each patch. In the clear deep dataset, the restored images exhibit high quality, with restoration errors mostly below the level of image noise. Despite the challenging conditions of the turbid coast dataset, characterized by poor visibility and very low SNR, some patches are even overexposed, the proposed method still provides a significant visual improvement after restoration, with the majority of patch errors kept below 25%.

Table 4.2: Pairwise error (in %) in RGB channels of each color checker patch, computed between the restored image to the ground truth values for the first test image in Fig. 4.9 and 4.10.

| Clear deep | Col 1 | Col 2 | Col 3 | Col 4 | Col 5 | Col 6 |
|---|---|---|---|---|---|---|
| Row 1 | [2.11, 0.73, 2.67] | [0.61, 1.39, 1.60] | [0.86, 0.21, 4.42] | [0.03, 1.21, 0.26] | [2.18, 2.38, 10.67] | [2.15, 4.45, 14.58] |
| Row 2 | [3.68, 2.79, 2.49] | [1.22, 0.81, 2.60] | [0.40, 0.06, 0.15] | [0.47, 0.84, 2.12] | [3.07, 3.14, 0.85] | [6.14, 3.44, 1.22] |
| Row 3 | [0.96, 0.95, 0.12] | [1.23, 2.37, 2.25] | [1.00, 0.67, 1.38] | [3.11, 1.27, 0.48] | [3.35 1.33, 6.24] | [9.68, 2.90, 10.79] |
| Row 4 | [4.37, 5.04, 5.08] | [1.52, 1.02, 8.74] | [9.25, 9.26, 6.75] | [0.06, 0.18, 2.35] | [7.89, 8.43, 9.43] | [7.82, 8.40, 6.95] |

| Turbid coast | Col 1 | Col 2 | Col 3 | Col 4 | Col 5 | Col 6 |
|---|---|---|---|---|---|---|
| Row 1 | [1.01, 7.55, 14.53] | [1.66, 0.55, 22.65] | [1.56, 0.85, 25.76] | [1.89, 1.16, 3.01] | [7.19, 4.16, 23.64] | [7.63, 6.74, 17.29] |
| Row 2 | [8.44, 3.76, 20.02] | [3.87, 5.21, 34.33] | [1.42, 3.86, 6.63] | [1.31, 7.36, 5.82] | [5.08, 2.89, 3.39] | [1.08, 5.46, 7.61] |
| Row 3 | [12.40, 14.85, 39.10] | [7.10, 2.53, 14.96] | [0.18, 12.39, 7.22] | [5.30, 0.90, 8.24] | [2.04, 3.02, 21.46] | [18.97, 3.00, 25.33] |
| Row 4 | [4.69, 4.69, 5.08] | [12.20, 12.23, 21.48] | [7.89, 5.92, 21.99] | [0.85, 1.64, 18.50] | [6.27, 9.54, 16.69] | [9.19, 13.14, 8.10] |

### Underwater Calibration by Using Single Board with Two Known Colors

A more practical approach for obtaining two known colors involves distributing them on a single board, such as a chessboard with black and white patches. This allows us to perform the lookup table calibration by filming only a single board. Additionally, using a chessboard offers the advantage of simultaneous camera geometrical calibration, which is particularly beneficial for real robotic missions with limited operation time and energy supply.

In the experiment, a custom underwater camera system enclosed in a dome-port waterproof housing (see Figure 4.12) was utilized. The system consisted of a Basler daA1600-60uc color camera equipped with an Evetar M118B029528W fisheye lens. Two rigidly co-moving light sources were positioned on the left and right sides of the camera, with a distance of approximately 15 cm from the camera. The camera was carefully adjusted to the center of the dome port using the techniques outlined in Section 2.4.1 to eliminated the underwater refraction effect. Similar to the previous experiments, the camera underwent both geometric and radiometric pre-calibration. Additional materials were added into the water tank to amplify the water effects, thereby intensifying the challenge for image restoration. For calibration, a standard chessboard was used as the target. Sample points were selected from the central region of each chessboard patch to calibrate the lookup table, and the relative poses between the camera and the board were estimated based on the chessboard corners, which were used to compute the depth information for each sample point.

As shown if Figure 4.13, the estimated lookup table effectively describes the light patterns generated by the two artificial light sources. two light cones are widely separated at close distance and gradually merge to the center when distance increases. The separation and merging of the two light cones with distance are clearly visible, and a slight shift of the right-side light cone towards the image center, indicating a greater tilt of the right-side light source towards the camera (see Light 1 in Figure 4.12). These observations affirm the accurate estimation of the lookup table. The test images in the same figure showcase the successful removal of strong lighting patterns and underwater effects, resulting in the recovery of texture and consistent appearance. The presence of colorful boundaries in the restored images is attributed to insufficient information on the dark region in calibration images, leading to erroneous parameter estimation. Furthermore, the dark regions exhibit a noticeably low SNR, thereby exacerbating the noise in these areas. Correspond-

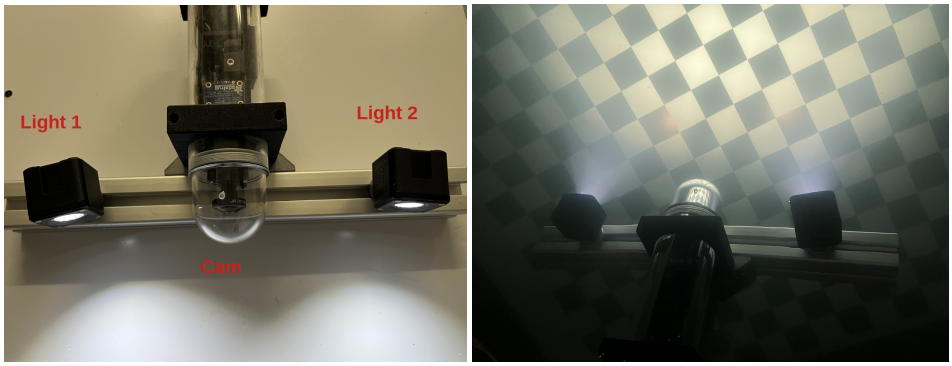



Figure 4.12: Left: The underwater camera system with dome port housing, accompanied by two rigid co-moving light sources. Right: A normal chessboard served as the calibration target and the center area of each chessboard patch was selected to facilitate the lookup table parameter estimation.

ing confidence maps are also computed and displayed in Figure 4.14, provide a visual representation of the confidence level for each pixel in the restored images. Higher intensity values indicate stronger confidence in the accuracy of the restored colors for those pixels. The confidence value is influenced by both the original color information and the results of the lookup table estimation. Black patches in restored images indicate the absence of valid calibration data in those specific voxels, resulting in incorrect estimation of the lookup table parameters. Additionally, certain pixels may be overexposed (mostly in blue and green channels), such as the bright spot in the first test image, causing low confidence values in the blue and green channels, while higher confidence is still maintained in the red channel for these pixels.

To demonstrate the effectiveness of the proposed method, it is compared with other three methodologies, including the popular image enhancement method CLAHE (Zuiderveld 1994), the well-known image dehazing method DCP (He et al. 2010), the state-of-the-art underwater image restoration approach Sea-thru (Akkaynak et al. 2019) [1], and the latest method Minimal Color Loss and Locally Adaptive Contrast Enhancement (MLLE) (Zhang et al. 2022). As shown in Fig. 4.15, CLAHE slightly improves the homogeneity of image brightness, but hazy backscatter persists, and the color of the image remains unchanged. DCP fails completely due to the sensitivity of the dark channel to illumination, such strong artificial lighting compromising the estimation of the dark channel. Sea-thru is heavily influenced by strong uneven lighting, resulting in overexposed restoration results. MLLE enhances image contrast and slightly improves color, but the enhanced color tends to be grayish. In contrast, the images restored using our method clearly exhibit the removal of complex dynamic light patterns and significant recovery of the object's true albedo, showcasing the effectiveness of our approach in restoring underwater images under strong complex illumination conditions.

It's important to note that most of these approaches were not designed for deep sea scenarios, and they use a single image as input (except Sea-thru, which requires an additional depth map). The proposed method has the advantage of a lookup table estimation procedure from multiple

[1] This thesis employed an unofficial implementation of the method via https://github.com/hainh/sea-thru

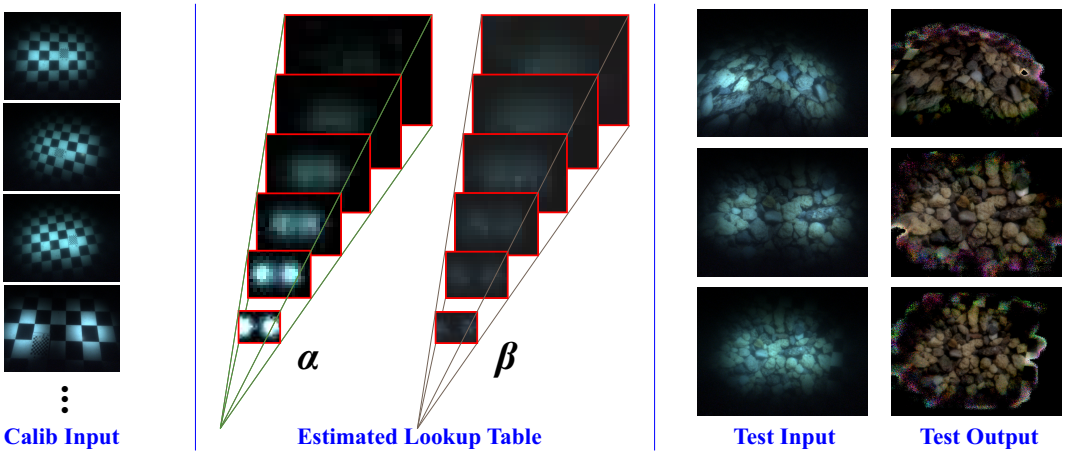



Figure 4.13: Experiment results on the real lab underwater dataset. Left: Forty-one input images of a single chessboard were utilized for the calibration of the lookup table. These images exhibit noticeable light patterns and strong water effects, resulting in poor visibility. Middle: The final estimated lookup table displaying the values of the transmission ($\alpha$) and backscatter ($\beta$) parameters within the viewing frustum. Right: Test images captured by the same system under identical water conditions, alongside the corresponding restored images obtained using the estimated lookup table. The presence of colorful boundaries in the restored images can be attributed to the lack of informative data in those areas during the calibration process, leading to erroneous estimation of lookup table parameters. Moreover, the dark region exhibits a notably low SNR, further exacerbating the noise in these area.

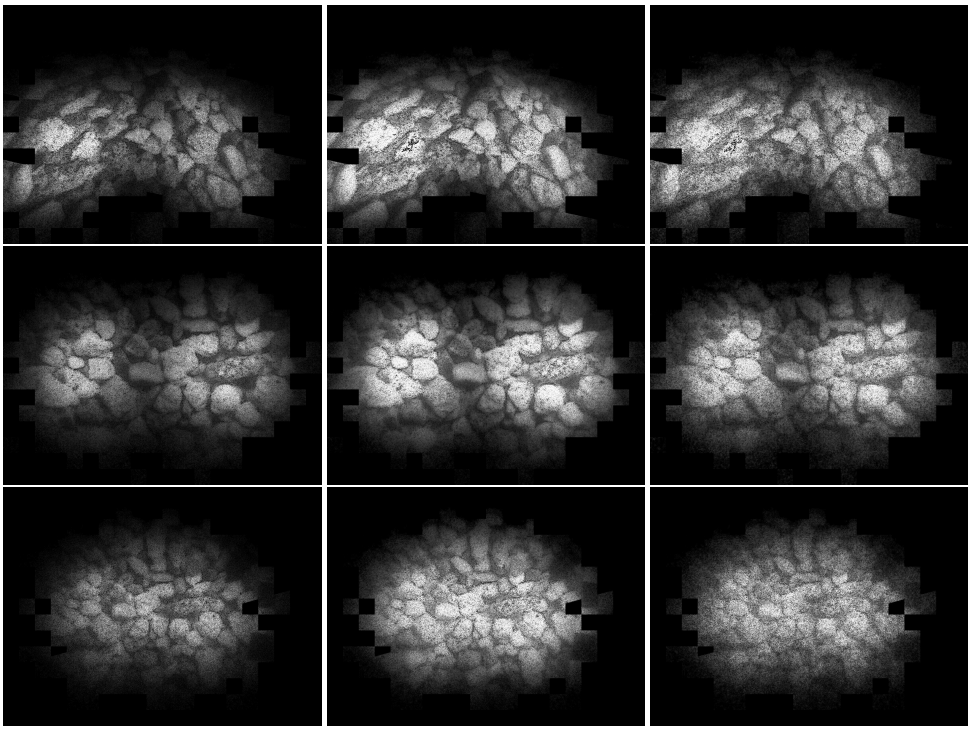


Figure 4.14: The confidence maps (R, G, B channels from left to right) for the corresponding restored images in Fig. 4.13 showcase the level of confidence in the restoration process, with brighter values indicating higher reliability. These maps offer visual representations of the accuracy of restored colors at each pixel. Dark boundaries result from low illumination in those areas, leading to low SNR in both calibration and test images. Black patches in the confidence maps signify regions with insufficient information for parameter estimation or noisy color data, leading to a lack of confident estimation in the lookup table. Notably, the green and blue channels exhibit brighter values than the red channel due to the stronger absorption of red color by water, resulting in weaker signals and lower SNR in red channel. Additionally, some overexposed areas, mainly in the green and blue channels, display low confidence, while the red channel retains a higher level of confidence, as its intensities remain within an optimal range.

calibration images, distinguishes itself with the ability to simultaneously address the removal of complex inhomogeneous illumination patterns and the restoration of the object's true albedo. The comparison emphasizes that existing methods fall short in handling both challenges concurrently. Despite requiring a pre-calibration procedure for lookup table estimation, the proposed method stands out as the only one currently capable of effectively tackling the complex underwater imaging conditions. This capability is particularly advantageous for autonomous mapping in deep ocean scenarios.

### Parameter Estimation from Correspondences

Previous experiments have demonstrated the viability of estimating the lookup table for underwater image restoration when utilizing known color calibration objects. In such instances, known color constraints serve as the primary source of information for estimating the lookup table parameters, with other constraints offering supplementary information in regions not covered by the known color constraints. This section delves into the scenario where known color calibration objects are unavailable and explore the potential of leveraging correspondence information from multi-view images of arbitrary scenes to calibrate the lookup table. In this case, the constraints mainly arise from correspondences.

Before delving into the methodology of the correspondence-based approach, it is essential to revisit the role of correspondence constraints within the known color-based approach and assess their influence on the estimation of lookup table parameters. The simulated turbid water dataset used previously is employed here to demonstrate the impact of correspondence constraints. In order to showcase this impact, the experiment focus on a specific $4\times4$ region within one of the slabs of the lookup table. Within this region, all known color information was intentionally removed. If one were to attempt the direct estimation of the lookup table without supplementary constraints, the parameters within this region would remain unaltered throughout the optimization process. (refer to Figure 4.16, second column). When solely employing smoothness constraints as supplementary factors, the empty region would be interpolated using information from neighboring regions with known color constraints (as seen in the third column of Figure 4.16). The calibrated values would gradually spread to the uncalibrated region over successive iterations. For a $4\times4$ area, this coverage would occur within just two iterations. On the other hand, when using only correspondence constraints, calibrated values from outside regions which are constrained by known colors would integrate with the uncalibrated parameters within the test region to form each correspondence constraint. In the uncalibrated area, the super-pixel centers are extracted and utilized to establish the correspondence constraints. Only those centers that have correspondences outside the test region with known parameters would be constrained with a unique solution, while other voxels within this region would possess unconstrained estimated values. When denser super-pixels are extracted within the test region, a greater number of correspondences are generated, resulting in more voxels' parameters being estimated with unique solutions. The influence of these correspondence constraints and their effects on parameter estimation are depicted in the fourth and fifth columns of Figure 4.16. The last column displays the outcomes obtained by integrating smoothness and correspondence constraints within the test region. Unlike the outcomes solely based on smoothness constraints, which involve straightforward value interpolation from neighboring voxels, and those relying solely on correspondence constraints, which may leave

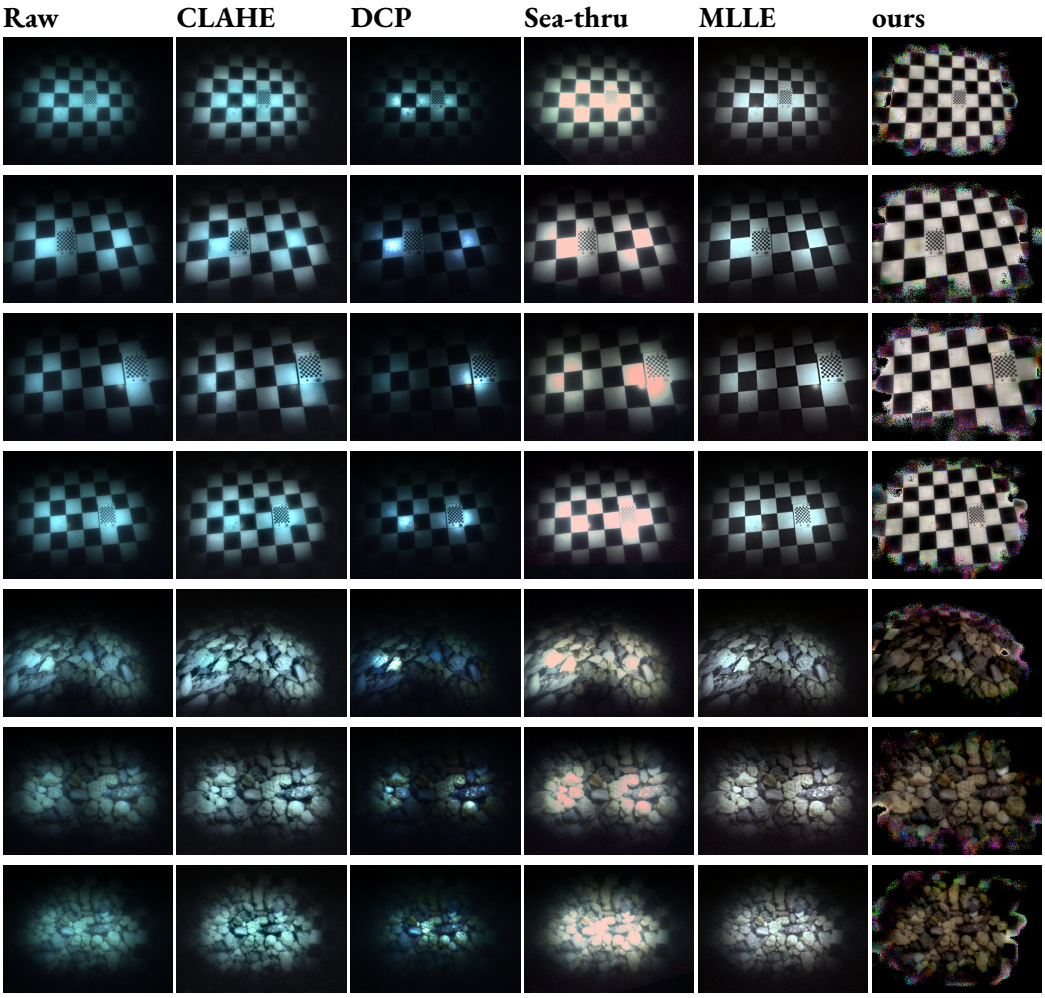



Figure 4.15: Underwater image restoration (or enhancement) results using CLAHE (Zuiderveld 1994), DCP (He et al. 2010), Sea-thru (Akkaynak et al. 2019), MLLE (Zhang et al. 2022) and our proposed method.

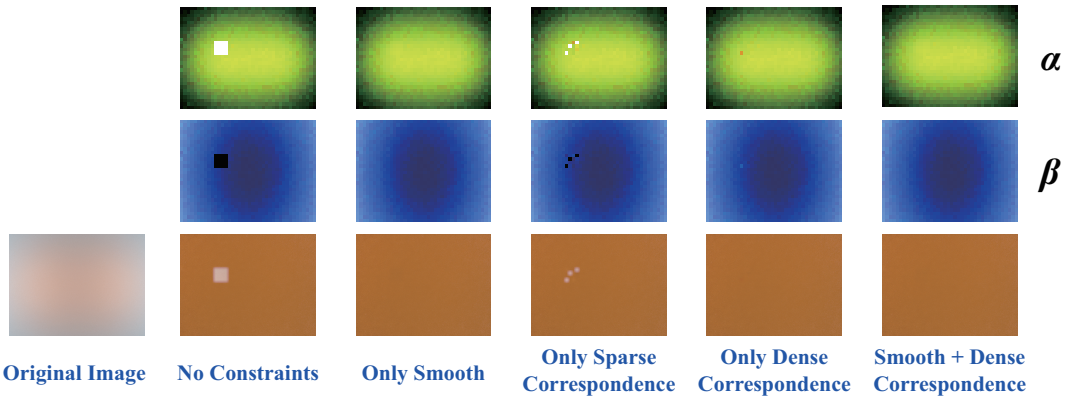



Figure 4.16: Comparison of lookup table estimation results using different supplementary constraints. From left to right: (1) Original turbid underwater test image. (2) Lookup table estimation results with only known color constraints, a deliberately chosen 4×4 test region where all constraints has been removed. (3) Results with only smoothness constraints in the test region. (4) Results with only sparse correspondence constraints within the test region, the correspondence constraints linking the unconstrained voxels inside the region with the constrained voxels outside. (5) Results with only dense correspondence constraints within the test region. (6) Result with both smoothness and dense correspondence constraints integrated into the lookup table parameters estimation.

uncovered voxels, the integrated approach offers a more comprehensive and precise estimation of the lookup table within the test region.

With the known color constraints, it is noteworthy that half of the unknown parameters in each correspondence constraint are already resolved. This simplifies the process of achieving a unique solution for the equation system, given that half of the unknown parameters are already estimated. However, when inspecting the equation system relying on correspondence constraints, two distinct degenerate solutions can be identified in Equation 4.5. The first solution is $\alpha_{1,2} = 0$. This implies that when filming an object without any illumination, the correspondence constraints are automatically satisfied. The second degenerate solution arises when $I_1 = \beta_1$ and $I_2 = \beta_2$, which signifies the filming of a black body object and the correspondence constraints are again fulfilled. To prevent all $\alpha$ values from becoming zero, an additional normalization constraint was imposed on them ($\sum_{i=0}^{n} \alpha_i = 1.$). Similarly, in order to avoid $\beta_i$ from becoming the observed color, it is necessary for each voxel to capture multiple distinct colors during the data acquisition. Moreover, considering the potential errors in the color observations, if each voxel captures similar colors, the ambiguities still remains in the equation system. To mitigate this, it is crucial to capture images in complex scenes with a diverse range of colors. This ensures that each voxel obtains sufficient color observations, enabling the accurate estimation of lookup table parameters. After estimating the lookup table parameters, all $\alpha$ values are still normalized, requiring them to be scaled to the appropriate scale. The scaling factor can be directly estimated from a single voxel with an absolute $\alpha$ value, which is obtained from known color constraints.

#### In-air Calibration Predominantly Based on Correspondences

To validate the correspondence-based lookup table parameter estimation approach, this thesis conducted a test in a simulated in-air scenario. As mentioned above, achieving a unique solution for the correspondence constraints requires diverse color observations. Therefore, a 3D plane with random unknown color patches texture was used as the object, providing a wide range of colors to satisfy the correspondence constraints. To simplify the experimental setup, it was ensured that all observed points were within one slab by simulating images from a fixed distance and viewing direction to the textured plane. The camera was constrained to shift and rotate on a virtual plane above the object, while a co-moving point light source was placed in front of the camera.

Eighteen test images were generated to estimate the one slab lookup table parameters. As backscatter ($\beta$) is negligible in in-air images, the focus was solely on estimating the transmission factor $\alpha$ for each voxel. Figure 4.17 illustrates the entire restoration procedure: 300 super pixels were extracted from each input image, and with the known extrinsics of each image, the center of each super pixel was projected into the corresponding paired image to construct the correspondence constraints. Based on these constraints, the one slab lookup table (size: 16×12×1) with normalized $\alpha$ values was estimated. Subsequently, a single point from one of the images was selected, and its true color served as the scale factor to compute the absolute value for the corresponding voxel. The entire $\alpha$ values in the lookup table were then re-scaled by this voxel. Using the re-scaled lookup table, the colors of all input images were corrected. The resulting corrected images demonstrated the successful removal of uneven illumination. Furthermore, the plotted intensity distributions along the lines in the images, before and after the correction, indicated relatively constant intensity in each patch of the corrected images. The quality of the estimated lookup table parameters for each voxel depended on the observed intensities, with higher robustness achieved when there were more observed colors and greater diversity among these colors.

In the underwater scenario, theoretically, it is possible to attain a unique solution for lookup table estimation when an ample number of correspondences are provided within the same voxel. However, each correspondence constraint encompasses four unknown parameters intertwined through multiplication. To achieve sufficient constraints for every voxel, an extraordinarily dense observation and an exceedingly complex scene with diverse colors are required. Especially when observations are prone to errors, a challenge arises where the optimizer is difficult to distinguish whether the effects stem from the $\alpha$ or $\beta$ terms. This predicament remains an unresolved question that warrants further investigation.

### 4.3.6 Simplified Approach for Large Scale Mapping

In the general underwater image formation model (see Section 4.1), all the intricate aspects of illumination, water attenuation, and camera radiometric effects are unified as a multiplicative factor $\alpha$, while the complex scattering effects caused by all light sources are aggregated as an additive factor $\beta$. The primary challenge in underwater image restoration lies in estimating these two parameters for each pixel, which is essential for determining the seafloor albedo. In many underwater mapping scenarios, especially when covering vast areas with a predominantly flat and uniform seafloor while imaging platforms are flying at high attitude, it is possible to simplify the estimation of the lookup table. This simplification involves focusing solely on the estimation of a single slab within the 3D lookup table. Such an approach proves particularly useful and practical

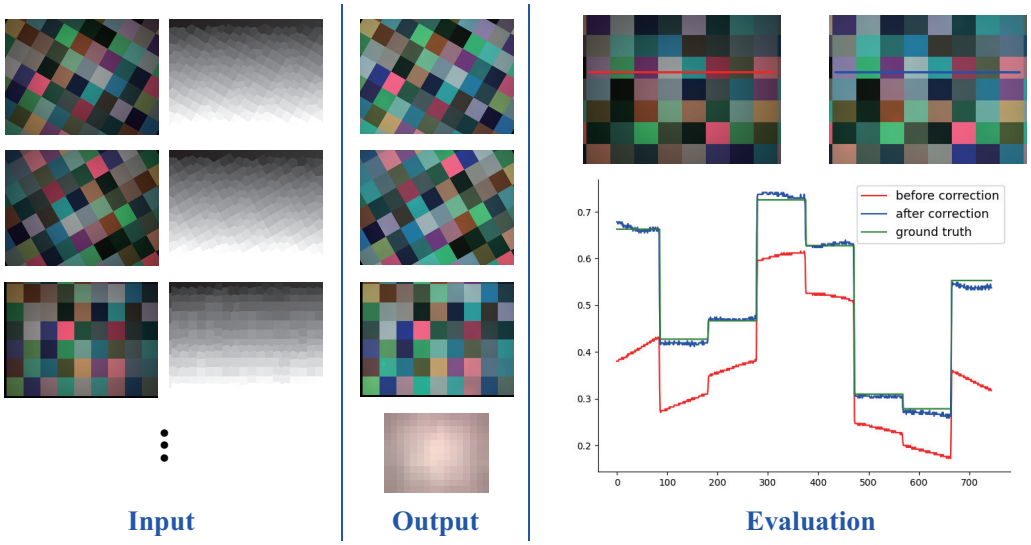



Figure 4.17: Experiment results on the simulated in-air dataset primarily utilizing image correspondences for image restoration. Left: Test images of a colorful plane used in the image restoration experiment, with 300 super pixels extracted from each image to construct correspondence constraints. These images exhibit uneven illumination due to a co-moving point light source. Middle: Corresponding restored images obtained using the estimated one slab lookup table (shown in bottom) from the correspondence constraints. Right: Evaluation of the restoration result. The top two images show examples before and after restoration, while the bottom figure displays the blue channel intensities sampled along the lines in these images. In the original images (in red), noticeable gradients are observed in each patch due to point light shading, and the values significantly deviate from the ground truth intensities (in green). After the correction, the intensities (in blue) become relatively constant in each patch, closely matching the ground truth values. This demonstrates the successful removal of uneven illumination and the accurate restoration of color in the images.

when calibration reference targets are unavailable throughout the mission. This section provides a brief summary of the approach developed in (Greinert 2015; Köser et al. 2021). The method predates the lookup table approach, with certain concepts serving as inspiration for the latter. Consequently, this work can be viewed as a simplified version of the lookup table approach under certain restrictions. Further details on the method and analysis can be referred to the original paper from the authors.

#### Estimation of the additive and multiplicative images

The additive (backscatter) factor actually depends solely on the relative pose of the light source with respect to the camera (rather than the ground) and the angular characteristics of the light source and water properties. As discussed in Section 3.3.2, in the context of deep-sea autonomous robotic mapping, where all the light sources are fixed relative to the camera, the total backscatter pattern remains stable and can be recorded in pure water images. These images contain no objects in the scene, only the water itself. Moreover, the distances between objects in the scene are significantly greater than the distances between the light sources and the camera, and very little light is scattered into the camera sensor from far distance. This trend continues with further distances, such that scattered light will be insignificant after a certain distance.

Before reaching the working altitude above the seafloor, the robot should already capture a certain number of these images that solely depict the water column. These pure water images offer valuable insights into the total scattering phenomena. In these specific pure water images only the backscatter effect is observable. Given that most of the visible scattering occurs in proximity to the camera, particularly when AUVs operate at high altitudes to achieve larger mapping footprints, the backscatter component present in the mapping image $\beta_{high}$ can be approximated by the pure water image $I_{pw}$:

$$\beta_{high}(\mathbf{x}) \approx I_{pw}(\mathbf{x}). \tag{4.11}$$

The pixel values within the pure water image are employed for the purpose of subtracting the backscatter component from other underwater images. In practice, the pure water images might also contain bright floating particles or dark regions located very close to the camera, which are not within the light cone. These atypical measurements should be treated as outliers, necessitating the use of a robust estimator to derive $I_{pw}$ from multiple observations at each pixel location. A common approach is to calculate the temporal median across several pure water images at each pixel position, which serves as a basis for inferring the ideal total scatter image.

Furthermore, based on practical experience, this scenario holds also when operating in murky waters with artificial illumination, albeit with reduced distances. In this case, the camera needs to be positioned closer to the seafloor to capture images of it. For various practical considerations such as ensuring homogeneous seafloor illumination, avoiding pronounced shadows, and facilitating robot maneuverability close to the ground, light sources that are also closer to the camera are used. This results in a similar relative geometry between the seafloor, camera, and lighting setup. As a consequence, the assumption is made that the majority of the scattering originates from the first few meters in front of the camera, and during seafloor mapping, a sufficient altitude is maintained to capture most of the scattering. Substituting Equation 4.11 into the general underwater image formation model 4.1 yields:

$$I(\mathbf{x}) \approx \alpha(\mathbf{x}) \cdot I_0(\mathbf{x}) + \beta_{high}(\mathbf{x}). \tag{4.12}$$

Let's first consider a scenario where the deep-sea floor is mostly flat and devoid of significant features, appearing primarily as sediments in the images. When AUVs fly at a fixed altitude with a stable pose above this flat seafloor and capture an "all-seafloor image" consisting exclusively of uniform sediment with a known color, the illumination pattern would typically remain visible in such an ideal image of the consistent seafloor. Using this image, which encodes water attenuation, the artificial illumination pattern and the camera radiometric characteristics, the transmission image $\alpha$ can be computed as follows:

$$\alpha(\mathbf{x}) = (I_{seafloor}(\mathbf{x}) - \beta_{high}(\mathbf{x}))/I_{sediment}(\lambda). \tag{4.13}$$

In Equation 4.13, the seafloor color $I_{sediment}$ serves a role analogous to a white balance reference in typical photographs. By setting it to grey, even if the actual seafloor color is brown, all other colors adjust proportionally in a consistent linear manner. $I_{sediment}$ remains correct with respect to a global scale factor, which implies that we can later correct all images with just one global scale factor per color channel or wavelength if needed. mapping and reconstruction, this implies that, without prior knowledge of the seafloor color, it can be assumed to be grey, resulting in the enhancement of all images in a consistent manner. This allows for tasks like matching, SLAM, and stereo reconstruction, similar to mapping on land with a camera using a fixed but unknown white balance.

The estimation of the ideal seafloor image $I_{seafloor}(\mathbf{x})$ can be approached similarly to the ideal pure water image. For instance, in case multiple all-seafloor images exist, each perturbed by Gaussian noise, it is suggested to average them for estimation of $I_{seafloor}(\mathbf{x})$. These images can be acquired at various locations, provided that the relative orientation between the AUV and the seafloor remains constant (with the same altitude, pitch, and roll). In the common case the seafloor is not of uniform color, but "contaminated", e.g. by stones, fauna or other objects that don't conform to Gaussian noise. In such cases, it is advisable to employ a robust estimation method for generating the all-seafloor image, such as applying a median filter over a sequence of images.

Due to variations in the seafloor terrain, as well as changes in altitude and vehicle pose, the relative orientation between the camera and the seafloor tends to vary over a several hours mission. Therefore, it is not advisable to compute just one all-seafloor image for an entire image sequence. Instead, it is recommended to calculate an individual all-seafloor image in a sliding window fashion for each image. This approach accounts for short periods during which the vehicle's pose with respect to the seafloor remains relatively stable.

#### Results of the Simplified Approach

One of the primary goal of this thesis is to develop an approach capable of efficiently handling tens of thousands of photographs to facilitate large-scale deep ocean mapping. Most of the operations within this approach are suitable for parallel scheduling, and consequently, an implementation was done using CUDA (Petersen 2018), allowing parallelization at the pixel level. When the median computation of the current central image is completed, the sliding window will shift to the next central image, and the "oldest" image is replaced by a new image from the stream. To enhance

the efficiency of this process and ensure robust estimation, and considering that the illumination patterns typically exhibit high smoothness when captured from high altitudes, the approach in this thesis incorporates a standard procedure of downsampling the undistorted input images (by a factor of 8 in this study). This downsampled set of images, consisting of 7 images, is then used for the temporal median computation before upsampling is applied.

In Figure 4.18, an experiment was conducted using a commercial 3D reconstruction software (Agisoft Metashape) on approximately 100 images captured by an AUV equipped with a self-designed camera and light system in the Baltic Sea. These images depict a World War II torpedo sunken into the sediment. The image pre-processing method proposed here was employed alongside other approaches prior to the 3D reconstruction, and a comparative analysis of their outcomes was conducted. (a) showcases a completely fail reconstruction directly from the raw images. (b) presents the reconstruction from "Fusion Enhanced" images (Ancuti et al. 2012), which partially corrects the color of the objects. However, due to the presence of a strong and uncorrected uneven lighting pattern, the 3D reconstruction ultimately fails. (c) reveals the results of the reconstruction from "Multi Exposure Fusion" enhanced images (Galdran 2018). While the image is sharper than the original, it fails to properly recover color and light patterns, leading to an unsuccessful 3D reconstruction. (d) illustrates the reconstruction from images processed using homomorphic filtering with a fourth-order polynomial (Singh et al. 2007). Here, the 3D reconstruction succeeds because the uneven lighting is adequately removed, although the texture remains suboptimal, resulting in a foggy appearance. (e) presents the outcomes achieved through the proposed method, where a detailed, consistent 3D model free of visible seams is obtained. The texture is sharp, as the scattering is effectively removed, and color accuracy is preserved.

In another experiment, a set of normalized images was registered from a sequence captured at a depth of over 4 km in the Pacific Ocean's polymetallic nodule fields (Greinert et al. 2017). Micro-navigation data for the deep-sea robot was obtained using SfM techniques applied to the enhanced images. This navigation information was subsequently employed to stitch together the images, including both raw and pre-processed versions, as illustrated in Figures 4.19 and 4.20. The raw mosaic exhibited pronounced illumination effects at the boundaries, primarily stemming from the illumination patterns. This resulted in unclear visibility of seafloor features, as the foggy effects caused by scattering persisted.

A comparative analysis was conducted to assess the effectiveness of our proposed approach in comparison to other image processing techniques, including "Optimized Contrast Enhance" (Kim et al. 2013), "Fusion Enhance" (Ancuti et al. 2012), "Remove Backscatter" (Zhang 2016) and "Multi Exposure Fusion" (Galdran 2018) using the implementation and default parameters provided in (Wang et al. 2019). Note that most of these approaches were not designed for deep sea scenarios. Nevertheless, it is interesting to qualitatively evaluate their effects. It is evident that none of these methods produced consistent results for the deep-sea light cone setting. Parameter tuning might potentially improve results, but it can be seen that all of these approaches struggled with inconsistencies between overlapping images. Notably, the proposed approach offers a distinct advantage in that it does not necessitate manual parameter adjustments or algorithm retraining after each mapping campaign. The only approach that yielded reasonable results was the one presented in (Singh et al. 2007), which involved fitting a 4th order polynomial in log space. However, this approach exclusively addressed multiplicative effects, resulting in a loss of contrast. Furthermore, the degree of the polynomial had to be adjusted to suit the specific conditions of the

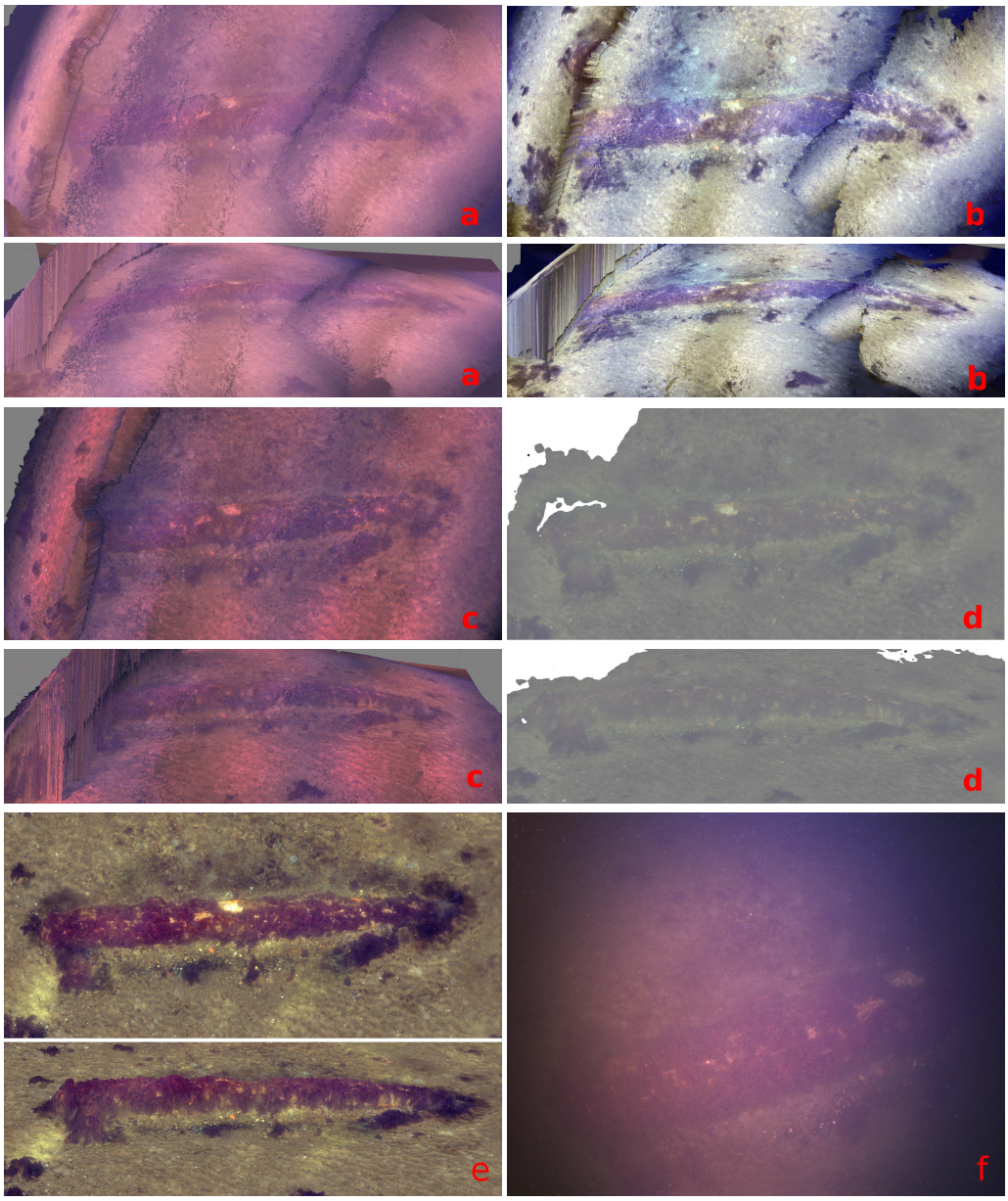



Figure 4.18: 3D reconstruction of a torpedo. The input images have been recorded with a white balance that stresses red too much. In (a) we see the 3D reconstruction using the raw images (each time a top view and a side view). It can be seen that the structure of the torpedo was not recovered correctly. (b) shows the reconstruction from the "Fusion Enhanced" (Ancuti et al. 2012) images, (c) from "Multi Exposure Fusion" (Galdran 2018), (d) from homomorphic filtering with a fourth order polynomial (Singh et al. 2007) and (e) from our proposed method. Only the reconstruction from our corrected images produces a straight torpedo model and a flat seafloor, whereas the reconstructions from all other image version have problems with the image data.

light cone. If the degree was too high, it led to the fitting and removal of scene structures, while if it was too small, it could not adequately account for the illumination pattern, particularly in the case of multi-LED setups.

This outcome underscores the value of proposed pre-processing approach for enhancing deep-sea visual mapping. It can significantly improves both sparse and dense correspondences and enhances overall mapping quality.

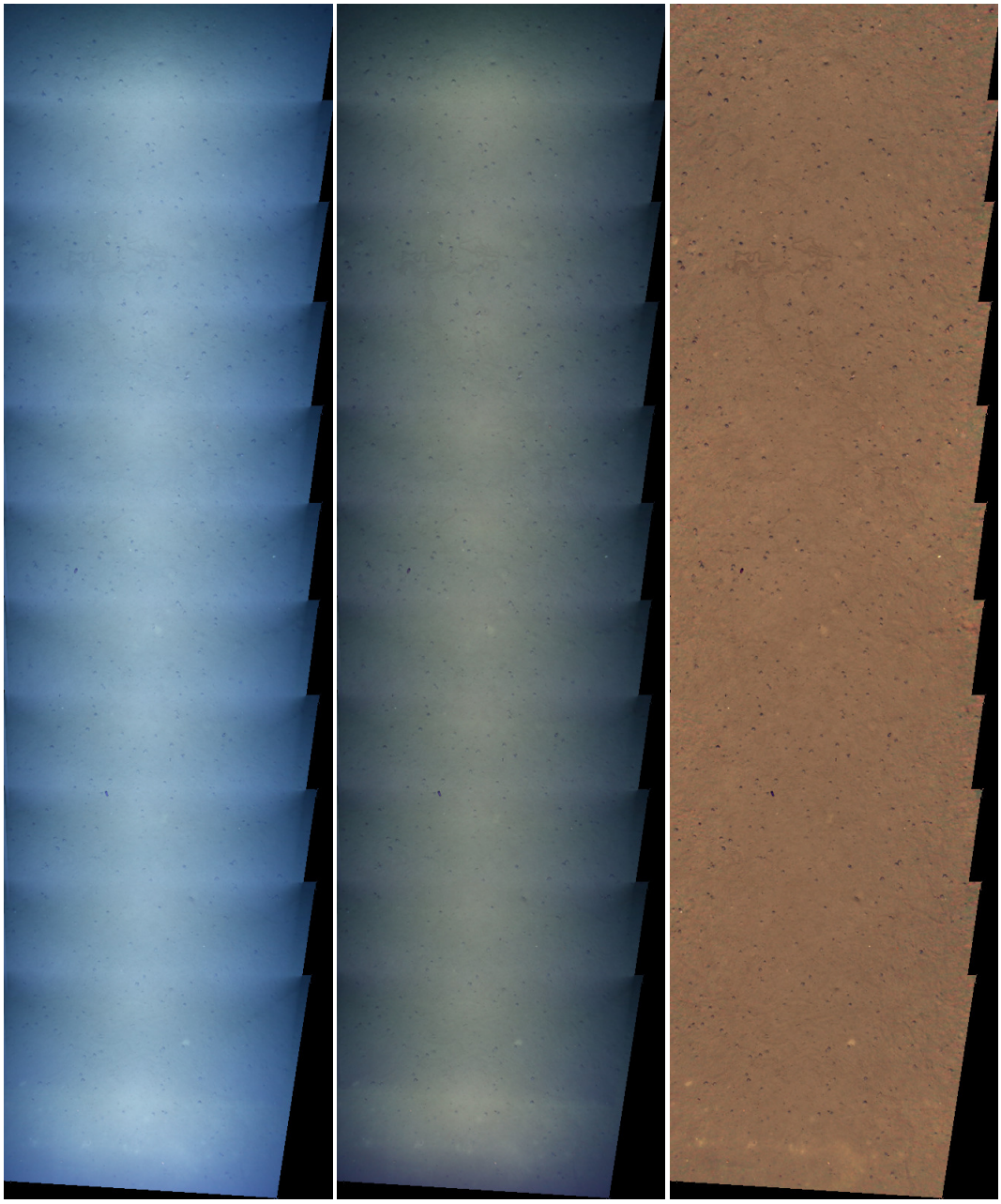


Figure 4.19: Deep sea mosaic stitched from raw photos (left), from intermediate images after removing backscatter (middle) and from images after backscatter removal and transmittance correction.

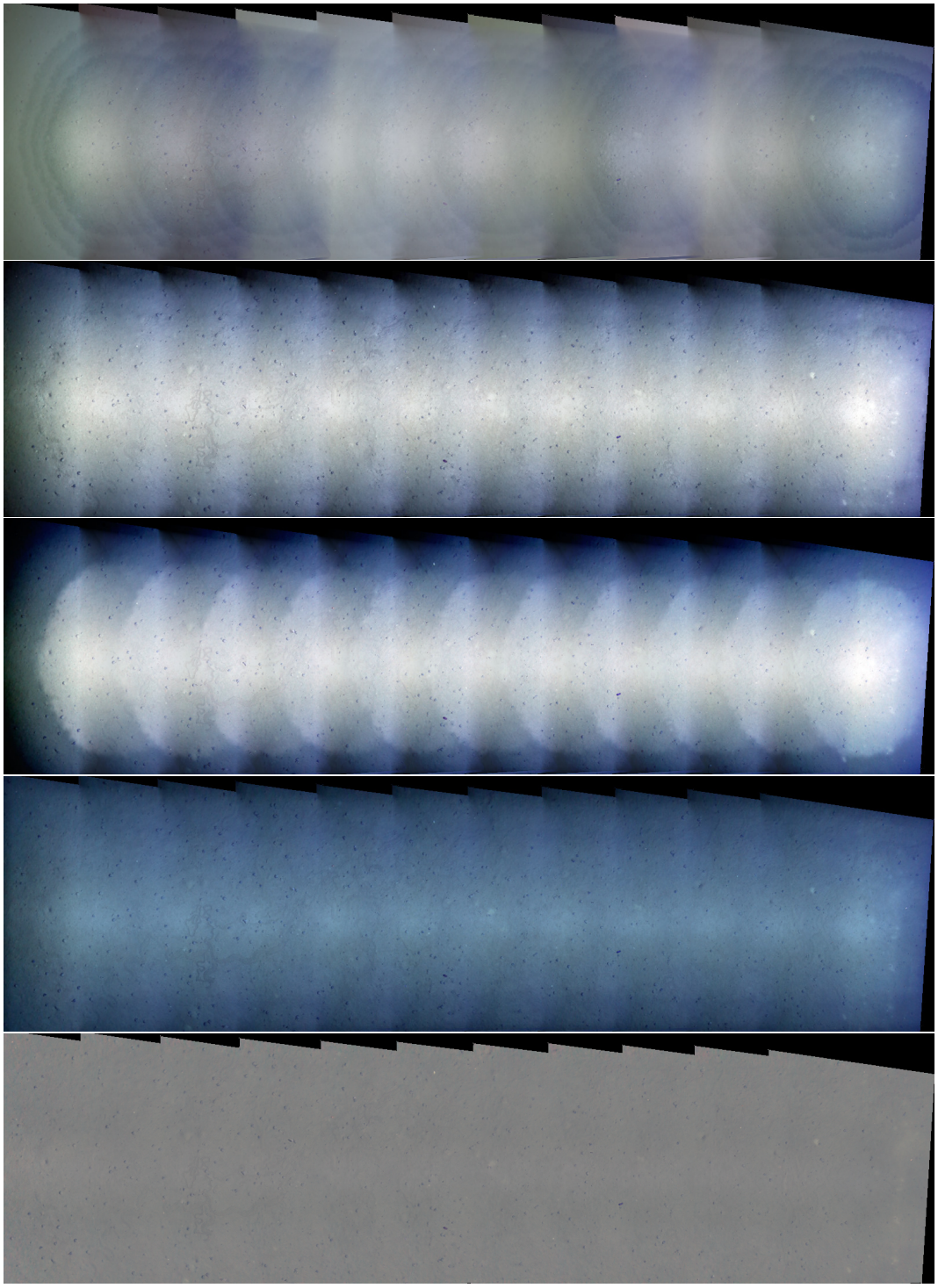


Figure 4.20: From top to bottom: Deep sea mosaic stitched using "Optimized Contrast Enhance" (Kim et al. 2013), "Fusion Enhance" (Ancuti et al. 2012), "Remove Backscatter" (Zhang 2016), "Multi Exposure Fusion" (Galdran 2018) and homomorphic filtering (Singh et al. 2007). All mosaics are stitched using the same geometry and blending parameters as Figure 4.19, they only differ in the pre-processing of the input images. It is evident that the quality of large-scale maps is heavily influenced by artifacts and lighting effects. Even if the uneven lighting is corrected, the presence of the backscatter effect can make it difficult to identify the actual seafloor structures.

# 5 Applications

The preceding sections have detailed various techniques and methods tailored for deep underwater vision. This section delves into their practical applications in the realm of deep sea imaging system designs and visual reconstruction applications. Specifically, the first application involves employing dome centering and calibration techniques for deep sea imaging systems to eliminate the refraction effect. This transforms refractive cameras into pinhole models, enabling the direct use of standard photogrammetry techniques for following processes. Later, this thesis demonstrates applications in underwater 3D reconstruction. In cases where the refraction effect cannot be avoided in the hardware aspect (e.g., flat port), the proposed iterative refinement reconstruction approach is employed to reconstruct underwater objects. In the radiometry domain, the deep sea simulation tools are not only utilized for simulating deep sea images but also for optimizing illumination systems. This optimization enhances the imaging quality for imaging systems mounted on AUV platforms.

## 5.1 Deep Sea Imaging System Design

In the previous geometry part, the geometric properties of the refractive dome geometry were explored. It was emphasized that when the camera's optical center aligns precisely with the center of the dome port window, the entire imaging system can be treated as a standard pinhole camera. This thesis introduces an adjustment approach aimed at centering the camera behind the dome port window, effectively eliminating refraction effects in captured images. For scenarios demanding very high accuracy where tiny refraction effects introduced by remaining centering offsets cannot be ignored, the proposed dome decentering offset calibration approach can be employed to fine-tune the system. Two representative systems are presented here: the first is a deep-sea survey camera system tailored for autonomous visual 3D mapping on AUV platforms, while the second is an underwater stereo camera system designed for quantifying marine bubble flow.

### 5.1.1 Survey Camera on AUVs

AUVs play a crucial role in ocean exploration and mapping due to their ability to operate independently for extended periods. This autonomy enables them to efficiently cover vast areas without direct human control, making them a cost-effective alternative to manned vehicles. Their independent operation eliminates the need for extensive support crews and expensive infrastructure. Operating autonomously is particularly advantageous in challenging and hazardous underwater environments, mitigating risks associated with deep-sea exploration. AUVs can continuously collect data over prolonged periods, making them invaluable for long-term environmental monitoring and large-scale area mapping. Their versatility is evident in their capacity to carry various sensing technologies tailored to different missions. In the realm of underwater mapping, AUVs

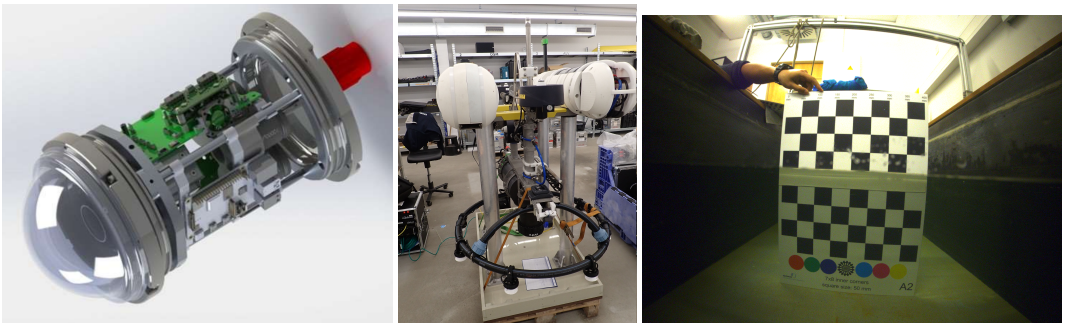


Figure 5.1: The CoraMo underwater survey camera system. Left: Technical drawing of the components of the CoraMo camera system (Image courtesy of GEOMAR AUV Team). Middle: Integration of the CoraMo system into the Girona 500 AUV platforms, featuring a ring of eight LEDs for illumination. Right: Visual feedback from the mechanical adjustment procedure, ensuring precise positioning of the camera at the center of the dome port.

equipped with underwater cameras offer a cost-efficient solution for achieving high-resolution 3D mapping of extensive seafloor areas, with the option to deploy multiple AUVs concurrently for parallel operations.

For high accuracy visual mapping with AUVs, it is required to know the exact position and orientation of the carried camera for each photo. Since water blocks the direct use of satellite navigation, and existing sensor such as Doppler Velocity Logs (DVLs) and IMUs on AUVs cannot fulfill the accuracy requirement, navigation information can only serve as initial poses for the image registration. The more accurate poses will be computed through image features matching. This requires the captured image to be calibrated w.r.t the optical distortion caused by lens system and multi-media refraction. For lens distortion it can be calibrated through a standard camera calibration procedure. For the refraction issue, Section 2.3 has pointed out that if the optical center of the camera is precisely aligned with the center of the dome port, incoming principal rays will not experience refraction, which can eliminate refraction effects in images.

The GEOMAR-AUV team developed the CoraMo[1] underwater survey camera system, featuring a Sony IMX253 1.1“ CMOS sensor with a f/3.5 fisheye lens that provides a 70 degree horizontal field of view and captures 12.34 megapixels (4104 x 3006) images. Housed within a pressure housing with a 100mm diameter spherical window, the camera is securely affixed directly to the port using a 3D-printed frame to maintain stable optical properties. The innovative mechanical adjustment approach outlined in Section 2.4 was employed to center the camera behind the dome port, as illustrated in Figure 5.1. This ensures the camera's suitability for survey and mapping purposes. The CoraMo camera systems are integrated into the Girona 500 AUV platforms by IQUA Robotics[2], eight high-performance LEDs arranged around the camera system in a ring to provide optimal illumination. Navigation information is directly sourced from the AUVs' compact C3 Inertial Navigation System (INS). The CoraMo system has been successfully utilized in seafloor mapping tasks, an illustrative example is shown in Figure 4.1.

[1] https://www.geomar.de/tlz/auv-autonome-unterwasserfahrzeuge/coramo
[2] https://iquarobotics.com/

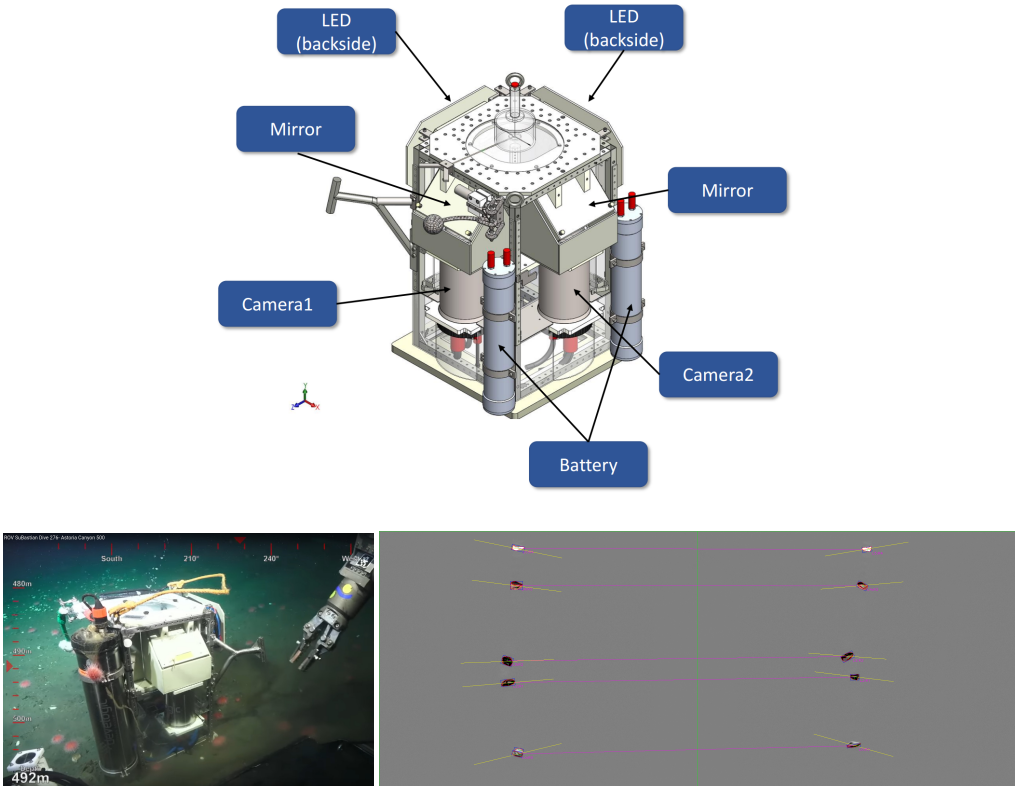



Figure 5.2: The Bubble Box system for marine bubble flow quantification. Top: System overview. Left: The system was deployed on a seep site at a depth of 1 km in the Pacific Ocean via an ROV robot arm. Right: Sample results of stereo epipolar geometry matching of bubble stream from left and right camera images (She et al. 2022b)

.

### 5.1.2 Stereo Camera for Marine Bubble Flow Quantification

Greenhouse gases, such as methane and CO2, contribute to climate change by trapping heat in the Earth's atmosphere. This trapped heat causes the planet to warm, resulting in shifts in climate patterns, rising temperatures, and other environmental changes. They are playing a significant role in enhancing the natural greenhouse effect, contributing to the ongoing global warming trend. Besides other sources, these gases are also released from seep sites on the ocean floor, and it is crucial to understand and accurately quantify gas emissions from these locations to address climate-related challenges. To address this, precise 3D measurement of released bubbles becomes imperative. The quantification of multi-phase flow parameters, involving the interactions of liquids and gases, is a key research focus with applications in both natural environments and industrial processes.

The GEOMAR Helmholtz Centre for Ocean Research Kiel has developed an advanced wide-baseline stereo camera system known as the Bubble Box, illustrated in Figure 5.2. Specifically designed for deep ocean bubble measurements, the Bubble Box can be deployed above a bubble

seep spot, allowing the bubble flow to pass through its central observation corridor. Two cameras, strategically positioned at 90-degree angles to the corridor, employ stereo photogrammetry to capture bubble images. To make the system size more compact, two cameras are placed to look upward through mirrors at 45deg angle, which are equivalent to two virtual horizontal views into the bubble rise corridor. Each camera is equipped with a Basler acA1300-200um mono color machine vision camera featuring an 8 mm C-Mount lens and rigidly mounted within a 10 mm diameter dome port. The cameras capture images at a resolution of $1024 \times 800$ pixels and record at a rate of 80 Hz.

Each camera underwent an initial centering process within the dome port, following the proposed techniques. This involved submerging the camera system halfway underwater and adjusting its position on the mount until the observed half-submerged object in the image became consistent at the air-water interface. Subsequently, the remaining decentering offset calibration was implemented using the technique outlined in Section 2.4.2 to fine-tune the system. A standard stereo camera calibration procedure was then applied to determine the cameras' intrinsics and the relative poses between them.

The algorithms employed for bubble measurement are the work of colleagues from GEOMAR and not part of this thesis. For more detailed information, please refer to (She et al. 2022b).

## 5.2 Refractive Underwater 3D Reconstruction and Mapping

Underwater visual 3D reconstruction plays a pivotal role in marine exploration, environmental monitoring, and resource management. It facilitates accurate mapping of underwater topography, supporting scientists in studying marine ecosystems and geological features. The technology aids underwater archaeology, preserving and documenting submerged cultural heritage. Moreover, it improves navigation for autonomous underwater vehicles, streamlines oil and gas industry inspections. Overall, underwater 3D reconstruction advances our understanding of the submerged world, informing conservation efforts and sustainable development in aquatic environments.

Section 2.4 has introduced a technique wherein centering the camera inside a dome port housing can effectively mitigate the refraction effects in images. The resulting system can be considered as a pinhole camera, enabling the application of standard photogrammetry pipelines for underwater 3D reconstruction. In this section, the thesis presents examples illustrating high-accuracy underwater reconstruction in scenarios where refraction effects cannot be mitigated. The examples utilize the iterative refinement reconstruction approach proposed in Section 2.5 to reconstruct underwater munitions from images captured by a camera within a flat port housing.

During and after the world wars, huge amounts of munitions were dumped into the sea. For instance, an estimated 1.6 million tons are still resting on the seafloor of German coastal waters of the Baltic Sea and the North Sea. These munitions not only endanger the environment, people and ships, but they also hinder extension of infrastructure. Besides the risk of explosions, the munitions also contain toxic substances or even chemical agents, which threaten the health of marine ecosystems as well as the food chain and people's lives. Thus, it is important to locate and map the munitions and to monitor potential drift, deformations and even appearance changes. To

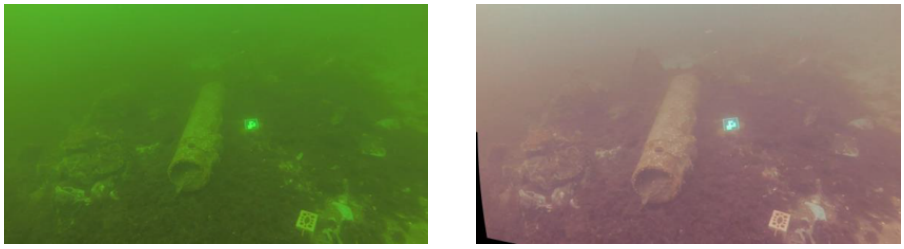


Figure 5.3: Original underwater munition image (left) and the refined image (right).

Figure 5.4: Reconstructed 3D model of a torpedo in the Baltic Sea via proposed approach.

monitor the munitions and provide the reference information (especially geometry information and the distribution) for government decisions, visual 3D reconstruction of these munitions is requested.

One critical site is located in Kolberger Heide, next to Kiel, Schleswig-Holstein, Germany. The early years test images of munitions (in this case, a torpedo) were taken by a GoPro Hero3 camera within a plastic underwater flat port housing. The 3D model was generated from 49 images in a challenging setting for 3D reconstruction, i.e. extremely greenish color in the Baltic Sea, bad visibility and all images were taken by divers in forward motion.

Early experimental images utilized a simple flat port camera system for underwater image acquisition, in which refraction effects cannot be avoided, the iterative refinement approach proposed in Section 2.5 was applied to this test dataset. Figure 5.3 (left) shows one of the original image, as well as the refined image (right, after the refraction correction, the image was further zoomed in to preserve the original resolution), after the geometry has been established during the first iteration of pinhole processing. A simple color adjustment was also applied to the images to further enhance image quality. This simple adjustment was chose because the radiometric calibration procedure was not conducted during data capture, and the metadata information was also missing, preventing the use of more comprehensive restoration techniques.

The final 3D model was established from the refined images. Due to the poor visibility, achieving a complete depth map for each image is becoming extremely difficult. Even though the ren-

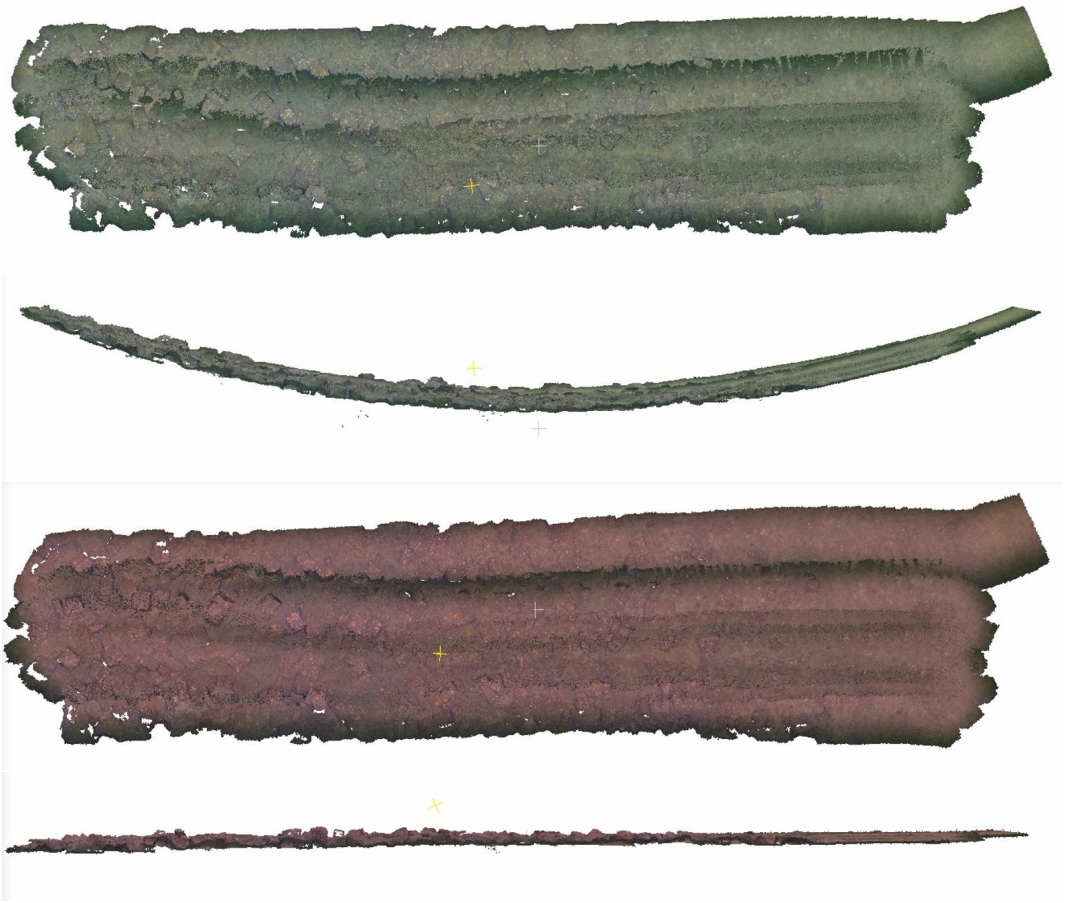


Figure 5.5: 3D reconstruction of the seafloor area with munition boxes. From top to bottom: top and side views of the direct reconstruction from original images, top and side views of the proposed iterative refined reconstruction.

dered depth map from the 3D model is used to correct the refraction effects for the images, there are still some noticeable discontinuities in the processed images which are so far unresolved. However, the discontinuity region does not cover the body of the torpedo, which didn't affect the 3D modeling of AOI. As it is shown in Figure 5.4, the refined images represent a useful quality of the 3D model with fine details.

Another dataset consisting of 488 images captured by the same flat port camera system was used to reconstruct a 3D model of a seafloor area approximately 30 meters by 7 meters, featuring munition boxes. The results are presented in Figure 5.5. The first two images depict the top and side views of the direct reconstruction using the original underwater images. It is evident that the camera intrinsics alone could not fully compensate for the refraction effect, resulting in significant bending of the 3D reconstruction over short distances. However, when the proposed iterative refinement was applied (see Figure 5.5, bottom), the bending effect was completely eliminated. These results clearly demonstrate the effectiveness of the proposed method in correcting the refraction effect for underwater 3D reconstruction.

## 5.3 AUV Lighting Optimization

As discussed in Section 3.3, once a proper image formation model is established, it becomes a valuable tool for estimating or optimizing various parameters within the model. A key application of the proposed deep underwater image formation model in this thesis is the optimization of complex illumination configurations for the underwater imaging system to enhance the quality of captured images.

In the early days, single light-source optimization could be devised through expert knowledge (e.g., (Jaffe 1990)) and possible subsequent human visual inspection of captured images for light configuration adjustments. This becomes impractical for more complex systems with several light sources. The emergence of systems with many flexibly attachable LEDs has even further complicated this process.

The specific lighting configuration significantly impacts the quality of captured underwater images, influencing the ease or difficulty of subsequent analysis. Traditional monolithic Xenon flashes are gradually being replaced by more flexible setups of multiple powerful LEDs. However, this raises the question of how to arrange these light sources, given different types of seawater and-depending-on different flying altitudes of the capture platforms. This thesis introduces a rendering-based coarse-to-fine approach to optimize contemporary multi-light setups for underwater vehicles, which has been preliminarily published in (Song et al. 2021b). It leverages physical underwater light transport models and considers ocean conditions and mission parameters to simulate underwater images as would be observed by a camera system with specific lighting setups. The thesis proposes systematically varying design parameters, such as each LED's orientation $\theta$, and analyzes rendered image properties (e.g., illuminated image area and light uniformity) to identify optimal light configurations.

### 5.3.1 Evaluation Factors

This subsection integrates the deep sea imaging simulator (see Section 3.3.1) into an optimization strategy that automatically synthesizes and analyses rendered images with different light configurations for an AUV. The optimization process suggests optimal lighting configurations tailored to specific flying altitudes of the robot, improving the image quality for different deep sea exploration tasks and aiding system designers in making decisions.

There is no gold standard approach to evaluate the quality of underwater photo illumination. It heavily depends on specific details the photos should depict, the observer's interests, or the intended analysis algorithm. Some people might look for subtle color differences of animals, while others would like to maximize the amount of light that falls into the image frame. This thesis focuses on generic mapping applications, aiming to utilize as much of the image as possible rather than optimizing for specific colors. Since the overlay of many light cones can create artificial patterns, potentially misinterpreted as scene patterns, a desirable property of illumination is also smoothness.

Consequently, an cost function is designed here, containing several terms quantifying desired illumination pattern properties. The pattern is analyzed by rendering a section of homogeneous seafloor from a predefined altitude and with predefined water properties using a target camera model. The resulting image is then inspected and represents a certain configuration (i.e., setting of

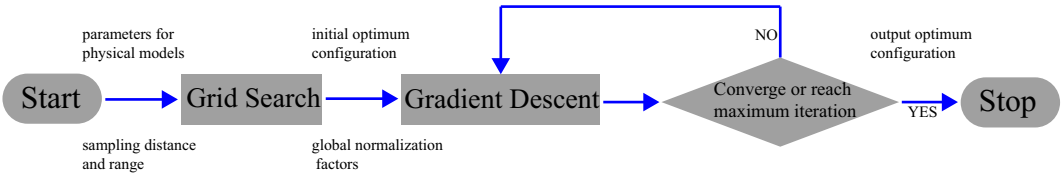



Figure 5.6: Workflow of the optimization algorithms.

illumination directions). A high-priority goal is to make as many pixels as possible well-lit. $W_{ratio}$ represents the ratio of pixels under good illumination (in the moderate intensity range). Additionally, considering the homogeneity of illumination for each pixel, two cost terms are computed from the local gradient ($W_{gradient}$) and from the image entropy ($W_{entropy}$). The final evaluation cost $W$ is defined as a weighted sum (with empirical weights $w_i$):

$$W(\theta) = w_1 \cdot W_{gradient} + w_2 \cdot W_{entropy} - w_3 \cdot W_{ratio}. \tag{5.1}$$

### 5.3.2 Optimization Algorithms

As the cost function defined in the last section for a multi-LED system is potentially non-convex, a two-step optimization approach is proposed for finding a suitable illumination configuration $\theta$ (see Figure 5.6). In the first step, a grid search is employed on the naturally confined parameter space, here only considering orientations of lights within physically realizable intervals. The grid search generates an image for each certain LED tilt angles, with fixed intervals. Each light can also be rotated orthogonal to the current dimension, adding more degrees of freedom to the entire setup. However, for mechanical reasons, lights are typically mounted in groups, restricting independent rotation in all directions. This reduces the dimension of the search space. For a discrete grid encompassing all possible tile angles, their cost values are calculated for the corresponding image (example shown in Figure 5.7, Left). The grid point corresponding to the image with the lowest cost is then chosen as the initial configuration for the next step, which involves local optimization.

For the second step of local optimization, a variant of simple gradient descent is implemented (Figure 5.7, Right). At the current best position in the parameter space from the first step, the numerical gradients of the cost with respect to the lighting configuration parameters were computed. This involves calculating the cost for small delta steps of all parameters and assembling the gradient from the small differences computed. The interval of the previous grid search, on the one hand, and the mechanical precision achievable for fixing physical LEDs to a robot, determine the approximate maximum and minimum useful step size in parameter space. The algorithm then iteratively takes steps in the gradient direction until convergence is achieved or the maximum number of iterations is reached. Many other downhill techniques can also be applied for fully automating the light configuration optimization.

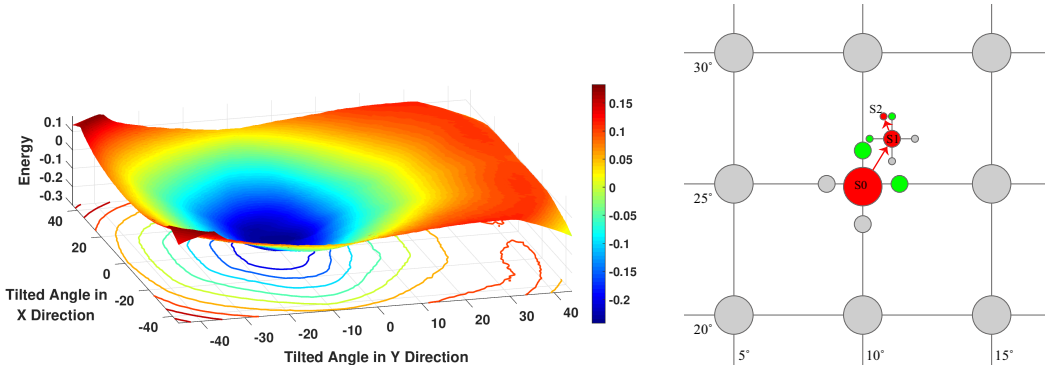



Figure 5.7: Left: Cost function results of a single light during grid search. (In this case, the camera is loacted 1m away from the light source along Y-axis and they all initially point perpendicular to the seafloor with 4m flying altitude.) Right: Optimization strategy in gradient descent steps of a single light.

### 5.3.3 Implementation and Test Results

The proposed method is implemented to optimize a real lighting system of a deep sea camera on an AUV (Kwasnitschka et al. 2016). This system has been designed for operation at high altitudes, up to 10 meters, and the optimization aims to find optimal lighting configurations for missions on operation altitudes of 4 meters and 7 meters, respectively. The AUV is equipped with a well-centered DSLR camera housed in a dome port. Twenty-four LEDs, arranged in 4 rows and 6 columns, are positioned 1.9 meters away from the camera (see Figure 5.8) on the lighting rigs. The images captured by this camera system in previous cruises at the same spot are accessible on PANGAEA[3], and these will be utilized as reference data in the simulation model.

Details regarding the parameter set for the renderer are provided below:

1. The peak wavelengths with respect to RGB channels are $\{650, 510, 440\}$ [nm], and the following values for the channels are referenced to these wavelengths.

2. The same LEDs described in Section 3.1.2 are utilized. The relative spectrum of the light source is $\{0.25, 0.35, 0.4\}$, measured by a spectrometer.

3. The expected Jerlov water type in the area is II, characterized by attenuation coefficients $\{0.37, 0.044, 0.035\}[\mathrm{m}^{-1}]$.

4. The seafloor is simulated as a flat Lambertian surface with a uniform brown albedo.

5. The camera in the AUV is a Canon EOS 6D with no significant vignetting. Intrinsic parameters are obtained through standard camera calibration. White balance parameters $k_{wb}(\lambda)$ are embedded in the metadata of the reference image. For relative sensitivity $k_{rs}(\lambda)$, the CIE-D50 Color Response of the Canon EOS 6D can be acquired from sources like dxomark[4].

[3] https://doi.pangaea.de/10.1594/PANGAEA.881850

[4] https://www.dxomark.com/Cameras/Canon/EOS-6D---Measurements

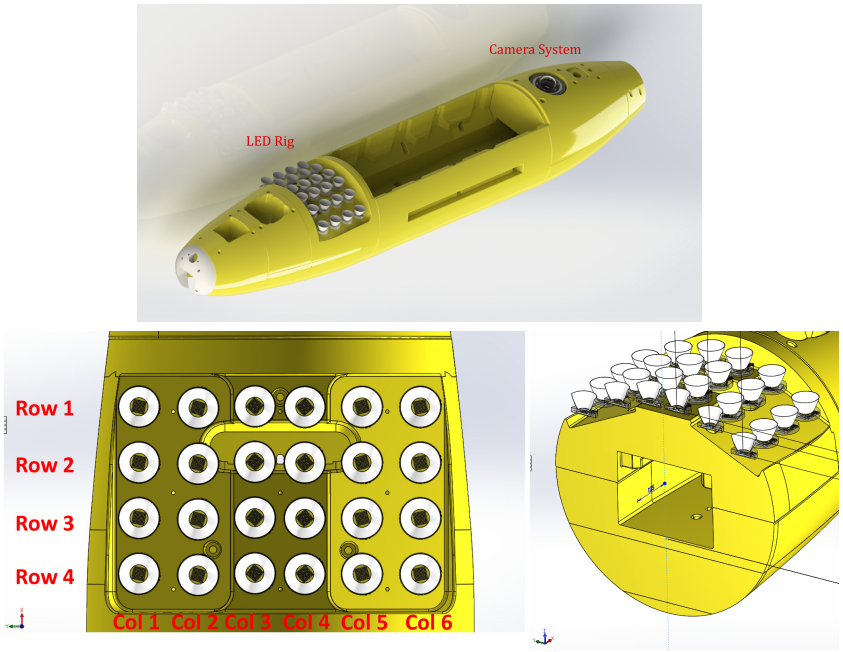



Figure 5.8: LED layout on the AUV rig, the camera is located 1.9m below the rig along the X-axis and its optical axis is along the Z-axis. The initial settings of all LEDs are perpendicular to the surface of the rig. Images courtesy of Emanuel Wenzlaff.

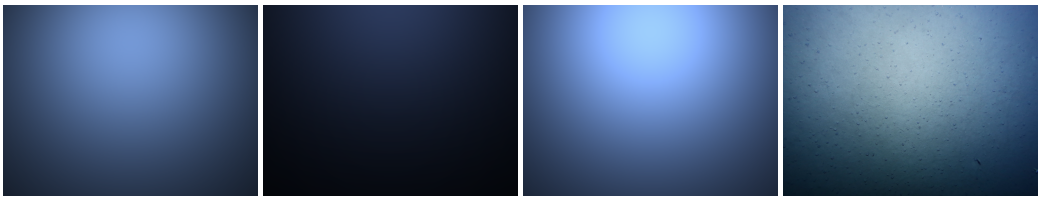


Figure 5.9: Simulation results of the reference image. From left to right: direct signal, back scatter, complete rendering result of the virtual seafloor, the reference image. Since the old LED rig was manually setup in the cruise, the precise light configuration of the reference image is not available. Only rough measurements are used for the simulator and the light pattern in the rendered image will be slightly different from the real one.

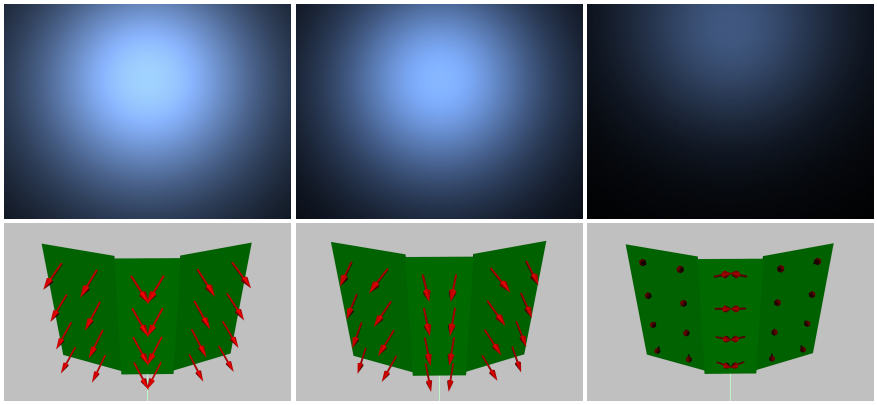


Figure 5.10: Final optimization results and their lighting configurations. First row: optimization result with 4m altitude (left),7m altitude (middle) and a bad example of illumination with 7m altitude (right). Second row: corresponding light layouts.

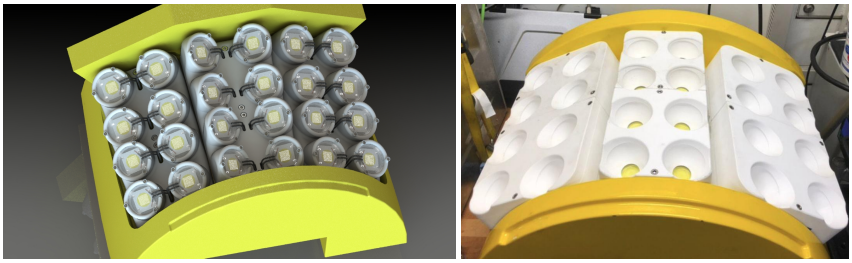


Figure 5.11: Left: Technical drawing of optimized LED configurations for 4m flying attitude mission in the Central Pacific. Right: Corresponding 3D printed LED rig bases. Courtesy: Emanuel Wenzlaff.

6. By comparing the reference pixel values with the computed absolute light irradiance values (see Figure 5.9), the global scaling factor $s$ in Equation 3.11 is determined.

7. The positions of all LEDs are obtained from the technical drawing of the AUV.

For different deep sea exploration tasks, images are simulated with the AUV operating at different distances above the seafloor to identify the respective optimal light configurations.

In the optimization, the sampling distance of the tilt angle was set to $15^\circ \in [-30^\circ, 30^\circ]$ in both the $x$ and $y$ directions for each LED (see Figure 5.8). It is assumed that LEDs on each row or column share the same tilt angle in corresponding dimensions, and the LED layout is symmetric along columns, continuously reducing the number of grid search samples from $5^{48}$ to $5^7$. Rendering of the scattering effect is the most time-consuming part, since the lights are fixed relative to the camera, the direct signal component pattern at the seafloor changes more significantly than the scattering component. Therefore, the scattering component in the grid search step was temporarily disabled, and only the direct signal was rendered during the initial optimization stage. The scattering was rendered only during the gradient descent step. The empirical weights $\{0.2, 0.2, 0.6\}$ were applied in the evaluation function (Equation 5.1). Once the initial optimum from the grid search was obtained, the gradient descent was run and always converged before 50 iterations. The

final optimal light configurations for 4m and 7m flying altitudes are illustrated in Figure 5.10. With the optimized lighting configurations, new lighting rigs were designed and manufactured accordingly, allowing for quick assembly and disassembly for different mission conditions during cruises (see Figure 5.11).

# 6 Conclusion and Future Work

The deep sea, covering the majority of Earth's surface, remains largely uncharted due to its challenging and inaccessible nature. While mapping using image sensors is common on land and in space, the unique conditions of the deep sea present significant obstacles to traditional mapping methods. Navigating the challenges of deep sea imaging demands innovative solutions tailored to the harsh conditions of extreme pressure, refractive complexities, and light absorption and scattering. This dissertation addresses these challenges, providing nuanced solutions encompassing both in hardware and software aspects.

Robust housings with transparent windows shield deep sea cameras from saltwater and high pressures, ensuring their functionality in harsh environments. Geometric challenges posed by refractive interfaces are discussed through an in-depth analysis of dome refraction properties in Chapter 2. A camera behind a dome port housing holds the potential to physically eliminate refraction effects, contingent upon precise alignment with the dome's center. A novel and practical mechanical adjustment approach, complemented by visual feedback, for aligning the camera within the dome port housing was discussed. This method effectively mitigates refraction effects, thereby enabling the application of standard photogrammetry pipelines. Furthermore, the thesis presents an approach to calibrate any remaining decentering offset, that can be used in applications demanding high accuracy. Recognizing the continued use of flat port cameras in underwater imaging, this thesis also proposed an iterative refinement reconstruction method, employing a standard photogrammetry tool, even in the presence of flat refraction effects.

Addressing radiometric challenges related to light absorption and scattering, particularly in the context of artificial illumination, this thesis delves into the deep sea underwater image formation. In Chapter 3, the adapted J-M model is proposed. This model is not only applicable for simulating deep-sea images but can also be employed in various other applications, including parameter estimation and optimization tasks, such as optimizing lighting configurations for deep sea imaging systems.

Subsequently, in Chapter 4, the thesis explores various techniques for underwater image restoration. Introducing a general model based on a novel 3D lookup table structure, this model is designed to restore diverse types of underwater images. Specifically tailored for the intricacies of complex artificial illumination in deep sea environments, the approach simultaneously eliminates lighting patterns and restores true colors. This innovation bears significant advantages for applications in deep underwater visual mapping.

The thesis concludes with practical demonstrations of the developed techniques. The dome port centering and calibration techniques enhance the efficiency of various camera system designs, ensuring their suitability for accurate 3D underwater reconstruction. Utilizing deep sea image simulation tools, complex illumination configurations for a deep sea imaging system on an AUV are optimized, ultimately improving the quality of captured images. The thesis under-

scores the importance of high-accuracy underwater 3D reconstruction and mapping, exemplified through two representative cases showcasing the outcomes of 3D reconstruction in two different environment.

In summary, this dissertation contributes to overcoming the complexities of deep sea visual 3D reconstruction, providing valuable tools and techniques for advancing exploration, mapping, and understanding of the submerged world.

**Future Work** There are scenarios where refraction cannot be entirely avoided and must be considered during 3D reconstruction. Currently, the thesis proposes the solution by introducing an extra refraction correction step after each iteration of reconstruction. While this approach avoids modifying each step in the standard photogrammetry pipeline to adapt to refraction, it comes at the cost of a significant increase in processing time, as each iteration involves a complete 3D reconstruction from images. Ideally, considering refraction at each step of the reconstruction pipeline could yield better performance. Additionally, prior research often distinguishes between flat and dome refraction, providing different approaches to address these refraction issues. Exploring the refractive geometry further could unveil a potential unified refraction projection model applicable to any refraction cases.

Another interesting avenue for future research involves improving the image formation model for image restoration. While the current solution is able to handles underwater images under complex illumination conditions, estimating the entire lookup table requires numerous observations and constraints. An option could be to reduce the number of estimated parameters, as $\alpha$ and $\beta$ are currently treated as independent parameters, although there could be some dependence in certain settings. The modified model should still manage restoration challenges under complex illumination conditions but with fewer parameters. Moreover, exploring the application of the latest inverse rendering or Neural Radiance Field (NeRF) techniques (Mildenhall et al. 2021) could present another promising direction for restoring different types of underwater images.

# Bibliography


Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk (2012). "SLIC superpixels compared to state-of-the-art superpixel methods". In: *IEEE transactions on pattern analysis and machine intelligence* 34.11, pp. 2274–2282.

Agarwal, S., K. Mierle, and Others. *Ceres Solver*. http://ceres-solver.org.

Agrawal, A., S. Ramalingam, Y. Taguchi, and V. Chari (2012). "A theory of multi-layer flat refractive geometry". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 3346–3353.

Akkaynak, D. and T. Treibitz (2018). "A revised underwater image formation model". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6723–6732.

– (2019). "Sea-thru: A method for removing water from underwater images". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1682–1691.

Akkaynak, D., T. Treibitz, T. Shlesinger, Y. Loya, R. Tamir, and D. Iluz (2017). "What is the space of attenuation coefficients in underwater computer vision?" In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 568–577.

Allais, A., M. Bouhier, T. Edmond, M. Boffety, F. Galland, N. Maciol, S. Nicolas, M. Chami, and K. Ebert (2011). "SOFI: A 3D simulator for the generation of underwater optical images". In: *OCEANS 2011 IEEE - Spain*, pp. 1–6.

Álvarez-Tuñón, O., A. Jardón, and C. Balaguer (2019). "Generation and Processing of Simulated Underwater Images for Infrastructure Visual Inspection with UUVs". In: *Sensors* 19.24, p. 5497.

Ancuti, C. O., C. Ancuti, C. De Vleeschouwer, and P. Bekaert (2017a). "Color balance and fusion for underwater image enhancement". In: *IEEE Transactions on image processing* 27.1, pp. 379–393.

Ancuti, C. O., C. Ancuti, C. De Vleeschouwer, L. Neumann, and R. Garcia (2017b). "Color transfer for underwater dehazing and depth estimation". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 695–699.

Ancuti, C. O., C. Ancuti, C. Hermans, and P. Bekaert (2010). "A fast semi-inverse approach to detect and remove the haze from a single image". In: *Asian Conference on Computer Vision*. Springer, pp. 501–514.

Ancuti, C., C. O. Ancuti, C. De Vleeschouwer, R. Garcia, and A. C. Bovik (2016). "Multi-scale underwater descattering". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, pp. 4202–4207.

Ancuti, C., C. O. Ancuti, T. Haber, and P. Bekaert (2012). "Enhancing underwater images and videos by fusion". In: *2012 IEEE conference on computer vision and pattern recognition*. IEEE, pp. 81–88.

Arnaubec, A., J. Opderbecke, A.-G. Allais, and L. Brignone (2015). "Optical mapping with the ARIANE HROV at IFREMER: The MATISSE processing tool". In: *OCEANS 2015-Genova*. IEEE, pp. 1–6.

Baker, S., D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski (2011). "A database and evaluation methodology for optical flow". In: *International journal of computer vision* 92.1, pp. 1–31.

Ballard, R. D. (1975). "Photography from a submersible during project FAMOUS". In: *Oceanus* 18.3, pp. 40–43.

Ballard, R. D., R. Archbold, R. Atcher, and W. Lord (1987). *The discovery of the Titanic*. Warner Books New York, NY.

Ballard, R. D., L. E. Stager, D. Master, D. Yoerger, D. Mindell, L. L. Whitcomb, H. Singh, and D. Piechota (2002). "Iron age shipwrecks in deep water off Ashkelon, Israel". In: *American Journal of Archaeology*, pp. 151–168.

Barker, B. A., I. Helmond, N. J. Bax, A. Williams, S. Davenport, and V. A. Wadley (1999). "A vessel-towed camera platform for surveying seafloor habitats of the continental shelf". In: *Continental Shelf Research* 19.9, pp. 1161–1170. DOI: https://doi.org/10.1016/S0278-4343(99)00017-5.

Barros, W., E. R. Nascimento, W. V. Barbosa, and M. F. Campos (2018). "Single-shot underwater image restoration: A visual quality-aware method based on light propagation model". In: *Journal of Visual Communication and Image Representation* 55, pp. 363–373.

Bay, H., T. Tuytelaars, and L. Van Gool (2006). "Surf: Speeded up robust features". In: *European conference on computer vision*. Springer, pp. 404–417.

Bazeille, S., I. Quidu, L. Jaulin, and J.-P. Malkasse (2006). "Automatic underwater image preprocessing". In: *CMM'06*, p. xx.

Beigpour, S., C. Riess, J. Van De Weijer, and E. Angelopoulou (2013). "Multi-illuminant estimation with conditional random fields". In: *IEEE Transactions on Image Processing* 23.1, pp. 83–96.

Bekerman, Y., S. Avidan, and T. Treibitz (2020). "Unveiling optical properties in underwater images". In: *2020 IEEE International Conference on Computational Photography (ICCP)*. IEEE, pp. 1–12.

Berman, D., S. Avidan, et al. (2016). "Non-local image dehazing". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1674–1682.

Berman, D., D. Levy, S. Avidan, and T. Treibitz (2020). "Underwater single image color restoration using haze-lines and a new quantitative dataset". In: *IEEE transactions on pattern analysis and machine intelligence*.

Berman, D., T. Treibitz, and S. Avidan (2017). "Diving into haze-lines: Color restoration of underwater images". In: *Proc. British Machine Vision Conference (BMVC)*. Vol. 1. 2.

Bertalmio, M., A. L. Bertozzi, and G. Sapiro (2001). "Navier-stokes, fluid dynamics, and image and video inpainting". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. IEEE, pp. I–I.

Bingham, B., B. Foley, H. Singh, R. Camilli, K. Delaporta, R. Eustice, A. Mallios, D. Mindell, C. Roman, and D. Sakellariou (2010). "Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle". In: *Journal of Field Robotics* 27.6, pp. 702–717. DOI: https://doi.org/10.1002/rob.20350.

Bitterli, B. and W. Jarosz (2017). "Beyond points and beams: Higher-dimensional photon samples for volumetric light transport". In: *ACM Transactions on Graphics (TOG)* 36.4, pp. 1–12.

Blanco-Claraco, J.-L., F.-A. Moreno-Duenas, and J. González-Jiménez (2014). "The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario". In: *The International Journal of Robotics Research* 33.2, pp. 207–214.

Blasinski, H., J. Breneman, and J. Farrell (2014). "A model for estimating spectral properties of water from RGB images". In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 610–614.

Bleier, M., C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Tröger, and A. Kaup (2011). "Color constancy and non-uniform illumination: Can existing algorithms work?" In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, pp. 774–781.

Blender Online Community (2021). *Blender - a 3D modelling and rendering package*. Blender Foundation. URL: http://www.blender.org.

Bodenmann, A., B. Thornton, and T. Ura (2017). "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, pp. 833–851.

Bongiorno, D. L., M. Bryson, and S. B. Williams (2013). "Dynamic spectral-based underwater colour correction". In: *2013 MTS/IEEE OCEANS-Bergen*. IEEE, pp. 1–9.

Borgetto, M., V. Rigaud, and J.-F. Lots (2003). "Lighting correction for underwater mosaicking enhancement". In: *Proceedings of the 16th International Conference on Vision Interface*.

Bouguet, J.-Y. (2004). "Camera calibration toolbox for matlab". In: *http://www. vision. caltech. edu/bouguetj/calib_doc/*.

Bradski, G. (2000). "The openCV library." In: *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25.11, pp. 120–123.

Brown, M. and D. G. Lowe (2007). "Automatic Panoramic Image Stitching using Invariant Features". In: *International Journal of Computer Vision* 74.1, pp. 59–73.

Bryant, M., D. Wettergreen, S. Abdallah, A. Zelinsky, et al. (2000). "Robust camera calibration for an autonomous underwater vehicle". In: *Proc. Australian Conf. on Robotics and Autom*, pp. 111–116.

Bryson, M., M. Johnson-Roberson, O. Pizarro, and S. B. Williams (2016). "True color correction of autonomous underwater vehicle imagery". In: *Journal of Field Robotics* 33.6, pp. 853–874.

Bui, T. M. and W. Kim (2017). "Single image dehazing using color ellipsoid prior". In: *IEEE Transactions on Image Processing* 27.2, pp. 999–1009.

Burri, M., J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart (2016). "The EuRoC micro aerial vehicle datasets". In: *The International Journal of Robotics Research* 35.10, pp. 1157–1163.

Cai, B., X. Xu, K. Jia, C. Qing, and D. Tao (2016). "Dehazenet: An end-to-end system for single image haze removal". In: *IEEE Transactions on Image Processing* 25.11, pp. 5187–5198.

Cao, G., Y. Zhao, R. Ni, and X. Li (2014). "Contrast enhancement-based forensics in digital images". In: *IEEE transactions on information forensics and security* 9.3, pp. 515–525.

Carlevaris-Bianco, N., A. Mohan, and R. M. Eustice (2010). "Initial results in underwater single image dehazing". In: *Oceans 2010 Mts/IEEE Seattle*. IEEE, pp. 1–8.

Chandrasekhar, S. (2013). *Radiative transfer*. Courier Corporation.

Chao, L. and M. Wang (2010). "Removal of water scattering". In: *2010 2nd international conference on computer engineering and technology*. Vol. 2. IEEE, pp. V2–35.

Chari, V. and P. Sturm (2009). "Multiple-view geometry of the refractive plane". In: *BMVC 2009-20th British machine vision conference*. The British Machine Vision Association (BMVA), pp. 1–11.

Chen, Y., Z. Zeng, and Y. Pan (2019). "A New Degradation Model for Imaging in Natural Water and Validation Through Image Recovery". In: *IEEE Access* 7, pp. 123244–123254.

Cheng, C.-Y., C.-C. Sung, and H.-H. Chang (2015). "Underwater image restoration by red-dark channel prior and point spread function deconvolution". In: *2015 IEEE international conference on signal and image processing applications (ICSIPA)*. IEEE, pp. 110–115.

Chiang, J. Y. and Y.-C. Chen (2011). "Underwater image enhancement by wavelength compensation and dehazing". In: *IEEE transactions on image processing* 21.4, pp. 1756–1769.

Corrigan, D., K. Sooknanan, J. Doyle, C. Lordan, and A. Kokaram (2018). "A low-complexity mosaicing algorithm for stock assessment of seabed-burrowing species". In: *IEEE Journal of Oceanic Engineering* 44.2, pp. 386–400.

Cozman, F. and E. Krotkov (1997). "Depth from scattering". In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 801–806.

Crane, K., I. Llamas, and S. Tariq (2007). "Real-time simulation and rendering of 3d fluids". In: *GPU gems* 3.1.

Drap, P., J. Seinturier, B. Hijazi, D. Merad, J.-M. Boi, B. Chemisky, E. Seguin, and L. Long (2015). "The ROV 3D Project: Deep-sea underwater survey using photogrammetry: Applications for underwater archaeology". In: *Journal on Computing and Cultural Heritage (JOCCH)* 8.4, pp. 1–24.

Drews, P., E. Nascimento, F. Moraes, S. Botelho, and M. Campos (2013). "Transmission estimation in underwater single images". In: *Proceedings of the IEEE international conference on computer vision workshops*, pp. 825–830.

Drews, P., E. R. Nascimento, M. F. Campos, and A. Elfes (2015). "Automatic restoration of underwater monocular sequences of images". In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1058–1064.

Duane, C. B. (1971). "Close-range camera calibration". In: *Photogramm. Eng* 37.8, pp. 855–866.

Dubreuil, M, P Delrot, I. Leonard, A. Alfalou, C. Brosseau, and A. Dogariu (2013). "Exploring underwater target detection by imaging polarimetry and correlation techniques". In: *Applied optics* 52.5, pp. 997–1005.

Durrant-Whyte, H. and T. Bailey (2006). "Simultaneous localization and mapping: part I". In: *IEEE robotics & automation magazine* 13.2, pp. 99–110.

Eigen, D., C. Puhrsch, and R. Fergus (2014). "Depth map prediction from a single image using a multi-scale deep network". In: *arXiv preprint arXiv:1406.2283*.

Elnashef, B. and S. Filin (2022). "Geometry, calibration, and robust centering procedures for refractive dome-port based imaging systems". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 194, pp. 132–151.

Emberton, S., L. Chittka, and A. Cavallaro (2015). "Hierarchical rank-based veiling light estimation for underwater dehazing". In:

Escartín, J., R Garcia, O. Delaunoy, J Ferrer, N Gracias, A. Elibol, X Cufi, L. Neumann, D. Fornari, S. Humphris, et al. (2008). "Globally aligned photomosaic of the Lucky Strike hy-

drothermal vent field (Mid-Atlantic Ridge, 37 18.5 N): Release of georeferenced data, mosaic construction, and viewing software". In: *Geochemistry, Geophysics, Geosystems* 9.12.

Eustice, R., O. Pizarro, H. Singh, and J. Howland (2002). "UWIT: Underwater Image Toolbox for optical image processing and mosaicking in MATLAB". In: *Proceedings of the 2002 International Symposium on Underwater Technology (Cat. No. 02EX556)*. IEEE, pp. 141–145.

Eustice, R., H. Singh, and J. Howland (2000). "Image registration underwater for fluid flow measurements and mosaicking". In: *OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No. 00CH37158)*. Vol. 3. IEEE, pp. 1529–1534.

Fabbri, C., M. J. Islam, and J. Sattar (2018). "Enhancing underwater imagery using generative adversarial networks". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 7159–7165.

Farbman, Z., R. Fattal, D. Lischinski, and R. Szeliski (2008). "Edge-preserving decompositions for multi-scale tone and detail manipulation". In: *ACM Transactions on Graphics (TOG)* 27.3, pp. 1–10.

Fattal, R. (2008). "Single image dehazing". In: *ACM transactions on graphics (TOG)* 27.3, pp. 1–9.

– (2014). "Dehazing using color-lines". In: *ACM transactions on graphics (TOG)* 34.1, pp. 1–14.

Ferrera, M., V. Creuze, J. Moras, and P. Trouvé-Peloux (2019). "AQUALOC: An underwater dataset for visual–inertial–pressure localization". In: *The International Journal of Robotics Research* 38.14, pp. 1549–1559.

Finlayson, G. D., B. V. Funt, and K. Barnard (1995). "Color constancy under varying illumination". In: *Proceedings of IEEE International Conference on Computer Vision*. IEEE, pp. 720–725.

Foley, B. P., K. Dellaporta, D. Sakellariou, B. S. Bingham, R. Camilli, R. M. Eustice, D. Evagelistis, V. L. Ferrini, K. Katsaros, D. Kourkoumelis, et al. (2009). "The 2005 Chios ancient shipwreck survey: New methods for underwater archaeology". In: *Hesperia*, pp. 269–305.

Fornari, D. J. and W. T. Group (2003). "A new deep-sea towed digital camera and multi-rock coring system". In: *Eos, Transactions American Geophysical Union* 84.8, pp. 69–73. DOI: https://doi.org/10.1029/2003EO080001.

Fournier, G. R. and J. L. Forand (1994). "Analytic phase function for ocean water". In: *Ocean Optics XII*. Vol. 2258. International Society for Optics and Photonics, pp. 194–201.

Fournier, G. R. and M. Jonasz (1999). "Computer-based underwater imaging analysis". In: *Airborne and In-Water Underwater Imaging*. Vol. 3761. International Society for Optics and Photonics, pp. 62–70.

Fryer, J. and C. Fraser (1986). "On the calibration of underwater cameras". In: *The Photogrammetric Record* 12.67, pp. 73–85.

Fu, X., J. Huang, X. Ding, Y. Liao, and J. Paisley (2017). "Clearing the skies: A deep network architecture for single-image rain removal". In: *IEEE Transactions on Image Processing* 26.6, pp. 2944–2956.

Fu, X., P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding (2014). "A retinex-based enhancing approach for single underwater image". In: *2014 IEEE international conference on image processing (ICIP)*. IEEE, pp. 4572–4576.

Fujimura, Y., M. Iiyama, A. Hashimoto, and M. Minoh (2018). "Photometric Stereo in Participating Media Using an Analytical Solution for Shape-Dependent Forward Scatter". In: *IEEE transactions on pattern analysis and machine intelligence* 42.3, pp. 708–719.

Galdran, A. (2018). "Image dehazing by artificial multiple-exposure image fusion". In: *Signal Processing* 149, pp. 135–147.

Galdran, A., D. Pardo, A. Picón, and A. Alvarez-Gila (2015). "Automatic red-channel underwater image restoration". In: *Journal of Visual Communication and Image Representation* 26, pp. 132–145.

Garcia, R., T. Nicosevici, and X. Cufí (2002). "On the way to solve lighting problems in underwater imaging". In: *OCEANS'02 MTS/IEEE*. Vol. 2. IEEE, pp. 1018–1024.

Geiger, A., P. Lenz, and R. Urtasun (2012). "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Geiger, A., M. Roser, and R. Urtasun (2010). "Efficient large-scale stereo matching". In: *Asian conference on computer vision*. Springer, pp. 25–38.

Glaeser, G. and H.-P. Schröcker (2000). "Reflections on refractions". In: *Journal for Geometry and Graphics* 4.1, pp. 1–18.

Godard, C., O. Mac Aodha, M. Firman, and G. J. Brostow (2019). "Digging into self-supervised monocular depth estimation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3828–3838.

Gracias, N. and J. Santos-Victor (2000). "Underwater video mosaics as visual navigation maps". In: *Computer Vision and Image Understanding* 79.1, pp. 66–91.

Gracias, N. R., S. Van Der Zwaan, A. Bernardino, and J. Santos-Victor (2003). "Mosaic-based navigation for autonomous underwater vehicles". In: *IEEE Journal of Oceanic Engineering* 28.4, pp. 609–624.

Greinert, J. (2015). "RV SONNE Fahrtbericht/cruise report SO242-1 [SO242/1]: JPI OCEANS ecological aspects of deep-sea mining, DISCOL revisited, Guayaquil-Guayaquil (Equador), 28.07.-25.08. 2015". In:

Greinert, J., T. Schoening, K. Köser, and M. Rothenbeck (2017). *Seafloor images and raw context data along AUV tracks during SONNE cruises SO239 and SO242/1*. data set. GEOMAR - Helmholtz Centre for Ocean Research Kiel. DOI: 10.1594/PANGAEA.882349. URL: https://doi.org/10.1594/PANGAEA.882349.

Grossberg, M. D. and S. K. Nayar (2001). "A general imaging model and a method for finding its parameters". In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 2. IEEE, pp. 108–115.

– (2005). "The raxel imaging model and ray-based calibration". In: *International Journal of Computer Vision* 61, pp. 119–137.

Guillemaud, R. (1998). "Uniformity correction with homomorphic filtering on region of interest". In: *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269)*. Vol. 2. IEEE, pp. 872–875.

Gupta, H. and K. Mitra (2019). "Unsupervised single image underwater depth estimation". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 624–628.

Haltrin, V. I. (1999). "Two-term Henyey-Greenstein light scattering phase function for seawater". In: *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293)*. Vol. 2. IEEE, pp. 1423–1425.

– (2002). "One-parameter two-term Henyey-Greenstein phase function for light scattering in seawater". In: *Applied Optics* 41.6, pp. 1022–1028.
Han, P., F. Liu, K. Yang, J. Ma, J. Li, and X. Shao (2017). "Active underwater descattering and image recovery". In: *Applied optics* 56.23, pp. 6631–6638.
Hartley, R. and A. Zisserman (2004). *Multiple View Geometry in Computer Vision*. 2nd ed. Cambridge University Press. DOI: 10.1017/CBO9780511811685.
Harvey, E. N. (1939). "Deep sea photography". In: *Science* 90.2330, p. 187.
Harvey, E. S. and M. R. Shortis (1998). "Calibration stability of an underwater stereo-video system: implications for measurement accuracy and precision". In: *Marine Technology Society Journal* 32.2, pp. 3–17.
Hautière, N., J.-P. Tarel, and D. Aubert (2007). "Towards fog-free in-vehicle vision systems through contrast restoration". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1–8.
He, D.-M. and G. G. Seet (2004). "Divergent-beam lidar imaging in turbid water". In: *Optics and lasers in engineering* 41.1, pp. 217–231.
He, K., J. Sun, and X. Tang (2010). "Single image haze removal using dark channel prior". In: *IEEE transactions on pattern analysis and machine intelligence* 33.12, pp. 2341–2353.
– (2012). "Guided image filtering". In: *IEEE transactions on pattern analysis and machine intelligence* 35.6, pp. 1397–1409.
Henke, B., M. Vahl, and Z. Zhou (2013). "Removing color cast of underwater images through non-constant color constancy hypothesis". In: *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, pp. 20–24.
Hou, W., D. J. Gray, A. D. Weidemann, G. R. Fournier, and J. Forand (2007). "Automated underwater image restoration and retrieval of related optical properties". In: *2007 IEEE international geoscience and remote sensing symposium*. IEEE, pp. 1889–1892.
Hu, H., L. Zhao, X. Li, H. Wang, and T. Liu (2018). "Underwater image recovery under the nonuniform optical field based on polarimetric imaging". In: *IEEE Photonics Journal* 10.1, pp. 1–9.
Hu, W. and K. Qin (2006). "Interactive approximate rendering of reflections, refractions, and caustics". In: *IEEE Transactions on Visualization and Computer Graphics* 13.1, pp. 46–57.
Huang, B., T. Liu, H. Hu, J. Han, and M. Yu (2016). "Underwater image recovery considering polarization effects of objects". In: *Optics express* 24.9, pp. 9826–9838.
Iscar, E., C. Barbalata, N. Goumas, and M. Johnson-Roberson (2018). "Towards low cost, deep water AUV optical mapping". In: *OCEANS 2018 MTS/IEEE Charleston*. IEEE, pp. 1–6.
Jaffe, J. S. (1990). "Computer modeling and the design of optimal underwater imaging systems". In: *IEEE Journal of Oceanic Engineering* 15.2, pp. 101–111.
Jakob, W., S. Speierer, N. Roussel, M. Nimier-David, D. Vicini, T. Zeltner, B. Nicolet, M. Crespo, V. Leroy, and Z. Zhang (2022). *Mitsuba 3 renderer*. Version 3.1.1. https://mitsuba-renderer.org.
Jerlov, N. (1968). "Irradiance optical classification". In: *Optical Oceanography*, pp. 118–120.
Johnson, L. J. (2012). "The underwater optical channel". In: *Dept. Eng., Univ. Warwick, Coventry, UK, Tech. Rep*.
Johnson-Roberson, M., M. Bryson, A. Friedman, O. Pizarro, G. Troni, P. Ozog, and J. C. Henderson (2017). "High-resolution underwater robotic vision-based mapping and three-dimensional

reconstruction for archaeology". In: *Journal of Field Robotics* 34.4, pp. 625–643. DOI: https://doi.org/10.1002/rob.21658.

Johnson-Roberson, M., O. Pizarro, S. B. Williams, and I. Mahon (2010). "Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys". In: *Journal of Field Robotics* 27.1, pp. 21–51. DOI: https://doi.org/10.1002/rob.20324.

Jones, D. O., B. J. Bett, R. B. Wynn, and D. G. Masson (2009). "The use of towed camera platforms in deep-water science". In: *Underwater Technology* 28.2, pp. 41–50. DOI: https://doi.org/10.3723/ut.28.041.

Jordt, A., K. Köser, and R. Koch (2016). "Refractive 3D reconstruction on underwater images". In: *Methods in Oceanography* 15-16, pp. 90 –113. ISSN: 2211-1220. DOI: https://doi.org/10.1016/j.mio.2016.03.001. URL: http://www.sciencedirect.com/science/article/pii/S2211122015300086.

Jordt-Sedlazeck, A. and R. Koch (2012). "Refractive calibration of underwater cameras". In: *European conference on computer vision*. Springer, pp. 846–859.

– (2013). "Refractive structure-from-motion on underwater images". In: *Proceedings of the IEEE international Conference on Computer Vision*, pp. 57–64.

Kaeli, J. W., H. Singh, C. Murphy, and C. Kunz (2011). "Improving color correction for underwater image surveys". In: *OCEANS'11 MTS/IEEE KONA*. IEEE, pp. 1–6. DOI: https://doi.org/10.23919/OCEANS.2011.6107143.

Kang, L., L. Wu, and Y.-H. Yang (2012). "Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the svp camera model". In: *Applied optics* 51.31, pp. 7591–7603.

Kim, J.-H., W.-D. Jang, J.-Y. Sim, and C.-S. Kim (2013). "Optimized contrast enhancement for real-time image and video dehazing". In: *Journal of Visual Communication and Image Representation* 24.3, pp. 410–425.

Kimmel, R., M. Elad, D. Shaked, R. Keshet, and I. Sobel (2003). "A variational framework for retinex". In: *International Journal of computer vision* 52.1, pp. 7–23.

Kopf, J., B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski (2008). "Deep photo: Model-based photograph enhancement and viewing". In: *ACM transactions on graphics (TOG)* 27.5, pp. 1–10.

Koschmieder, H. (1924). "Theorie der horizontalen Sichtweite". In: *Beitrage zur Physik der freien Atmosphare*, pp. 33–53.

Köser, K., Y. Song, L. Petersen, E. Wenzlaff, and F. Woelk (2021). "Robustly Removing Deep Sea Lighting Effects for Visual Mapping of Abyssal Plains". In:

Kotowski, R. (1988). "Phototriangulation in multi-media photogrammetry". In: *International Archives of Photogrammetry and Remote Sensing* 27.B5, pp. 324–334.

Kratz, L. and K. Nishino (2009). "Factorizing scene albedo and depth from a single foggy image". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 1701–1708.

Kunz, C., C. Murphy, R. Camilli, H. Singh, J. Bailey, R. Eustice, M. Jakuba, K.-i. Nakamura, C. Roman, T. Sato, et al. (2008a). "Deep sea underwater robotic exploration in the ice-covered arctic ocean with AUVs". In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 3654–3660.

Kunz, C. and H. Singh (2008b). "Hemispherical refraction and camera calibration in underwater vision". In: *OCEANS 2008*. IEEE, pp. 1–7.

Kwasnitschka, T., K. Köser, J. Sticklus, M. Rothenbeck, T. Weiß, E. Wenzlaff, T. Schoening, L. Triebe, A. Steinführer, C. Devey, et al. (2016). "DeepSurveyCam—a deep ocean optical mapping system". In: *Sensors* 16.2, p. 164. DOI: https://doi.org/10.3390/s16020164.

Kölle, M., D. Laupheimer, S. Schmohl, N. Haala, F. Rottensteiner, J. D. Wegner, and H. Ledoux (2021). "The Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from UAV LiDAR and Multi-View-Stereo". In: *ISPRS Open Journal of Photogrammetry and Remote Sensing* 1, p. 11. ISSN: 2667-3932. DOI: https://doi.org/10.1016/j.ophoto.2021.100001.

Land, E. H. (1977). "The retinex theory of color vision". In: *Scientific american* 237.6, pp. 108–129.

Land, E. H. and J. J. McCann (1971). "Lightness and retinex theory". In: *Josa* 61.1, pp. 1–11.

Lavest, J.-M., G. Rives, and J.-T. Lapresté (2000). "Underwater camera calibration". In: *European Conference on Computer Vision*. Springer, pp. 654–668.

Lembke, C., S. Grasty, A. Silverman, H. Broadbent, S. Butcher, and S. Murawski (2017). "The Camera-Based Assessment Survey System (C-BASS): A towed camera platform for reef fish abundance surveys and benthic habitat characterization in the Gulf of Mexico". In: *Continental Shelf Research* 151, pp. 62–71. DOI: https://doi.org/10.1016/j.csr.2017.10.010.

Li, C. Y. and A. Cavallaro (2018a). "Background light estimation for depth-dependent underwater image restoration". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 1528–1532.

Li, C.-Y., J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang (2016a). "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior". In: *IEEE Transactions on Image Processing* 25.12, pp. 5664–5677.

Li, C., S. Anwar, and F. Porikli (2020). "Underwater scene prior inspired deep underwater image and video enhancement". In: *Pattern Recognition* 98, p. 107038.

Li, C., C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao (2019). "An underwater image enhancement benchmark dataset and beyond". In: *IEEE Transactions on Image Processing* 29, pp. 4376–4389.

Li, C., J. Guo, C. Guo, R. Cong, and J. Gong (2017a). "A hybrid method for underwater image correction". In: *Pattern Recognition Letters* 94, pp. 62–67.

Li, C., J. Quo, Y. Pang, S. Chen, and J. Wang (2016b). "Single underwater image restoration by blue-green channels dehazing and red channel correction". In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 1731–1735.

Li, H., X. Wang, T. Bai, W. Jin, Y. Huang, and K. Ding (2009). "Speckle noise suppression of range gated underwater imaging system". In: *Applications of Digital Image Processing XXXII*. Vol. 7443. International Society for Optics and Photonics, 74432A.

Li, J., K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson (2017b). "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images". In: *IEEE Robotics and Automation letters* 3.1, pp. 387–394.

Li, Z. and N. Snavely (2018b). "Megadepth: Learning single-view depth prediction from internet photos". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2041–2050.

Lirman, D., N. R. Gracias, B. E. Gintert, A. C. R. Gleason, R. P. Reid, S. Negahdaripour, and P. Kramer (2007). "Development and application of a video-mosaic survey technology to doc-

ument the status of coral reef communities". In: *Environmental monitoring and assessment* 125.1, pp. 59–73.

Liu, Q., X. Gao, L. He, and W. Lu (2018a). "Single image dehazing with depth-aware non-local total variation regularization". In: *IEEE Transactions on Image Processing* 27.10, pp. 5178–5191.

Liu, R., M. Hou, J. Liu, X. Fan, and Z. Luo (2019). "Compounded Layer-Prior Unrolling: A Unified Transmission-Based Image Enhancement Framework". In: *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, pp. 538–543.

Liu, R., L. Ma, Y. Wang, and L. Zhang (2018b). "Learning converged propagations with deep prior ensemble for image enhancement". In: *IEEE Transactions on Image Processing* 28.3, pp. 1528–1543.

Liu, Z., Y. Yu, K. Zhang, and H. Huang (2001). "Underwater image transmission and blurred image restoration". In: *Optical Engineering* 40.6, pp. 1125–1131.

Lowe, D. G. (2004). "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2, pp. 91–110.

Lu, H., Y. Li, and S. Serikawa (2013). "Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction". In: *2013 IEEE international conference on image processing*. IEEE, pp. 3412–3416.

Lu, H., Y. Li, L. Zhang, and S. Serikawa (2015). "Contrast enhancement for images in turbid water". In: *JOSA A* 32.5, pp. 886–893.

Lu, J., F. Yuan, W. Yang, and E. Cheng (2021). "An Imaging Information Estimation Network for Underwater Image Color Restoration". In: *IEEE Journal of Oceanic Engineering*.

Łuczyński, T. and A. Birk (2017a). "Underwater image haze removal with an underwater-ready dark channel prior". In: *OCEANS 2017-Anchorage*. IEEE, pp. 1–6.

Łuczyński, T., M. Pfingsthorn, and A. Birk (2017b). "The pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings". In: *Ocean Engineering* 133, pp. 9–22.

Ludvigsen, M., B. Sortland, G. Johnsen, and H. Singh (2007). "Applications of geo-referenced underwater photo mosaics in marine biology and archaeology". In: *Oceanography* 20.4, pp. 140–149.

Maas, H.-G. (1995). "New developments in multimedia photogrammetry". In: *Optical 3D measurement techniques III*.

Mallios, A., E. Vidal, R. Campos, and M. Carreras (2017). "Underwater caves sonar data set". In: *The International Journal of Robotics Research* 36.12, pp. 1247–1251.

Mangeruga, M., M. Cozza, and F. Bruno (2018). "Evaluation of underwater image enhancement algorithms under different environmental conditions". In: *Journal of Marine Science and Engineering* 6.1, p. 10.

Manhães, M. M. M., S. A. Scherer, M. Voss, L. R. Douat, and T. Rauschenbach (2016). "UUV Simulator: A Gazebo-based package for underwater intervention and multi-robot simulation". In: *OCEANS 2016 MTS/IEEE Monterey*. IEEE. DOI: 10.1109/oceans.2016.7761080. URL: https://doi.org/10.1109%2Foceans.2016.7761080.

Marks, R. L., S. M. Rock, and M. J. Lee (1995). "Real-time video mosaicking of the ocean floor". In: *IEEE Journal of Oceanic Engineering* 20.3, pp. 229–241.

Mathias, A. and D. Samiappan (2019). "Underwater image restoration based on diffraction bounded optimization algorithm with dark channel prior". In: *Optik* 192, p. 162925.

Maybank, S. J. and O. D. Faugeras (1992). "A theory of self-calibration of a moving camera". In: *International journal of computer vision* 8.2, pp. 123–151.

McGlamery, B. (1975). "Computer analysis and simulation of underwater camera system performance". In: *SIO ref* 75, p. 2.

– (1980). "A computer model for underwater camera systems". In: *Ocean Optics VI*. Vol. 208. International Society for Optics and Photonics, pp. 221–231.

McGlone, J. C., ed. (2004). *Manual of Photogrammetry*. 5th. ASPRS. Chap. 2, pp. 98–102.

Menna, F., E. Nocerino, F. Fassi, and F. Remondino (2016). "Geometric and Optic Characterization of a Hemispherical Dome Port for Underwater Photogrammetry". In: *Sensors* 16.1. ISSN: 1424-8220. DOI: https://doi.org/10.3390/s16010048.

Menna, F., E. Nocerino, and F. Remondino (2017). "Optical aberrations in underwater photogrammetry with flat and hemispherical dome ports". In: *Videometrics, Range Imaging, and Applications XIV*. Vol. 10332. International Society for Optics and Photonics, p. 1033205.

Mertens, L. E. and F. S. Replogle (1977). "Use of point spread and beam spread functions for analysis of imaging systems in water". In: *JOSA* 67.8, pp. 1105–1117.

Mie, G. (1976). "Contributions to the optics of turbid media, particularly of colloidal metal solutions". In: *Contributions to the optics of turbid media* 25.3, pp. 377–445.

Mildenhall, B., P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng (2021). "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1, pp. 99–106.

Mobley, C., E. Boss, and C. Roesler (2010). "Ocean optics web book". In: *URL http://www.oceanopticsbook.info* 740, pp. 7484–7504.

Mobley, C. D. (1994). *Light and water: radiative transfer in natural waters*. Academic press.

Möller, T. and B. Trumbore (2005). "Fast, minimum storage ray/triangle intersection". In: *ACM SIGGRAPH 2005 Courses*, 7–es.

Murez, Z., T. Treibitz, R. Ramamoorthi, and D. Kriegman (2015). "Photometric stereo in a scattering medium". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3415–3423.

Nakath, D., M. She, Y. Song, and K. Köser (2021). "In-Situ Joint Light and Medium Estimation for Underwater Color Restoration". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. IEEE, pp. 0–0.

Narasimhan, S. G., M. Gupta, C. Donner, R. Ramamoorthi, S. K. Nayar, and H. W. Jensen (2006). "Acquiring scattering properties of participating media by dilution". In: *ACM SIGGRAPH 2006 Papers*, pp. 1003–1012.

Narasimhan, S. G. and S. K. Nayar (2002). "Vision and the atmosphere". In: *International journal of computer vision* 48.3, pp. 233–254.

Narasimhan, S. G. and S. K. Nayar (2003a). "Contrast restoration of weather degraded images". In: *IEEE transactions on pattern analysis and machine intelligence* 25.6, pp. 713–724.

Narasimhan, S. G. and S. K. Nayar (2003b). "Interactive (de) weathering of an image using physical models". In: *IEEE Workshop on color and photometric Methods in computer Vision*. Vol. 6. 6.4. France, p. 1.

Narasimhan, S. G. and S. K. Nayar (2005a). "Structured light methods for underwater imaging: light stripe scanning and photometric stereo". In: *Proceedings of OCEANS 2005 MTS/IEEE*. IEEE, pp. 2610–2617.

Narasimhan, S. G., S. K. Nayar, B. Sun, and S. J. Koppal (2005b). "Structured light in scattering media". In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Vol. 1. IEEE, pp. 420–427.

Nayar, S. K. and S. G. Narasimhan (1999). "Vision in bad weather". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. IEEE, pp. 820–827.

Negahdaripour, S., H Zhang, and X Han (2002). "Investigation of photometric stereo method for 3-d shape recovery from underwater imagery". In: *OCEANS'02 MTS/IEEE*. Vol. 2. IEEE, pp. 1010–1017.

Nex, F, F Remondino, M Gerke, H.-J. Przybilla, M Bäumker, and A Zurhorst (2015). "ISPRS BENCHMARK FOR MULTI-PLATFORM PHOTOGRAMMETRY." In: *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 2.

Nocerino, E., F. Menna, and A. Grün (2021). "Bundle adjustment with polynomial point-to-camera distance dependent corrections for underwater photogrammetry". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 43, pp. 673–679.

Novák, J., I. Georgiev, J. Hanika, and W. Jarosz (2018). "Monte Carlo methods for volumetric light transport simulation". In: *Computer Graphics Forum*. Vol. 37. 2. Wiley Online Library, pp. 551–576.

Park, D., H. Park, D. K. Han, and H. Ko (2014a). "Single image dehazing with image entropy and information fidelity". In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 4037–4041.

Park, J., S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. So Kweon (2014b). "Calibrating a non-isotropic near point light source using a plane". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2259–2266.

Pegoraro, V., M. Schott, and S. G. Parker (2010). "A Closed-Form Solution to Single Scattering for General Phase Functions and Light Distributions". In: *Computer Graphics Forum*. Vol. 29. 4. Wiley Online Library, pp. 1365–1374.

Peng, Y.-T., K. Cao, and P. C. Cosman (2018). "Generalization of the dark channel prior for single image restoration". In: *IEEE Transactions on Image Processing* 27.6, pp. 2856–2868.

Peng, Y.-T. and P. C. Cosman (2017). "Underwater image restoration based on image blurriness and light absorption". In: *IEEE transactions on image processing* 26.4, pp. 1579–1594.

Peng, Y.-T., X. Zhao, and P. C. Cosman (2015). "Single underwater image enhancement using depth estimation based on blurriness". In: *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 4952–4956.

Petersen, L. (2018). "Beschleunigung von Algorithmen zur Meeresbodenkartierung durch Parallelisierung auf einer GPU".

Petzold, T. J. (1972). *Volume scattering functions for selected ocean waters*. Tech. rep. Scripps Institution of Oceanography La Jolla Ca Visibility Lab.

Peukert, A., T. Schoening, E. Alevizos, K. Köser, T. Kwasnitschka, and J. Greinert (2018). "Understanding Mn-nodule distribution and evaluation of related deep-sea mining impacts using AUV-based hydroacoustic and optical data". In: *Biogeosciences* 15.8, pp. 2525–2549.

Phong, B. T. (1998). "Illumination for computer generated pictures". In: *Seminal graphics: pioneering efforts that shaped the field*, pp. 95–101.

Pillai, S., R. Ambruş, and A. Gaidon (2019). "Superdepth: Self-supervised, super-resolved monocular depth estimation". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 9250–9256.

Pizarro, O. and H. Singh (2003). "Toward large-area mosaicing for underwater scientific applications". In: *IEEE journal of oceanic engineering* 28.4, pp. 651–672.

Pizer, S. M., E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld (1987). "Adaptive histogram equalization and its variations". In: *Computer vision, graphics, and image processing* 39.3, pp. 355–368.

Prados, R., R. Garcia, N. Gracias, J. Escartin, and L. Neumann (2012). "A novel blending technique for underwater gigamosaicing". In: *IEEE Journal of Oceanic Engineering* 37.4, pp. 626–644.

Prats, M., J. Perez, J. J. Fernández, and P. J. Sanz (2012). "An open source tool for simulation and supervision of underwater intervention missions". In: *2012 IEEE/RSJ international conference on Intelligent Robots and Systems*. IEEE, pp. 2577–2582.

Preisendorfer, R. (1964). "Physical aspect of light in the sea". In: *Univ. Hawai. Press. Honolulu Hawaii 51* 60.

Purser, A., Y. Marcon, S. Dreutter, U. Hoge, B. Sablotny, L. Hehemann, J. Lemburg, B. Dorschel, H. Biebow, and A. Boetius (2018). "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, pp. 87–99. DOI: https://doi.org/10.1109/JOE.2018.2794095.

Queiroz-Neto, J. P., R. Carceroni, W. Barros, and M. Campos (2004). "Underwater stereo". In: *Proceedings. 17th Brazilian Symposium on Computer Graphics and Image Processing*. IEEE, pp. 170–177.

Rahman, Z.-u., D. J. Jobson, and G. A. Woodell (2004). "Retinex processing for automatic image enhancement". In: *Journal of Electronic imaging* 13.1, pp. 100–110.

Ramalingam, S., S. K. Lodha, and P. Sturm (2006). "A generic structure-from-motion framework". In: *Computer Vision and Image Understanding* 103.3, pp. 218–228.

Ren, W., L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang (2018). "Gated fusion network for single image dehazing". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3253–3261.

Roser, M., M. Dunbabin, and A. Geiger (2014). "Simultaneous underwater visibility assessment, enhancement and improved stereo". In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 3840–3847.

Rzhanov, Y, L. M. Linnett, and R Forbes (2000). "Underwater video mosaicing for seabed mapping". In: *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*. Vol. 1. IEEE, pp. 224–227.

Sarafraz, A. and B. K. Haus (2016). "A structured light method for underwater surface reconstruction". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 114, pp. 40–52.

Scharstein, D., H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling (2014). "High-resolution stereo datasets with subpixel-accurate ground truth". In: *German conference on pattern recognition*. Springer, pp. 31–42.

Scharstein, D. and R. Szeliski (2003). "High-accuracy stereo depth maps using structured light". In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* Vol. 1. IEEE, pp. I–I.

Schechner, Y. Y. and Y. Averbuch (2007). "Regularized image recovery in scattering media". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.9, pp. 1655–1660.

Schechner, Y. Y. and N. Karpel (2004). "Clear underwater vision". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1. IEEE, pp. I–I.

– (2005). "Recovery of underwater visibility and structure by polarization analysis". In: *IEEE Journal of oceanic engineering* 30.3, pp. 570–587.

Schechner, Y. Y., S. G. Narasimhan, and S. K. Nayar (2001). "Instant dehazing of images using polarization". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.* Vol. 1. IEEE, pp. I–I.

– (2003). "Polarization-based vision through haze". In: *Applied optics* 42.3, pp. 511–525.

Schops, T., T. Sattler, and M. Pollefeys (2019). "Bad slam: Bundle adjusted direct rgb-d slam". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 134–144.

Schops, T., J. L. Schonberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, and A. Geiger (2017). "A multi-view stereo benchmark with high-resolution images and multi-camera videos". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3260–3269.

Sedlazeck, A. and R. Koch (2011). "Simulating deep sea underwater images using physical models for light attenuation, scattering, and refraction". In:

Sedlazeck, A., K. Köser, and R. Koch (2009). "3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions". In: *OCEANS 2009-EUROPE*. IEEE, pp. 1–10.

Seitz, S. M., B. Curless, J. Diebel, D. Scharstein, and R. Szeliski (2006). "A comparison and evaluation of multi-view stereo reconstruction algorithms". In: *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*. Vol. 1. IEEE, pp. 519–528.

Serikawa, S. and H. Lu (2014). "Underwater image dehazing using joint trilateral filter". In: *Computers & Electrical Engineering* 40.1, pp. 41–50.

She, M., D. Nakath, Y. Song, and K. Köser (2022a). "Refractive geometry for underwater domes". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 183, pp. 525–540.

She, M., Y. Song, J. Mohrmann, and K. Köser (2019). "Adjustment and Calibration of Dome Port Camera Systems for Underwater Vision". In: *German Conference on Pattern Recognition*. Springer, pp. 79–92. DOI: https://doi.org/10.1007/978-3-030-33676-9_6.

She, M., T. Weiß, Y. Song, P. Urban, J. Greinert, and K. Köser (2022b). "Marine bubble flow quantification using wide-baseline stereo photogrammetry". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 190, pp. 322–341.

Shortis, M. (2015). "Calibration techniques for accurate measurements by underwater camera systems". In: *Sensors* 15.12, pp. 30810–30826.

Shortis, M. R., E. S. Harvey, and D. A. Abdo (2009). "A review of underwater stereo-image measurement for marine biology and ecology applications". In: *Oceanography and marine biology: an annual review* 47, pp. 269–304.

Shukla, A. and H. Karki (2016). "Application of robotics in offshore oil and gas industry—A review Part II". In: *Robotics and Autonomous Systems* 75, pp. 508–524.

Shwartz, S. and Y. Schechner (2006). "Blind haze separation". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. IEEE, pp. 1984–1991.

Sibson, R. (1981). "A brief description of natural neighbour interpolation". In: *Interpreting multivariate data*, pp. 21–36.

Simon-Lledó, E., B. J. Bett, V. A. Huvenne, K. Köser, T. Schoening, J. Greinert, and D. O. Jones (2019). "Biological effects 26 years after simulated deep-sea mining". In: *Scientific reports* 9.1, pp. 1–13.

Singh, H., R. Armstrong, F. Gilbes, R. Eustice, C. Roman, O. Pizarro, and J. Torres (2004a). "Imaging coral I: imaging coral habitats with the SeaBED AUV". In: *Subsurface Sensing Technologies and Applications* 5.1, pp. 25–42.

Singh, H., A. Can, R. Eustice, S. Lerner, N. McPhee, and C. Roman (2004b). "Seabed AUV offers new platform for high-resolution imaging". In: *Eos, Transactions American Geophysical Union* 85.31, pp. 289–296.

Singh, H., J. Howland, and O. Pizarro (2004c). "Advances in large-area photomosaicking underwater". In: *IEEE Journal of Oceanic Engineering* 29.3, pp. 872–886. DOI: https://doi.org/10.1109/JOE.2004.831619.

Singh, H., J. Howland, D. Yoerger, and L. Whitcomb (1998). "Quantitative photomosaicking of underwater imagery". In: *IEEE Oceanic Engineering Society. OCEANS'98. Conference Proceedings (Cat. No. 98CH36259)*. Vol. 1. IEEE, pp. 263–266.

Singh, H., C. Roman, O. Pizarro, R. Eustice, and A. Can (2007). "Towards high-resolution imaging from underwater vehicles". In: *The International journal of robotics research* 26.1, pp. 55–74.

Skinner, K. A. and M. Johnson-Roberson (2017). "Underwater image dehazing with a light field camera". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 62–69.

Solonenko, M. G. and C. D. Mobley (2015). "Inherent optical properties of Jerlov water types". In: *Applied optics* 54.17, pp. 5392–5401.

Song, Y., K. Köser, T. Kwasnitschka, and R. Koch (2019). "Iterative refinement for underwater 3d reconstruction: Application to disposed underwater munitions in the baltic sea". In: *ISPRS-international archives of the photogrammetry, remote sensing and spatial information sciences* 42, pp. 181–187.

Song, Y., D. Nakath, M. She, F. Elibol, and K. Köser (2021a). "Deep sea robotic imaging simulator". In: *Pattern Recognition. ICPR International Workshops and Challenges*. Cham: Springer International Publishing, pp. 375–389. DOI: https://doi.org/10.1007/978-3-030-68790-8_29.

Song, Y., D. Nakath, M. She, and K. Köser (2022a). "Optical imaging and image restoration techniques for deep ocean mapping: a comprehensive survey". In: *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 90.3, pp. 243–267.

Song, Y., M. She, and K. Köser (2022b). "Virtually throwing benchmarks into the ocean for deep sea photogrammetry and image processing evaluation". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4, pp. 353–360.

Song, Y., M. She, and K. Köser (2024). "Advanced Underwater Image Restoration in Complex Illumination Conditions". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 209, pp. 197–212. DOI: https://doi.org/10.1016/j.isprsjprs.2024.02.004.

Song, Y., J. Sticklus, D. Nakath, E. Wenzlaff, R. Koch, and K. Köser (2021b). "Optimization of Multi-LED Setups for Underwater Robotic Vision Systems". In: *Pattern Recognition. ICPR International Workshops and Challenges*. Cham: Springer International Publishing, pp. 390–397. DOI: https://doi.org/10.1007/978-3-030-68790-8_30.

Spier, O., T. Treibitz, and G. Gilboa (2017). "In situ target-less calibration of turbid media". In: *2017 IEEE International Conference on Computational Photography (ICCP)*. IEEE, pp. 1–9.

Streckel, B. and R. Koch (2005). "Lens model selection for visual tracking". In: *Joint Pattern Recognition Symposium*. Springer, pp. 41–48.

Strutt, J. W. (1871). "XV. On the light from the sky, its polarization and colour". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41.271, pp. 107–120.

Sturm, J., N. Engelhard, F. Endres, W. Burgard, and D. Cremers (2012). "A benchmark for the evaluation of RGB-D SLAM systems". In: *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, pp. 573–580.

Sturm, P. and S. Ramalingam (2004). "A generic concept for camera calibration". In: *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part II 8*. Springer, pp. 1–13.

Tan, C., G. Seet, A. Sluzek, and D. He (2005). "A novel application of range-gated underwater laser imaging system (ULIS) in near-target turbid medium". In: *Optics and lasers in engineering* 43.9, pp. 995–1009.

Tan, C., A Sluzek, G. S. GL, and T. Jiang (2006). "Range gated imaging system for underwater robotic vehicle". In: *OCEANS 2006-Asia Pacific*. IEEE, pp. 1–6.

Tan, R. T. (2008). "Visibility in bad weather from a single image". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1–8.

Tao, M. W., S. Hadap, J. Malik, and R. Ramamoorthi (2013). "Depth from combining defocus and correspondence using light-field cameras". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 673–680.

Tarel, J.-P. and N. Hautiere (2009). "Fast visibility restoration from a single color or gray level image". In: *2009 IEEE 12th international conference on computer vision*. IEEE, pp. 2201–2208.

Telem, G. and S. Filin (2010). "Photogrammetric modeling of underwater environments". In: *ISPRS journal of photogrammetry and remote sensing* 65.5, pp. 433–444.

Tian, J., Z. Murez, T. Cui, Z. Zhang, D. Kriegman, and R. Ramamoorthi (2017). "Depth and image restoration from light field in a scattering medium". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2401–2410.

Torres-Méndez, L. A. and G. Dudek (2005). "Color correction of underwater images for aquatic robot inspection". In: *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, pp. 60–73.

Treibitz, T., Y. Schechner, C. Kunz, and H. Singh (2011). "Flat refractive geometry". In: *IEEE transactions on pattern analysis and machine intelligence* 34.1, pp. 51–65.

Treibitz, T. and Y. Y. Schechner (2006). "Instant 3descatter". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. IEEE, pp. 1861–1868.

– (2008). “Active polarization descattering”. In: *IEEE transactions on pattern analysis and machine intelligence* 31.3, pp. 385–399.

Trucco, E. and A. T. Olmos-Antillon (2006). “Self-tuning underwater image restoration”. In: *IEEE Journal of Oceanic Engineering* 31.2, pp. 511–519.

Tsiotsios, C., M. E. Angelopoulou, T.-K. Kim, and A. J. Davison (2014). “Backscatter compensated photometric stereo with 3 sources”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2251–2258.

Ueda, T., K. Yamada, and Y. Tanaka (2019). “Underwater Image Synthesis from RGB-D Images and its Application to Deep Underwater Image Restoration”. In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 2115–2119.

Vincent, A., N. Pessel, M. Borgetto, J. Jouffroy, J. Opderbecke, and V. Rigaud (2003). “Real-time geo-referenced video mosaicking with the MATISSE system”. In: *Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492)*. Vol. 4, 2319–2324 Vol.4. DOI: https://doi.org/10.1109/OCEANS.2003.178271.

Wang, G., B. Zheng, and F. F. Sun (2011). “Estimation-based approach for underwater image restoration”. In: *Optics letters* 36.13, pp. 2384–2386.

Wang, K., E. Dunn, J. Tighe, and J.-M. Frahm (2014). “Combining semantic scene priors and haze removal for single image depth estimation”. In: *IEEE Winter Conference on Applications of Computer Vision*. IEEE, pp. 800–807.

Wang, T.-C., A. A. Efros, and R. Ramamoorthi (2015). “Occlusion-aware depth estimation using light-field cameras”. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3487–3495.

Wang, Y., W. Song, G. Fortino, L.-Z. Qi, W. Zhang, and A. Liotta (2019). “An experimental-based review of image enhancement and image restoration methods for underwater imaging”. In: *IEEE access* 7, pp. 140233–140251.

Wang, Y., H. Liu, and L.-P. Chau (2017a). “Single underwater image restoration using adaptive attenuation-curve prior”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 65.3, pp. 992–1002.

– (2017b). “Single underwater image restoration using attenuation-curve prior”. In: *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, pp. 1–4.

Weber, M. and R. Cipolla (2001). “A Practical Method for Estimation of Point Light-Sources.” In: *BMVC*. Vol. 2001. 2, pp. 471–480.

Wen, H., Y. Tian, T. Huang, and W. Gao (2013). “Single underwater image enhancement with a new optical model”. In: *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, pp. 753–756.

Winters, G, R Holzman, A Blekhman, S Beer, and Y Loya (2009). “Photographic assessment of coral chlorophyll contents: implications for ecophysiological studies and coral monitoring”. In: *Journal of Experimental Marine Biology and Ecology* 380.1-2, pp. 25–35.

Wolff, K. (2007). “Zur Approximation allgemeiner optischer Abbildungsmodelle und deren Anwendung auf eine geometrisch basierte Mehrbildzuordnung am Beispiel einer Mehrmedienabbildung”. PhD thesis. Universitäts-und Landesbibliothek Bonn.

Wu, M., K. Luo, J. Dang, and D. Li (2017). “Underwater image restoration using color correction and non-local prior”. In: *OCEANS 2017-Aberdeen*. IEEE, pp. 1–5.

Wyman, C. (2005). "An approximate image-space approach for interactive refraction". In: *ACM transactions on graphics (TOG)* 24.3, pp. 1050–1053.

Xiao, C. and J. Gan (2012). "Fast image dehazing using guided joint bilateral filter". In: *The Visual Computer* 28.6, pp. 713–721.

Yang, H.-Y., P.-Y. Chen, C.-C. Huang, Y.-Z. Zhuang, and Y.-H. Shiau (2011). "Low complexity underwater image enhancement based on dark channel prior". In: *2011 Second International Conference on Innovations in Bio-inspired Computing and Applications*. IEEE, pp. 17–20.

Yang, M., A. Sowmya, Z. Wei, and B. Zheng (2019). "Offshore underwater image restoration using reflection-decomposition-based transmission map estimation". In: *IEEE Journal of Oceanic Engineering* 45.2, pp. 521–533.

Yoerger, D. R., M. Jakuba, A. M. Bradley, and B. Bingham (2007). "Techniques for deep sea near bottom survey using an autonomous underwater vehicle". In: *The International Journal of Robotics Research* 26.1, pp. 41–54.

Yoerger, D. R., D. S. Kelley, and J. R. Delaney (2000). "Fine-scale three-dimensional mapping of a deep-sea hydrothermal vent site using the Jason ROV system". In: *The International Journal of Robotics Research* 19.11, pp. 1000–1014.

Yu, X., Y. Qu, and M. Hong (2018). "Underwater-GAN: Underwater image restoration via conditional generative adversarial network". In: *International Conference on Pattern Recognition*. Springer, pp. 66–75.

Zhang, H. (2016). "Removing backscatter to enhance the visibility of underwater object". MA thesis.

Zhang, K., W. Zuo, S. Gu, and L. Zhang (2017a). "Learning deep CNN denoiser prior for image restoration". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3929–3938.

Zhang, S., T. Wang, J. Dong, and H. Yu (2017b). "Underwater image enhancement via extended multi-scale Retinex". In: *Neurocomputing* 245, pp. 1–9.

Zhang, W., P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li (2022). "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement". In: *IEEE Transactions on Image Processing* 31, pp. 3997–4010.

Zhang, Z. (2000). "A flexible new technique for camera calibration". In: *IEEE Transactions on pattern analysis and machine intelligence* 22.11, pp. 1330–1334.

Zhang, Z., H. Rebecq, C. Forster, and D. Scaramuzza (2016). "Benefit of large field-of-view cameras for visual odometry". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 801–808.

Zhao, X., T. Jin, and S. Qu (2015). "Deriving inherent optical properties from background color and underwater image enhancement". In: *Ocean Engineering* 94, pp. 163–172.

Zhou, J., Z. Liu, W. Zhang, D. Zhang, and W. Zhang (2021a). "Underwater image restoration based on secondary guided transmission map". In: *Multimedia Tools and Applications* 80.5, pp. 7771–7788.

Zhou, J., T. Yang, W. Ren, D. Zhang, and W. Zhang (2021b). "Underwater image restoration via depth map and illumination estimation based on a single image". In: *Optics Express* 29.19, pp. 29864–29886.

Zhou, Y., Q. Wu, K. Yan, L. Feng, and W. Xiang (2018). "Underwater image restoration using color-line model". In: *IEEE Transactions on Circuits and Systems for Video Technology* 29.3, pp. 907–911.

Zhu, Q., J. Mai, and L. Shao (2015). "A fast single image haze removal algorithm using color attenuation prior". In: *IEEE transactions on image processing* 24.11, pp. 3522–3533.

Zuiderveld, K. (1994). "Contrast limited adaptive histogram equalization". In: *Graphics gems IV*, pp. 474–485.

Zwilgmeyer, P. G. O., M. Yip, A. L. Teigen, R. Mester, and A. Stahl (2021). "The VAROS Synthetic Underwater Data Set: Towards realistic multi-sensor underwater data with ground truth". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3722–3730.