Underwater Visual Multi-Modal 3D Sensing

Alexander Duda
Fachbereich 3 (Mathematik und Informatik)
Universität Bremen

Dissertation zur Erlangung des Grades eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
November 2019

Gutachter:
Prof. Dr.-Ing. Udo Frese
Dr.-Ing. Kevin Köser

Datum des Kolloquiums:
23. April 2020

Weitere Mitglieder des Prüfungsausschusses:
Prof. Dr. Dr. h.c. Frank Kirchner
Dr. rer. nat. Tom Kwasnitschka
Tom Lucas Koller
Fabian Hauschildt

Acknowledgements

First, I would very much like to express my deepest gratitude to my supervisor Prof. Udo Frese for his encouraging support and his way to always find time for discussions. Based on his broad experience in machine vision and multi-sensor fusion, he was excellent guidance for me and the resulting thesis.

I would also like to especially thank Prof. Frank Kirchner for introducing me to the underwater domain and his excellent supervision while working with the Robotics Innovation Center (RIC) at the German Research Center for Artificial Intelligence (DFKI). Only because of his constant support for me and my work during many years of exciting project work, it was possible to start working towards this thesis before founding a spin-off focusing on the commercializing of underwater laser systems. Therefore, I would also very much thank my former and new colleagues Jakob Schwendner, Patrick Paranhos, and Jan Albiez, for their help and their advice during this transition and beyond. Here, I would also like to mention my other colleagues at DFKI, including but not limited to Leif Christensen, Peter Kampmann, Martin Fritsche, Christopher Gaudig, and Sylvain Joyeux.

In preparing for the final experiments, I would also like to thank my new colleges Nathan Smith for his support and his dedication to perfection, Andrew Edwards, Glenn Healey, Bethany Randell, and Nick Graham for electrical help and implementation. I am also very grateful to Karl Kenny, Greg Reid, and David Shea for supporting me in finishing this thesis and driving the vision of measuring the ocean.

Further, I would like to thank Kevin Köser for inspirational discussions as well as agreeing to be my second evaluator.

Finally, my sincere thanks belong to Maria and my whole family for all their tireless encouragement, inspiration, and motivation, especially during the last years when it was needed the most.

Abstract

In the underwater domain, optical sensors are extremely limited with respect to range, resolution, and accuracy in comparison to most terrestrial remote sensors. The reason for this is the medium water, which heavily interacts with electromagnetic signals and therefore reduces their corresponding signal-to-noise ratio. Also, many underwater areas can only be visited by remotely operated vehicles due to water pressure, turbidity, and or strong currents, posing a high risk for humans. This combination considerably increases the complexity of underwater metrology, and many applications currently require highly skilled personnel and large support vessels. Here, a simplification of these applications is presently effectively prevented by the performance gap of underwater optical systems in comparison to their terrestrial counterparts.

Motivated by the above limitations, this research work evaluates different optical sensing modalities when applied to the underwater domain and identifies their possible sweet spots. Based on these considerations, several novel fusion strategies for passive-active optical systems are presented able to reconstruct whole underwater scenes with high accuracy without relying on additional navigation systems. For their evaluation, a novel self-referenced optical 3D underwater scanner is implemented and used for several test setups as well as real-world scenarios. The implementation also includes a novel method for in-air calibration of flat-port cameras and integration into bundle adjustment frameworks for visual pose estimation. Here, the evaluation demonstrates that passive-active optical systems outperform standard methods when underwater sensor motion is a critical design parameter. The most significant advantage of self-referenced optical 3D scanners is that they compensate sensor motion in the same sensor domain as 3D measurements take place. This reduces the complexity of sensor co-calibration, ensures a similar accuracy for sensor pose and scene depth estimation, and broadens their possible application to smaller sensor platforms.

Contents

List of Figures

Chapter 1

The Vision of Measuring the Ocean

Measuring heights, widths, distances and positions with millimeter precision are essential for construction projects, inspection tasks or documentation. The exact knowledge of these quantities allows to stake-out buildings and infrastructure, calculate volumes and angles and to answer complex research questions. Over the last several hundred years sophisticated mathematical methods and technologies have been developed to improve the accuracy for ever more demanding applications.

Nowadays, satellite-based positioning systems like NAVSTAR-GPS (USA), Galileo (Europe), GLONASS (Russia) or Beidou (China) allow estimating the position of a receiver around the world with an accuracy of up to several meters. In the case of, known reference stations the accuracy can be even further improved down to 1-2 centimeter using, for example, Real Time Kinematic (RTK) positioning. In combination with 3D sensors such as lasers and or high-resolution cameras, local metric measurements with millimeter resolution can easily be referenced to each other as a standard tool for all kind of terrestrial applications.

However, in the case of marine applications, the medium water heavily absorbs electromagnetic waves used by most modern sensor systems such as satellite-based positioning or infrared laser systems. Also, the visible spectrum of light is strongly affected, reducing sight distances to several dozen of meters which is further degenerated to a few meters or even centimeters in case of turbid water.

Taking away the most powerful tools for performing metric measurements, leaves the underwater community with a half-empty toolbox. Therefore, many technologies were adapted or developed to fill the gap such as acoustic-based positioning systems, synthetic aperture sonar, laser systems optimized for the visible spectrum of light or scanning systems minimizing backscattering. Here, one of the biggest challenges is still the accessibility of underwater sides. Often areas in questions can only be visited by remotely operated vehicles due to water pressure, turbidity and or strong currents,

posing a high risk for humans. Having floating underwater vehicles as the primary platform for performing underwater metrology is therefore still today a very time consuming and error prone process which requires highly skilled personnel and large support vessels. Also, due to a lack of a global reference system, underwater sides cannot easily be referenced to each other, and low-resolution acoustic positioning has to be deployed to translate positions on the seafloor or in the water column to locations on sea level covered by global positioning systems such as NAVSTAR-GPS.

1.1 High Interest for Exploration and Surveys

Finding objects underwater can be still considered as looking for a needle in a haystack even for objects as big as submarines or airplanes. The reason for this is that all modern remote sensing technologies cannot penetrate larger volumes of water with a reasonable resolution making it impossible to perform surveys of deeper underwater areas via satellites or airplanes. Even ship based high resolution surveys are limited to shallow water regions and autonomous or remotely operated vehicles must be deployed to reduce the distance between sensor and seafloor for deep sea operation.

The result is a vivid market around underwater survey, exploration and metrology services for all kind of applications ranging from underwater mappings of black smokers to inspections of damaged ship hulls, mooring chains, pipelines, and foundations or simply locating missing airplanes or submarines. Here, depending on the application and the desired resolution, optical systems become more and more popular for inspection tasks due to their order of magnitude higher resolution and accuracy with respect to sonar based systems when operated under good visibility conditions.

However, there is still no universal tool available which can be configured according to the mission goals to automatically deliver the best possible resolution and accuracy under given circumstances. In contrast, highly trained personnel with a rich toolset of different technologies are required to solve a single task such as measuring the relative position of several underwater assets to each other or finding and inspecting cathodic protections on underwater pipelines in a more or less manual fashion.

1.2 Limited Remote Sensing Capabilities

Underwater sensors are extremely limited with respect to range, resolution and accuracy in comparison to most terrestrial remote sensors. The reason for this is the

medium water which heavily interacts with any signal reducing its amplitude and increasing the noise level. The only exceptions are acoustic based systems which benefit from smaller transmission losses in water in comparison to in-air but have a several order of magnitude lower working range than microwaves in-air or vacuum.

State of the art synthetic aperture sonar (SAS) delivers an along-track resolution which is range-independent and allows covering large underwater areas with a resolution down to a few centimeters. The working principle is similar to synthetic aperture radar (SAR) combining several measurements while moving along a known path. However, while a typical SAS can cover around two square kilometers per hour with centimeter resolution a state of the art SAR system like the Sentinel-1 can cover around one million times more area in the same time frame at a few meter resolution.

Unfortunately, the same analogy applies to other techniques like LIDAR or photogrammetry. While satellites can usually cover more than half a million square kilometers per day, an autonomous underwater vehicle equipped with a high-resolution camera would take more than 400 years to survey a similar area. In combination, with a lack of a global underwater positioning system, it becomes reasonable why underwater areas are very demanding to map, and large areas remain still untouched.

1.3 Previous Works

The first underwater sensing technologies date back to ancient civilization when ship navigators used sounding leads to measure the water depth manually. These point measurements helped to navigate in shallow waters and on rivers. Several improvements and automation were carried out over the next centuries. However, it took until the beginning of the 20th century when sounding was replaced by echo sounding measuring the amount of time a sound signal requires to travel from an emitter to the seabed and back to a receiver. By knowing the speed of sound in water, the exact depth can be calculated. All modern active sonar systems are based on this principle even if they measure or control additional quantities like the phase and the amplitude to perform beam steering or to synthetically enlarge their antenna.

In the last decade, the resolution of underwater sonar systems has continuously been improved by the usage of phased arrays and increased processing power allowing to embed real-time synthetic aperture sonar within medium-sized autonomous underwater vehicle reaching resolutions around $3cm$ while maintaining a $600m$ swath discussed in the work from Shea et al. [2013].

Furthermore, underwater LIDAR, photogrammetry and structured light systems are nowadays more and more used by the underwater community. However, the main driving factors for these technologies are still terrestrial applications, and many sacrifices must be made if these technologies are applied to the underwater domain without adaptions to the medium water. In order to overcome these limitations, different techniques were suggested over the last decades addressing performance issues such as light refraction, backscattering, and light absorption discussed in Hannon [2013], Balletti et al. [2015] and McLeod et al. [2013].

One of the challenges using underwater cameras, or to be more precise light-sensitive detectors or emitters, is their protection against the medium water. In the simplest case, a planar window is used to separate the detector/emitter from the medium also referred to as flat-port. However, this invalidates the classical pinhole camera model, or in general these sensor systems have no longer a single viewpoint showed by Agrawal et al. [2012] and Treibitz et al. [2012]. This none single viewpoint leads to large model errors for many machine vision algorithm using this constraint as a foundation to derive underlying mathematical models. To relax this, the primary challenge is not the back projection which can be solved using ray tracing approaches but the forward projection of 3D points onto the detector required for calculating re-projection errors.

Modeling the simpler back-projection of 2D image points onto their corresponding 3D rays in the water layer allows triangulating 3D scene points observed by multiple cameras or from different camera positions. Here, Treibitz et al. [2012] suggest modelling the distortion of a flat-port with the help of a ray map describing the light rays with respect to the flat-port interface. However, for calibration, it is assumed the principal ray of the camera is orthogonal to the flat-port interface and also the resulting underlying ray model is incompatible with methods relying on a perspective camera model. A similar approach is proposed by Fan et al. [2017] to model the back-projection of 2D image points onto 3D rays subject to refraction and to triangulate 3D scene points using active stereo.

To handle the more complex forward projection which allows to predict the position of a 3D scene point in the image domain and to apply iterative refinement methods Agrawal et al. [2012] derived an analytic solution resulting in a 12^{th} degree equation for a planar glass-water interface. They also showed that pinhole cameras behind a flat-port convert to an axial camera where all camera rays no longer intersect in a common point but with a common line. Here, solving the 12^{th} equation insight an iterative refinement for every single 3D point is usually not feasible. Therefore,

Jordt-Sedlazeck and Koch [2013] proposed to use virtual cameras for each 2D point allowing to project the corresponding 3D point and to minimize the effect of refraction. However, the proposed virtual camera error function adds additional degrees of freedoms for every image pixel during calibration and masks the actual focal length of the real camera.

Yau et al. [2013] also built on the work from Agrawal et al. [2012] to perform underwater camera calibration using the dispersion of light. However, instead of solving the analytic forward projection equation, they computed the plane of refraction combined with a 1D bisection search for the angle of a camera ray lying on that particular plane in such a way that the back-projected ray intersects with the given point.

Łuczyński et al. [2017] derived a so-called pinax model for calibration and image rectification from underwater cameras in flat-port housings. The pinax model is based on virtual cameras also used by Jordt-Sedlazeck and Koch [2013], but instead of using multiple virtual cameras for modeling the axial camera they search for a single virtual camera which fits best for all camera rays. This search is supported by constraining the distance of the camera focus point to the window. Furthermore, they propose a 2D lookup table to correct distortion due to refraction. However, as shown in the following chapters this introduces non-linear re-projection errors depending on the image location and scene depth becoming increasingly problematic when used in combination with iterative refinements.

With the help of refractive camera models, supporting efficient forward projection, it is possible to use them within a bundle adjustment framework for iterative refinements. However, for this also a good initial guess for each of the camera poses and 3D points is required. Therefore, Haner and Astrom [2015] derived non-minimal five- and six-point PnP solvers using only co-planarity constraints for flat refractive interfaces which can be used in a random sample conses framework. Also, solvers for solving the refractive plane parameters are included using 11 and 8 points recovering all camera parameters except for the perpendicular distance of the camera to the plane assuming zero distortion.

Jordt et al. [2015] proposes a complete structure from motion pipeline including the virtual camera error explained in Jordt-Sedlazeck and Koch [2013] to calculate the residual during bundle adjustment and a refractive relative pose solver for calculating the initial guess of each camera pose. However, as stated by the authors the proposed refractive direct solver is less reliable than solvers for the standard pinhole case.

In addition to refraction, the overall performance of underwater optical systems is primarily limited by the absorption and scattering properties of the surrounding

medium water. These fundamental limits to underwater imaging were already established in the 1960s by Duntley [1963]. In general, systems using wide-angle illumination sources are limited by backscattering especially when the source and receiver are placed close to each other. The reason for this is that this configuration increases the amount of light reaching the receiver reflected by particles in the water column before reaching the object of interest leading to contrast limited systems. Other more optimized systems may become bounded by forward scattering or by absorption limiting the strength of the returning signal and leading to a power limitation of the system explained by Klepsvik et al. [1990], Jaffe [1990]. As a result, the best performance can be expected for systems combining a narrow field-of-view for the source and at the same time for the receiver element minimizing the common volume between the illumination beam and the receiver. This setup reduces back- and forward scatter by a spatial rejection of light from the outside of this volume (Moore et al. [2000]). These, so-called laser linescan synchronous scanning systems have shown superior contrast resolution over an area or line illumination strategy which is pointed out by Jaffe [2010].

An additional degree of freedom is achieved by using either pulsed or modulated light sources. In the case of short pulses, the system is usually referred to as Light Detection And Ranging or short LIDAR (McLeod et al. [2013]). Here, a significant advantage of LIDAR systems is that they can ignore backscattering originated in the near-field which can quickly saturate other sensor systems. Another advantage of LIDAR systems is that the range to an object reflecting the laser pulse can directly be estimated by measuring the round-trip time of the pulse reaching the object and traveling back to the light sensor. However, because of a none homogeneous water column and the need for a precise timing due to the speed of light complex setups are required to reach superior range resolution (Jaffe [2015]).

More recent underwater visual imaging systems combine spatial and temporal separation between emitting and reviving signal to further increase the achievable working range. Such a system is briefly mentioned in Jaffe [2015] and is referred to as Pulsed Time-Resolved Laser Line Scan System. However, these systems rely on expensive hardware and are usually not available to the broader community.

On the other side underwater laser systems based on off the shelf components become increasingly popular in the underwater vision community. One of the most prominent systems are line structured light systems which are solely based on fixed line lasers and a machine vision camera and used for example by Klepsvik et al. [1990], Roman et al. [2010], Liu et al. [2010], Albiez et al. [2015], Lucht et al. [2018].

These systems face a contrast reduction in turbid water in comparison to line scanning systems explained by Jaffe [2010]. However, they are relatively simple to build, have no moving parts, and are a significant improvement over areal illumination based systems. For reconstructing larger scenes, the whole system must be moved in order to sweep the projected laser line over the scene. This sweeping is either accomplished by mounting the system onto a pan/tilt unit or by moving the whole sensor carrier. In both cases, the exact sensor position must be known in relation to a fixed coordinate frame while scanning to be able to merge the reconstructed 3D profiles into one consistent 3D point cloud. In the case the system is mounted to a tripod, the pose estimation is relatively straight forward. However, in the cases the system is attached to moving platforms the platform motion must be compensated while scanning. This compensation is achieved by fusing the 3D profiles with the pose estimation of the navigation system of the carrier vehicle based on Doppler velocity logs and military grade fiber optic gyros. This usually leads to reconstruction errors considerable larger than the accuracy of the optical systems itself due to sensor drifts and available sensor resolution of the navigation system. Therefore, Brignone et al. [2011] proposed a self-contained optical mapping payload which uses scene features in the image domain to track the sensor motion while also using the same camera for structured light scene depth estimation. This combination was also proposed by Strobl and Mair [2009] for in air applications. However, while for in-air applications more advanced commercial products (RGB-D cameras) are now available using areal light patterns like the Kinect One or the Intel Real Sense, for marine applications no such improvements were made. The main reasons for this are:

- Underwater systems using areal illumination face considerable more backscattering than ones using narrow light beams.
- Infrared light is heavily absorbed by water and cannot be used to separate the RGB camera from active depth estimation.
- Underwater light sources need considerable more power to reach comparable working ranges to in air sources due to electromagnetic absorption in the medium water.

Unfortunately, these constraints are contradictory for underwater systems able to estimate their sensor motion while performing active 3D depth estimation. The highest contrast for underwater active depth estimation is achieved by using a narrow field of view for emitter and receiver elements pointed out by Jaffe [1990]. However, to track scene features in the same sensor domain, areal sensors must be deployed to allow overlapping measurements. Therefore, the system proposed by Brignone et al. [2011] can be considered as an excellent compromise to allow motion tracking

while performing active depth estimation. Such self-referenced optical 3D sensors could significantly reduce the complexity of underwater metrology currently limited by the performance gap between underwater optical systems in comparison to their terrestrial counterparts and broaden its applications.

1.4 Contributions

This research work evaluates different visual sensing technologies for the underwater domain and identifies possible adaptions and fusion strategies to improve their accuracy and availability for various applications ranging from point measurements to autonomous navigation and mapping. A particular focus is on active and passive optical methods including a novel method for in-air calibration of flat-port cameras and their integration into bundle adjustment frameworks for visual pose estimation. Here, the integration is achieved by building on the framework proposed by Agrawal et al. [2012], but unlike Jordt-Sedlazeck and Koch [2013] a Taylor expansion is proposed to simplify the 12^{th} degree equation for planar glass water interfaces and to allow a tight integration into structure from motion pipelines.

Following this, several novel fusing strategies between active and passive vision are presented able to reconstruct whole underwater scenes with high accuracy from moving platforms without relying on additional navigation systems. For their evaluation tank as well as sea trial data were collected and processed to show the influence of different environmental parameters onto the reconstruction and pose accuracy. As a summary this research work makes contributions in the following areas where many vital parts were already presented on peer-reviewed conferences:

- Evaluation of active and passive optical 3D sensing methods for the underwater domain including experiments showing the influence of design decisions and scene properties onto their performance.
- A novel algorithm for accurate detection and localization of checkerboard corners for demanding applications like flat-port camera calibration - Duda and Frese [2018].
- A novel approximation of the flat-refractive forward projection with direct support for bundle adjustment frameworks for in-air, in-water camera calibration and structure from motion pipelines - Duda and Gaudig [2016].
- Novel fusion algorithms to fuse active with passive vision for building self-referenced underwater 3D scanner - Duda et al. [2015], Duda et al. [2016].
- Evaluation of several novel hardware and software setups of underwater self-referenced 3D scanners in real world use cases - Duda et al. [2016].

Chapter 2

Signal Degradation

Any system measuring a value of a quantity relies on a signal which is varied in time or space depending on the state of the quantity. In general, signals are subject to disturbances which are an unwanted modification to an ideal signal and result in measurement uncertainties and errors. These disturbances can be grouped into ones which can be evaluated by statistical methods (random uncertainties/errors) and into ones which are evaluated by other means (systematic uncertainties/errors).

In the case of underwater measurements, the medium water has a high impact on these errors and uncertainties of the measurement process. This reduction in performance is especially prominent for vision systems not taking the optical properties of water into account introducing large systematic errors. The reason for this are violations of the assumptions used for modeling standard vision systems. In the following chapter, the most prominent effects due to the medium water are described changing the optical path of light rays or reducing their intensity.

2.1 Optical Properties of Water

Most optical properties of water can be divided into inherent and apparent properties. Inherent optical properties (IOP's) only depend on the medium itself and are independent of the ambient light. According to Mobley [1994] the two fundamental IOP's are the absorption coefficient and the volume scattering function. Other IOP's include the index of refraction, the beam attenuation coefficient and the single-scattering albedo. Apparent optical properties (AOP's) depend both on the medium and the ambient light field and are usually used as descriptors of the water body. In the following only inherent optical properties are regarded, and apparent optical properties are ignored.

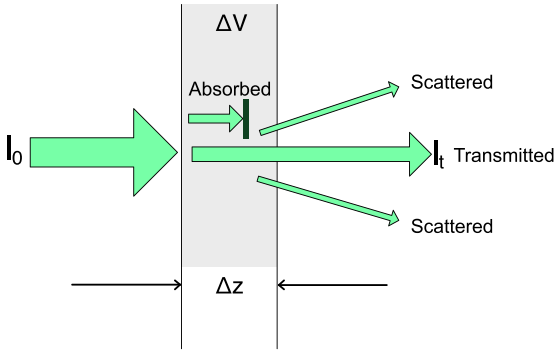


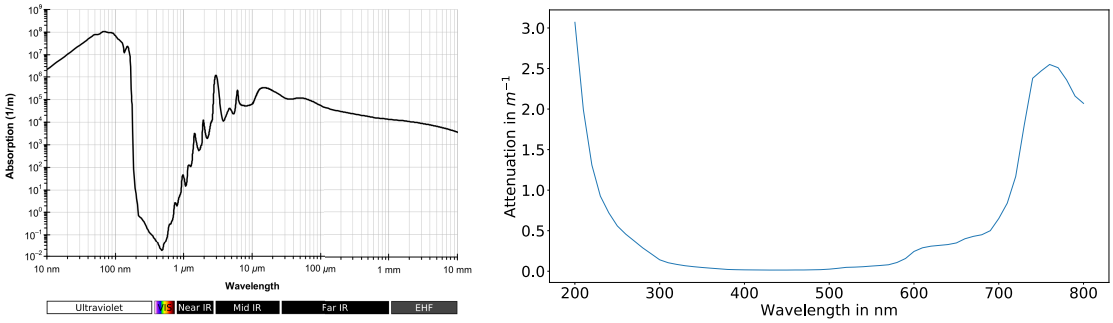
Figure 2.1: Inherent optical properties.

Light Attenuation and Scattering in Water

In general, if a small volume ΔV of water is illuminated by a narrow collimated beam of monochromatic light some part is absorbed within the volume, some part is scattered out, and the remaining part is transmitted through without facing any change which is displayed in Fig. 2.1. Here, attenuation is defined as the gradual loss of flux intensity through a medium. Based on the Beer-Lambert law, the remaining intensity I of a signal I_0 through a material is directly related to the material thickness z and its attenuation coefficient α.

$$I_t(z, \alpha) = I_0 e^{-z\alpha} \tag{2.1}$$

The attenuation coefficient α is the sum of all absorption and scattering, which are defined analog to each other. The absorption usually only accounts for effects transferring energy from a photon to the medium. Whereas scattering accounts for effects deviating the photon from its original path. In the case of pure water, there is a narrow band from near-ultraviolet to near-infrared, where the electromagnetic absorption decreases by over nine orders of magnitude in comparison to wavelengths outside of this band. The reason for this is that at blue wavelengths, photons do not have the right energy to interact efficiently with the water molecules or to boost electrons into higher states. At smaller wavelengths, photons gain enough energy to excite atomic transitions, and the absorption rapidly increases. The same is valid for wavelength beyond red, where photons have the right energy to excite whole water molecules. However, even in the narrow band of visible light, the absorption heavily dependents on the wavelength. Here, the absorption coefficients, as determined by Smith and Baker [1981], are displayed in Fig. 2.2, which gives a lower bound for



(a) Wide spectrum by Kebes [2008] (b) Visible spectrum by Smith and Baker [1981]

Figure 2.2: Electromagnetic absorption coefficient for liquid sea water.

wavelength-dependent absorption in clear water. Interesting enough, this narrow band overlaps with the wavelengths of the sun's maximum energy output and with a corresponding window in atmospheric absorption. It is only this overlap of open windows with the energy source, which has enabled aquatic life to develop in the form we see on earth.

In addition to the absorption of pure water, additional absorption takes place due to dissolved organic matter like yellow matter, phytoplankton, and organic detritus. Here, the additional absorption can easily be the dominant part of the total absorption depending on the water conditions. The absorption spectrum for selected waters is displayed in Fig. 2.3.

The second fundamental IOP is the scattering of photons changing their path due to interactions between them and atoms or molecules. Here, small-scale scattering by water molecules serves as a minimum value for the scattering properties. However, the scattering properties also strongly depend on organic and inorganic particles and by fluctuations in the index of refraction. Therefore, because of more significant fluctuations in the index of refraction seawater also scatters light around 30% more than pure water pointed out by Mobley [1994].

As a result of absorption in combination with scattering, water has a relatively high opaqueness outside of the band of visible light, and all wavelengths outside of this band are usually unsuitable for measuring scene properties in distances exceeding the size of the measuring system. Therefore, this opaqueness usually leaves only the visible spectrum of light for active or passive optical underwater sensing.

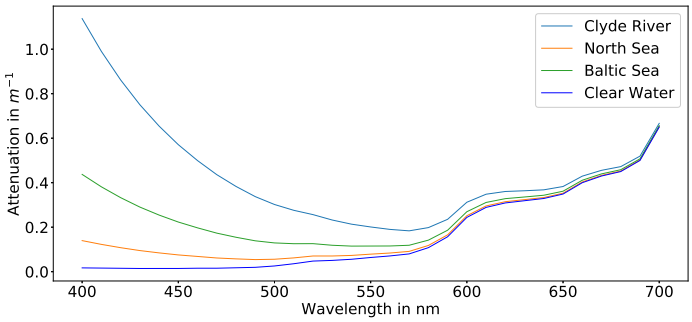


Figure 2.3: Absorption coefficients for selected waters with an increase in yellow matter Mobley [1994].

Refraction Index

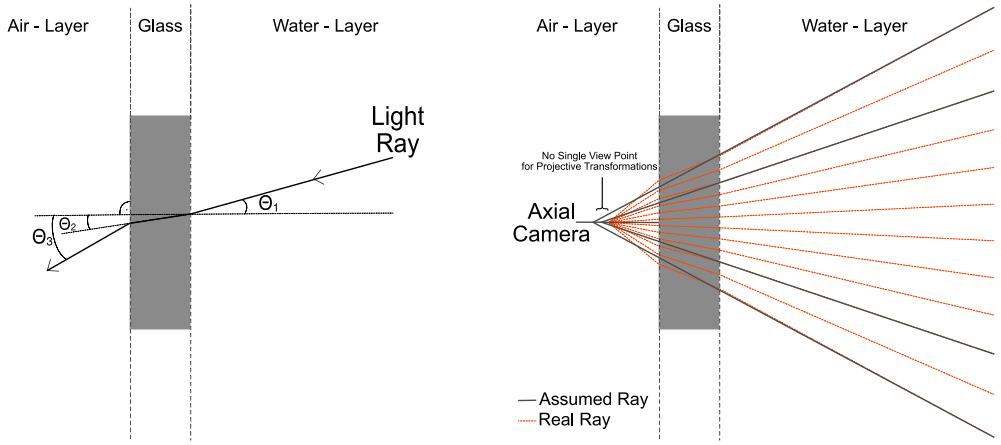
The refraction index n of a material is a dimensionless quantity describing how fast light propagates through the medium with respect to light traveling through vacuum expressed in Eq. 2.2 where c is the speed of light in vacuum.

$$n_{medium} = \frac{c}{v_{medium}} \tag{2.2}$$

In the case of pure water having a refraction index of $n_w = 1.333$, light travels 1.333 times faster in a vacuum than in the water body. However, the refraction is usually not truly constant, and small variations of n within a material such as random thermal fluctuations result in small-scale fluctuations in the index of refraction, causing light scattering. In general, the refraction index depends mainly on the four parameters temperature, wavelength, pressure, and salinity. Here, n increases with decreasing wavelength or temperature and an increase in salinity or pressure. The relevant extreme values for hydrologic optics is usually between 1.329 and 1.367, translating to a maximal variation of around three percent discussed in Mobley [1994].

2.2 Refraction and Reflection at Housings

A light sensor is a sensitive element that must be protected from pressure, dust, water, and other elements. This protection is usually achieved by placing the sensor into a pressure housing and behind a transparent glass port allowing light rays to enter the optical system. Here, when a light ray enters a medium with a different refraction index than water, the change in speed leads to bending the light ray according to Snell's law of refraction summarized in Eq. 2.3 where θ_1 and θ_2 are the angles of incidence and n_1 and n_2 are the refraction indices.



(a) Refraction on a flat-port (b) Axial-Camera due to refraction

Figure 2.4: Refraction of light rays entering an underwater flat-port transforming a pinhole camera into an axial camera.

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \tag{2.3}$$

In the case of a simple glass window, the refraction invalidates the Single Viewpoint Constraint, which requires that all incoming principal light rays of a lens intersect at a single point. In fact, a camera system behind an underwater flat-port converts to an axial camera where all rays meet in a common line instead, as shown by Agrawal et al. [2012]. This fundamental difference to the usual assumed underlying pinhole camera model is visualized in Fig. 2.4 and introduces large model errors depending on the application. In addition to refraction, the refraction index also describes the amount of light that is reflected when entering a medium with a different refraction index. In the simplest case where the light ray is orthogonal to the surface, and the permeability is one for all involved mediums the intensity of reflected light is given by the reflection coefficient R.

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 \tag{2.4}$$

This reflection translates to a transmission loss of up to ten percent when light travels through a window surrounded by air if no special anti-reflection coating is used, matching the refraction indices of both mediums. In general, R also depends on the angle of incidence and the permeability of the mediums and strongly increases for shallow angles. Its exact value can be calculated using the Fresnel equation, which is also used to calculate the light reflection for incoming rays entering an underwater

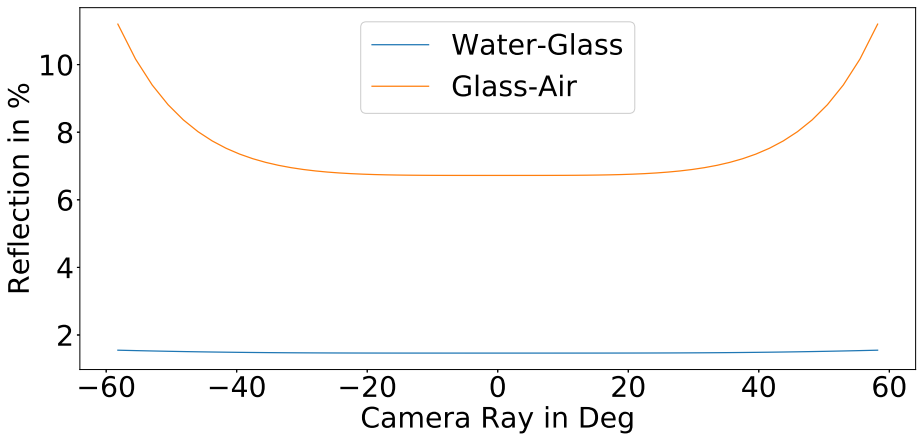


Figure 2.5: Light reflection of incoming light rays on underwater flat-ports with $n_{water} = 1.333$, $n_{glass} = 1.7$ and a permeability of one for all mediums.

flat-port visualized in Fig. 2.5. Here, the dominant transition takes place on the glass-air interface, where the difference between both involved refraction indices is the highest. Therefore, no special anti-reflection coating is required on the window side facing the medium water. However, on the other side, transmission losses as high as ten percent are present due to reflection if no special coatings are applied.

2.3 Image Degradation

Standard area image sensors face different types of noise, which can be mainly grouped into temporal and spatial noise. Here, temporal noise is defined as a temporal variation in pixel output values under constant illumination due to supply, substrate noise, and quantization effects being most pronounced at low signal values. Spatial noise also called fixed pattern noise is constant for a given sensor, but varies from sensor to sensor. It consists of offset- and gain-components for each pixel, which are dependant on the pixel location and causes the most degradation in image quality at low illumination. Following Suess [2016] and combining temporal and the fixed pattern noise, the total input noise Q_n can be expressed as:

$$Q_n = Q_{shot} + Q_{reset} + Q_{readout} + Q_{fpn} \tag{2.5}$$

Where Q_{shot} is the dark current shot noise with a Gaussian distribution with zero-mean. Q_{reset} is the reset noise, and $Q_{readout}$ is the readout circuit noise. Both are independent of the signal. Q_{fpn} is the fixed pattern noise and grows with the signal.

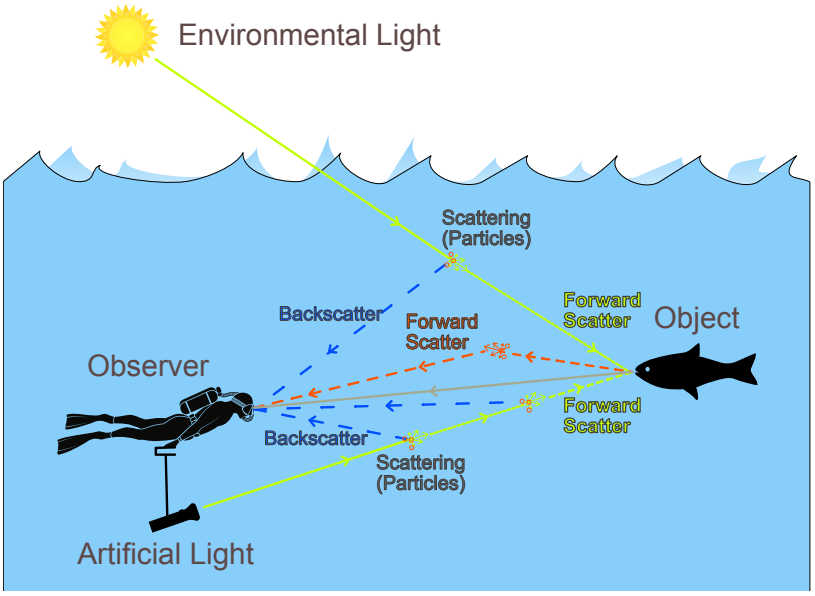


Figure 2.6: Image degradation due to forward and backscattering.

Also, the medium water acts as an additional noise source. Based on the theoretical foundations of the optical underwater image formation model laid out by McGlamery [1980] and extended by Jaffe [1990] the underwater image can be presented as a linear superposition of mainly three components. Following this, the total irradiance E_T of every pixel is the sum of the light E_d directly reflected by the object, light E_f reflected by the object but has been scattered at a small angle afterward (forward scattering) and light E_b which has been scattered before reaching the object (backscattering).

$$E_T = E_d + E_f + E_b \tag{2.6}$$

The dominance of the components E_b and E_f in comparison to in-air applications result in an additional image blur, which can be modeled with the help of a point spread function (PSF) reducing the contrast of the image respectively the signal-to-noise level shown by Schettini and Corchs [2010]. In the ideal case, a vertical laser line imaged by a camera would have a single well-defined peak value for each image row. This is, for example, the case for in-air or underwater applications with low water turbidities. With increasing turbidity, the well-defined peak value is going to be blurred by the forward scattering, and the backscattering increases background illumination of the common volume between the camera and the laser projector. This circumstance is visualized in Fig. 2.7 for three different turbidity levels.

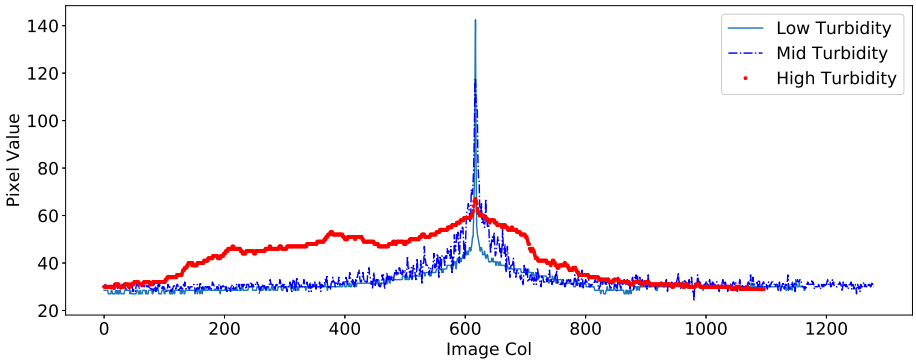
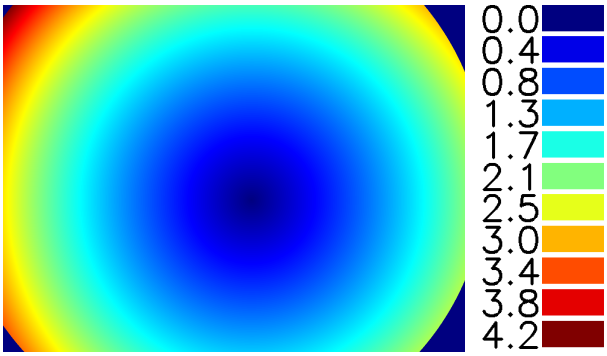


Figure 2.7: Measured brightness distribution along an image row for a projected laser line for different levels of water turbidity.

In addition to the blur due to scattering, underwater images are often dominated by green and blue colors and considerably missing red tones. The reason for this is the wavelength-dependent light attenuation in water. Here, the red color disappears first at a scene depth of around $3m$, followed by green and blue. Therefore, most underwater laser systems concentrate on the green and blue wavelengths. The same is true for underwater photogrammetry or mosaicking, where often only the green channel is utilized for visual feature detection and matching due to its lower noise level in comparison to the other channels. Here, a critical sensor property is its dynamic range allowing imbalanced color channels without saturation and to allow for color correction in post-processing.

However, different colors are not only subject to different absorption, but they are also refracted differently. The reason for this is the wavelength-dependent refraction index of a medium leading to not only a geometric distortion but also to chromatic aberration where each color is focused differently, leading to rainbow edges in areas of contrast. This chromatic aberration is especially problematic for monochromatic cameras where this cannot be compensated by offsetting each color channel resulting in unsharp images for larger radial distances. However, even for color cameras, this cannot be fully compensated because each color channel has a bandwidth resulting in a de-focus across the channel, limiting the maximal resolution of the optical system. In the case of a flat-port color camera with a $1/3$ inch sensor (1280×960 pixel, $2.9mm$ focal length), and a $5mm$ sapphire glass window, the image blur due to wavelength-dependent refraction insight each color channel is visualized in Fig. 2.8.

In the case of dome ports or special wet lenses, chromatic aberration and geometric distortions can be reduced, but the remaining errors are often difficult to model. Also, image restoration and enhancement algorithms can be used to somehow mitigate the



(a) (b)

Figure 2.8: Simulated image blur in pixel for the green color channel ($480 - 600nm$) due to wavelength dependant refraction on a flat-port window.

effects of the medium water onto the image quality discussed in work by Schettini and Corchs [2010].

Conclusion

The medium water alters the light signal on many levels before reaching the image sensor. First, the light signal is refracted when entering the medium water. After this, the light signal is partly absorbed by the medium and partially scattered before reaching an object reflecting it back to the sensor. The reflected light is again subject to scattering and absorption before reaching the sensor housing. Here, the signal is refracted again, leading to geometric and chromatic distortions before finally reaching the image sensor. Also, floating particles in the path of the light yield further altering of the signal. Most of these effects can be explicitly modeled. However, in practice, it is difficult to obtain the correct model parameters, and estimations must be used to correct the measured signal and remove unwanted effects. In any case, the signal-to-noise ratio rapidly drops with increasing penetration depths making it infeasible to use electromagnetic waves to measure scene properties in greater distances usually used for remote sensing via satellites or airplanes. Therefore, the usage of underwater vision techniques requires sensor platforms able to maintain a low altitude to visually inspected objects. In return, visual inspections can outperform sonar-based technologies with respect to resolution and accuracy reaching sub-millimeter resolution depending on the enviromental conditions.

Chapter 3

Impact onto Optical Systems

The medium water has a direct influence on signal propagation like discussed in the previous Chapter 2. This influence leads to all sort of errors for standard optical methods summarized in Fig. 3.1. In the following, these errors are outlined for the methods on the bottom of the figure including active and passive triangulation based as well as indirect- and direct-time of flight methods.

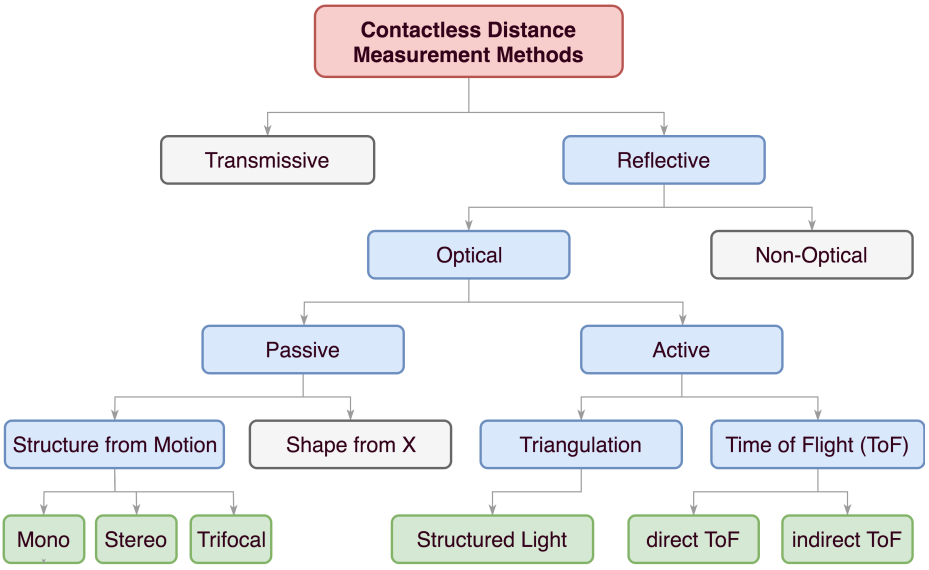


Figure 3.1: Taxonomy of distance measurements derived from Luhmann [2010].

3.1 Stereo

A stereo camera system is a sensor system that is very similar to the three dimensional perception of a human. It uses two projections of the same scene observed from two slightly different physical locations to derive the scene depth with the help of

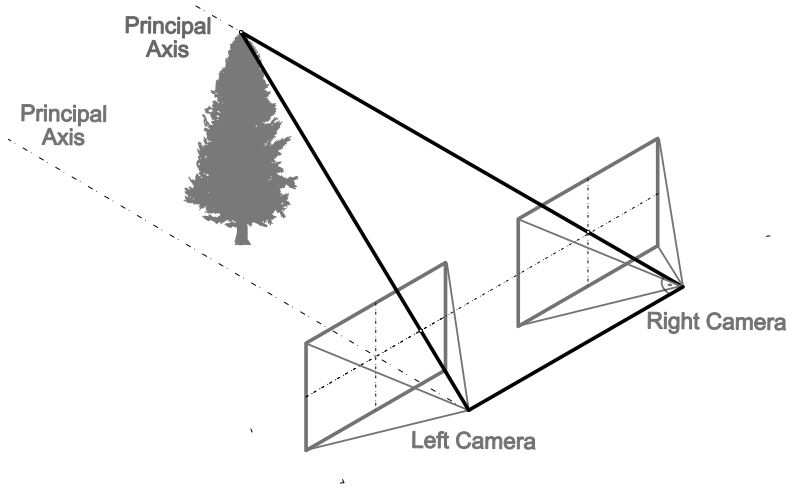


Figure 3.2: Schematic of a stereo vision system.

triangulation, which is visualized in Fig. 3.2. This triangulation is possible because the variation in the camera locations leads to a difference between both projections, also known as parallax, which is scene-depth dependant and is caused by perspective distortion. In the case of a visual stereo system, usually two identical cameras are used, having a fixed baseline to each other to record two images of the same scene at the same point in time, also referred to as stereo image pair.

Scene Depth Estimation

The scene depth calculation of a stereo system is based on the triangulation principle, which has been used for centuries to calculate distances to objects with the help of trigonometric functions and by measuring two angles and a baseline. This principle is demonstrated in Fig. 3.3 for calculating the distance z to a remote object by determining the angles α and β and the baseline c, which can often more accurately be measured than directly measuring the longer distance z.

In the case of a stereo system, this translates to finding common scene points in both camera images. Here, the position of a scene point in the image domain determines the angle for the triangulation, and the distance z can be calculated with the help of Eq. 3.1 visualized in Fig. 3.3 using the known baseline c between both cameras.

$$z = \frac{c}{\frac{1}{\tan(\alpha)} + \frac{1}{\tan(\beta)}}. \tag{3.1}$$

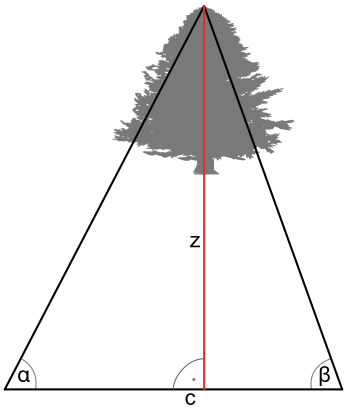


Figure 3.3: Triangulation of the unknown distance z based on the known baseline c and the two angles α and β.

In addition, most machine vision cameras can be approximated by a pinhole camera model P, linearly relating the coordinate of a point X in the three dimensional space to its projection x for a specific camera pose $R, \widetilde{C}$.

$$\begin{aligned} x &= PX \\ P &= KR[I - \widetilde{C}] \\ K &= \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \end{aligned} \tag{3.2}$$

Based on the camera model given in Eq. 3.2, the required angles α and β can be expressed in terms of the focal lengths f_a and f_b of both cameras and the location p_a and p_b of the observed object in both image planes according to Eq. 3.3. Here, to account for that β has a different orientation than α a minus must be added. Both angles are measured against the image plane and are not identical with their corresponding camera ray angle measured against the principal axis of the camera.

$$\begin{aligned} \tan(\alpha) &= \frac{f_a}{p_a} \\ \tan(\beta) &= -\frac{f_b}{p_b} \end{aligned} \tag{3.3}$$

A more in-depth explanation of the pinhole camera model, not required at this point, is given in Section 4.2. Here, assuming both cameras have the same focal

length $f_a = f_b$, the distance z of an object seen in both cameras as p_a and p_b can be determined by the following equation with c being the known and fixed baseline between the cameras.

$$z = \frac{cf}{p_a - p_b} \tag{3.4}$$

This assumption also allows calculating the depth resolution as the absolute value of the partial derivative of Eq. 3.4 with respect to p_b translating small changes of the projected object location dp_b to distance changes dz.

$$\begin{aligned} |dz| &= \left| \frac{\partial z(p_b)}{\partial p_b} \right| dp_b \\ &= \frac{cf}{p_b^2} dp_b \\ &= \frac{cf}{\left(\frac{cf}{z} + p_a\right)^2} dp_b \end{aligned} \tag{3.5}$$

In the case the object is projected onto the image center of the first camera, p_a is zero and Eq. 3.5 can be further simplified to:

$$|dz| = \frac{z^2}{cf} dp_b \tag{3.6}$$

This simplified case for stereo triangulation is also demonstrated in Fig. 3.4. For real-world cameras, the differential dp_b is limited by the ability to determine the exact sub-pixel location of an object in the camera image. Here, sub-pixel peak detection algorithms can localize strong features up to $\frac{1}{100}$ of a pixel. However, due to the relation $|dz| \propto z^2$, the depth resolution quadratically scales up with the object distance requiring large and larger baselines c to maintain a constant depth resolution.

Underwater Challenges

Underwater stereo vision systems face multiple challenges due to the signal degradation introduced by the medium water. One very prominent error source is the camera housing, which has unwanted effects onto the light propagation depending on the shape of its air-water interface.

In the case, a flat-port camera housing is used, the refraction of the light on the water glass and the glass-air interface invalidates the pinhole camera model because

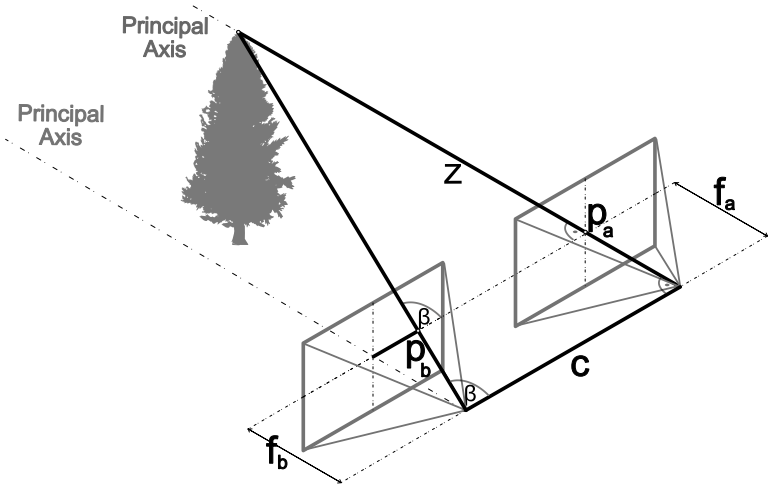
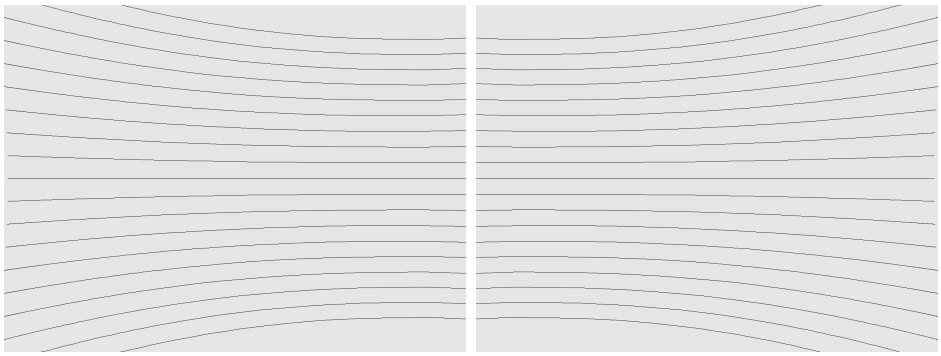


Figure 3.4: Passive stereo camera triangulation of the scene depth z.

the camera becomes, in fact, an axial camera shown by Agrawal et al. [2012]. Besides, the refraction also leads to chromatic distortion due to the dependency of the refraction indices on the wavelength. This introduces serious shortcomings of stereo algorithms when applied to underwater flat-port cameras. The underlying reason is that most stereo algorithms use the fact that in the case of two pinhole cameras, corresponding feature points must lie on a plane which is defined by the position of both camera centers and the position of the feature in one of the images. This so-called Epipolar constraint helps to reduce the search space in the other image to a search along a straight line. However, in the case of flat-port housings, these lines become curved (Li et al. [2018]), and the search for corresponding features will fail if not taken into account, which is visualized in Fig. 3.5.



(a) Left Camera Image (b) Right Camera Image

Figure 3.5: Curved Epipolar lines of an ideal underwater stereo flat-port camera system with $10cm$ baseline and a $3cm$ thick glass port.

Also, the triangulation of scene points is strongly affected due to the change in

refraction. Even if the camera is calibrated while being submerged to compensate for these effects, the pinhole model is usually not able to compensate for the depth-dependent distortion introduced by the flat port housing. The best mode of operation with small error margins is to calibrate the camera to a fixed distance, which is equivalent to the range later also used during an operation. Any depth deviation to it will introduce errors that usually cannot be compensated by standard lens distortion models (Łuczyński et al. [2017]), and flat-refractive projection must be taken into account explained in the next chapter.

In the case of dome-ports or special wet lenses, the single viewpoint constraints can be ideally maintained using standard stereo algorithms, and excellent results can be achieved by the cost of increasing hardware/assembly costs and a reduced depth of focus (Drap et al. [2015a]). Here, a more in-depth comparison between different camera ports is given in Section 4.1.

Assuming the camera housing is introducing no additional image distortion, the remaining effects are due to forward and backward scattering as well as electromagnetic absorption limiting the contrast and working range. These limitations are especially prominent if the light source is located in the near vicinity of the cameras, increasing the common water volume between the emitter (light source) and receivers (cameras), which considerably increases the backscattering of the light signal.

Underwater Applications

Currently, there are mainly off the shelf underwater stereo camera systems available for stereoscopy. These systems either display image pairs on a 3D display or, with the help of 3D glasses, create the illusion of depth in an image for the human observer. Other commercial systems mainly address the research community and allow to size underwater objects by manually selecting features in both images and manually match them. However, none of these solutions take the refraction explicitly into account nor support the generation of metric correct depth images, which is usually provided by in-air stereo camera systems. In research, underwater stereo systems are more widely used to demonstrate underwater mapping and localization tasks with promising results for the near field (Drap et al. [2015b], Łuczyński et al. [2017]). An overview of underwater stereo vision in the underwater domain is given in Tab. 3.1.

Camera Port/Model	Application	Authors
Flat Port Pinhole	Dataset / Obj. Detection	Oleari et al. [2015]
	Dense matching	Massot-Campos et al. [2015] Huo et al. [2018]
	Visual Odometry	Nawaf et al. [2018]
	SLAM	Schattschneider et al. [2011] Hildebrandt [2014] Carrasco et al. [2016] Carrasco et al. [2015]
Flat Port Refractive	Dense matching	Li et al. [2018] Fan et al. [2017] Jordt-Sedlazeck et al. [2013]
	Active dense matching	Kuo and Nobuhara [2017]
Dome Port Pinhole	Dense matching	Bruno et al. [2011] Bianco et al. [2012] Drap et al. [2015a]
	Active dense matching	Detry et al. [2018]
	Visual Odometry	Drap et al. [2015a]
	SLAM	Rossi et al. [2018]

Table 3.1: Survey: Underwater stereo vision systems.

Strengths and Weaknesses

Stereo systems can be bought off the shelf, usually delivering several dense depth images per second for in-air application by the use of specialized hardware for real-time processing and triggering the cameras at the same point in time. The achievable depth resolution mainly depends on the baseline between both cameras, the camera resolution, and the availability of suitable scene features. In the case the system is facing homogeneous regions, it is unable to triangulate depth information for these regions and has to fill these areas with surrounding measurements. Also, the working volume is smaller than for structure from motion systems because it is equivalent to only the overlapping field of view of both cameras.

Its strengths lie primarily in the usages of standard machine vision cameras and the ability to generate high-resolution 3D data for small working volumes with a relatively high frame rate without any strong constraints on the sensor motion (Drap et al. [2015a]). However, because the whole scene must be illuminated, stereo systems use one of the worst illumination strategies for underwater vision, and they usually become contrast limited due to backscattering (Jaffe [2010]). Therefore, in literature, usually underwater stereo is used for small working distances between $1m$ (Drap et al.

[2015a]) and $2m$ (Oleari et al. [2015]) and with rather poor results with distances up to $3.2m$ (Pfingsthorn et al. [2016], Detry et al. [2018])

3.2 Structure from Motion

Structure from motion describes the process of estimating the motion of a camera and 3D scene points by solely taking the parallax of the estimated 3D points into account while moving through the scene as visualized in Fig. 3.6. It is the generalization of a stereo system or a trifocal system Torr and Zisserman [2002] replacing the observation of additional physical cameras by observations of the same camera taken at a different point in time.

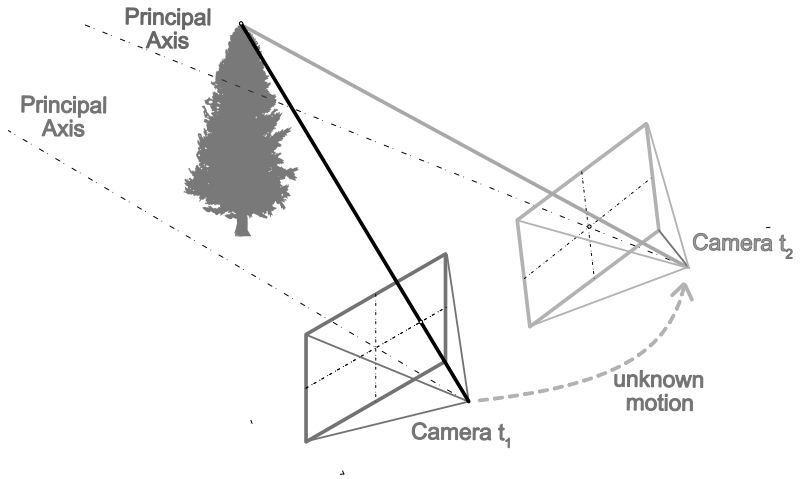


Figure 3.6: Schematic of a structure from motion system using multiple views captured by the same camera.

Scene Depth Estimation

In the case of a static scene, a structure from motion system is equivalent to a stereo system with an unknown baseline between both cameras. Therefore, because of this, the scale of the scene cannot be recovered by a simple structure from motion system, and it can't detect if an imaged object is, for example, a real house or a doll's house. However, because the measurements from multiple locations can be fused, the depth resolution is no longer limited by a fixed baseline and virtually growths with the number of observations, which is also referred to as multi-view geometry by Hartley and Zisserman [2004]. Based on this, the depth resolution can be bounded by Eq. 3.7, where the baseline c is the maximal distance between camera locations from where

identical scene features were observed. Note the additional scalar factor s for fixing the unknown scale of the reconstruction.

$$|dz| = \frac{z^2}{cfs} dp_b \tag{3.7}$$

Underwater Challenges

In principle, the same challenges apply for underwater structure from motion like for underwater stereo systems. However, on one side, structure from motion systems cannot take advantage of limiting the search space by utilizing Epipolar-Constraints because the pose between subsequent positions is unknown. On the other side, by not limiting the search space, they will be also less affected by refraction during the feature detection and matching step. Only during structure estimation and a subsequent bundle adjustment step, scene depth dependant triangulation errors, introduced by the refraction, will alter the camera pose and 3D scene point locations, when refraction is not considered properly. Here, image features located further away from the principal point are more affected due to larger angles of refraction, reducing the global accuracy of the final 3D reconstruction in comparison to dome ports discussed by Menna et al. [2017]. These un-modeled errors usually lead to a curved reconstruction of previously straight structures, and larger sea-floor stripes seem like rolled up similar to a corkscrew. The reason for this is that the remaining residuals due to refraction can be further minimized by the bundler adjustment step by falsely bending these regions visualized in Fig. 3.7.

Underwater Applications

Like stereo systems, structure from motion systems are mainly used for underwater mapping tasks in research at low altitudes. Commercial systems usually try to minimize the effect of the medium water during image acquisition to be able to apply off the shelf software suites for photogrammetry in a post-processing step. Otherwise, larger regions tend to be bent due to inhomogeneous errors across the image domain. However, good results have been achieved for underwater archaeological sites, which also include ground control points to bound the reconstruction error (Yamafune et al. [2017], Kan et al. [2018], Menna et al. [2017]). An overview of underwater structure from motion system is given in Tab. 3.2.

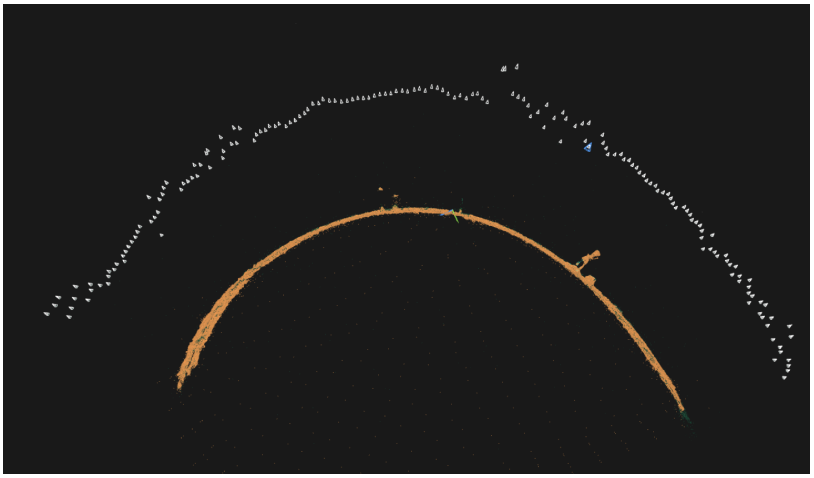


Figure 3.7: 3D scene reconstruction of a towing tank heavily distorted due to an un-modelled flat-port housing (white: camera poses, orange: tank floor).

Camera Port/Model	Application	Authors
Flat Port Pinhole	Visual Odometry	Botelho et al. [2009]
	SLAM (real-time)	Concha et al. [2015] Leonard et al. [2012]
	SfM / SLAM (offline processing)	Beall et al. [2011] Menna et al. [2017]
Flat Port Refractive	SfM / SLAM (offline Processing)	Jordt et al. [2015] Jordt-Sedlazeck and Koch [2013] Kang et al. [2012]
Dome Port Pinhole	Visual Odometry	Ferrera et al. [2018]
	SLAM (real-time)	Meireles et al. [2015] Ferrera et al. [2018]
	SfM / SLAM (offline processing)	Kan et al. [2018] Menna et al. [2017] Yamafune et al. [2017] Drap et al. [2015b] Arnaubec et al. [2015]

Table 3.2: Survey: Underwater structure from motion (SfM) and related technologies.

Strengths and Weaknesses

In its basic configuration, a structure from motion system consists of a single camera and has the lowest hardware requirements under all-optical systems. However, the software must simultaneously estimate the pose of the camera and estimate 3D scene features. Therefore, it also has the weakest constraints between measurements and is likely to fail under challenging conditions. To improve the robustness, it is often fused with an inertial measurement unit to stabilize it during rotation and to introduce some weak constraints for estimating the scale of the reconstruction. Here, for real-time 3D perception, no real commercial solutions are currently available, although they gain more and more popularity in the mobile phone market to enrich application with virtual reality capabilities.

Its strength is its offline processing capability and to generate baseline independent depth information from images like aerial photos. However, it relies on unique scene features that can be tracked over an extended period, which is not always possible. Also, any dynamic in the observed scene is adding additional challenges. Furthermore, similar to a stereo system, the whole scene must be illuminated, leading to contrast limitation due to backscattering (Jaffe [2010]). Therefore, good results can usually only be achieved for working distances smaller than $3m$ similar to stereo systems.

3.3 Structured Light

A structured light system is an active vision sensor that projects a known light pattern onto the scene observed by one or more machine vision cameras to estimate the scene depth. They are used in a large variety of use cases, like in industrial applications for quality control, in consumer products for gesture control or robotics for simultaneous localization and mapping. Depending on the application, many different configurations and light patterns like dots, lines, or Gray codes exist, including coding in the temporal and or spatial domain. Here, a more in-depth consideration is given in the work by Chen [2015].

In the case of marine applications, usually, lines or dot patterns are used to minimize backscattering (Jaffe [2010]). These light patterns also have the advantage that they can be easily be generated using high power laser diodes to compensate light absorption in the medium water. Other light patterns often require a light engine that is either very bulky or cannot provide highly focused high power light patterns and are mostly used in research to obtain close range underwater reconstructions

(Bianco et al. [2012]). Here, a table of different structured light methods used in the underwater domain is given in Tab. 3.3.

Scene Depth Calculation

A structured light system is equivalent to a stereo system where one of its cameras is replaced by a light projector, which can be understood as an inverse camera. Here, the projector actively projects a known light pattern onto the scene. The other camera observes this light pattern as artificial features perspectively distorted due to the baseline between camera and projector.

After the positions of artificial scene features are detected in the image domain, their position in the known light pattern must be determined. This detection is required to derive the angle α of their corresponding light rays emitted by the projector and generating the observed artificial features, also referred to as a decoding step. The complexity of this decoding step massively depends on the used light pattern. For example, in the case of a random point pattern, additional neighboring information in the pattern must be used. In contrast, a single point or line pattern does not require any further associations. Here, it is sufficient to know the pose of the projector with respect to the camera and to identify the light pattern in the image domain because its projection angle α is uniquely given by the projector to camera arrangement.

Following the determination of the emitting rays, the distance z of each artificial feature can be triangulated by additionally utilizing the angle β of the corresponding camera rays projecting the artificial features onto the image plane and the known baseline c between camera and projector. This triangulation process is demonstrated in Fig. 3.8 for a laser sheet and is analog to the stereo triangulation case.

The depth resolution of the active vision system is also analog to the passive case and estimated by Eq. 3.6 assuming the resolution of the projector is not the limiting factor which is, for example, the case for laser line projectors usually following a Gaussian dispersion model. For a Quad VGA camera (FOV 60°) its estimated depth resolution in relation to the baseline is visualized in Fig. 3.9.

Underwater Challenges

In general, structured light systems are subject to refraction similar to stereo systems. However, there are some simplifications possible, depending on the light pattern. For example, in the case of a laser line pattern, the distortion on a flat-port can be

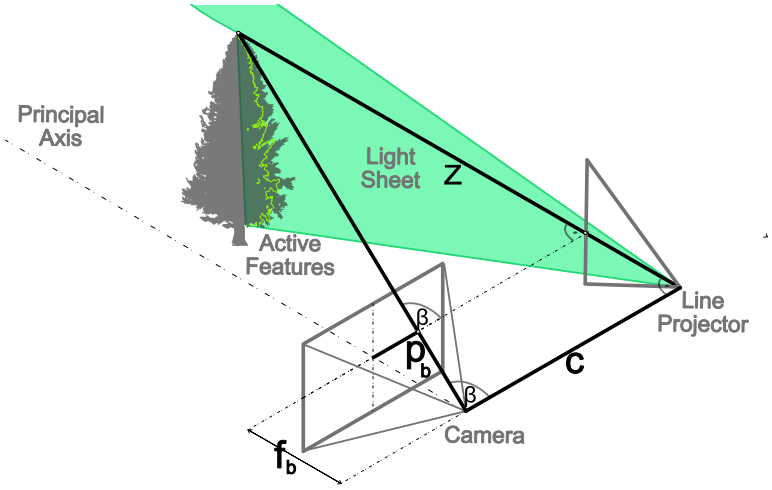


Figure 3.8: Active triangulation using line structured light.

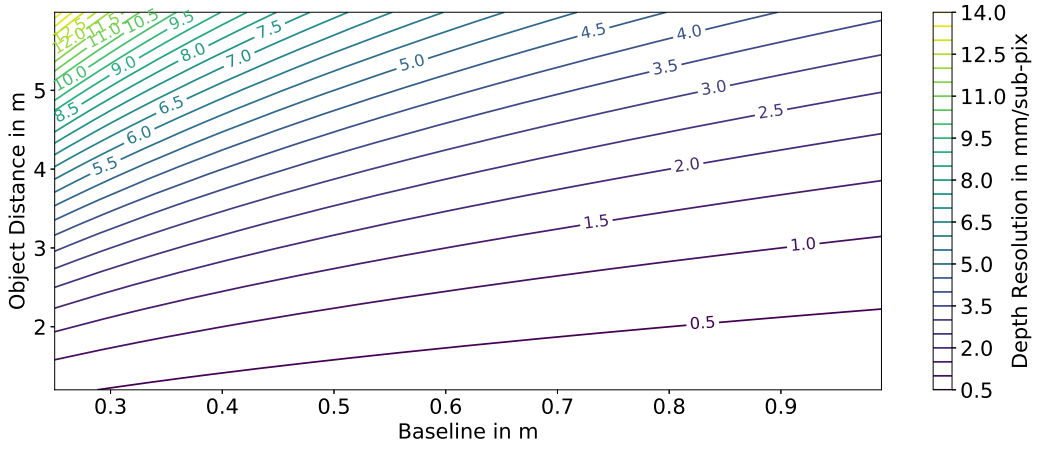


Figure 3.9: Estimated depth resolution in mm/pixel of a line structured light system in relation to its baseline and the distance to the observed object (FOV 60°).

neglected entirely if the laser sheet is orthogonal to the glass port. The reason for this is that the light refraction on the flat-port only compresses the laser along its projected line and because the absolute brightness of the laser line is usually not used for the depth estimation, the refraction has no effect onto the depth estimation. However, in the case the laser sheet is not entirely orthogonal to the flat-port interface, the laser sheet will be geometrically distorted. The result is a bent laser sheet in the medium water, altering estimated scene depths. For small angles, this can be compensated using a radial distortion term, which significantly reduces the computational burdens for scene point triangulation. Here, a more in-depth analysis is given in the next chapter.

Furthermore, the light pattern itself, used for illumination, has a significant influence on the backscattering leading to contrast or power-limited systems. Here the illumination schemas with the lowest contrast are areal illuminations while point illuminations are the ones with the highest. A line-pattern can be seen as somewhere in the middle between both extremes allowing to recover multiple 3D scene points at once while having less contrast than point patterns. An in-depth analysis between different illumination schemas for active underwater sensing is given in the work from Jaffe [1988], Jaffe [2010], and Ouyang and Dalgleish [2012] including simulations of the optical path in the medium water.

In addition, most light projectors have very little power in comparison to electromagnetic radiation received at the Earth's surface due to sunlight. This in-balanced radiation is especially the case if the power of the structured light projector is distributed along a line or other light patterns generated by none collimated light beams. In shallow water, this leads to a reduced signal to noise ratio, and the performance of underwater structured light systems is usually regarded as profoundly affected in the first $20m$ of the water column during the daytime, as discussed by Brignone et al. [2011]. The same is true for situations where the structured light system is used for areas subject to intense external lights such as lights from remotely operated vehicles or artificial illumination of harbor piers. In the case of dark objects, this is even more prominent because the structured light system must illuminate all scene points sufficiently enough to measure a return value and to be able to triangulate the scene depth. Otherwise, scene texture dependent holes will be present in the resulting point cloud. Therefore, underwater structured light systems often use high power lasers to compensate for the absorption of the medium water and to be able to measure a return value even for dark objects. However, this high irradiation poses a significant danger to operators with the risk of damaging eyesight if the laser beam is exposed

to the human eye, which is especially problematic during calibration. Here, it is often required to perform the calibration after the system is mounted on the carrier system. Therefore, different strategies are used to reduce this risk. One approach is, for example, to set the laser in a low power mode during calibration like performed by Lucht et al. [2018] or the use of an additional stereo camera system to do underwater on sight calibration described by Roman et al. [2010].

Underwater Applications

Underwater structured light systems are mainly used to perform high-resolution seabed mapping from low altitudes or scanning static scenes to measure the three dimensional shape of pipes, chains, or foundations. It is also often used to document underwater archaeological sites, and shape changes of underwater habitats. Here, mainly, two survey strategies exist. In the first case, a pan-unit equipped with a line-structured light system is mounted onto a tripod, and a full scene scan is performed before moving the tripod to a new position. This approach usually gives the best results because multiple measurements can be aggregated without any sensor pose change in between. Still, it is also very time consuming and requires intervention capability. In the second case, a line-structured light system is mounted onto a moving platform, and either a pan-unit is used to sweep the line laser across the scene while the platform is hovering or the whole platform is moved. In both cases, the motion of the platform has to be compensated with an accuracy matching the sensor resolution of the structured light system, or the performance will significantly be reduced, posing a major challenge for this kind of operation discussed in Duda et al. [2016]. Other structured light patterns than line patterns can be mainly found in research laboratories as they are usually not robust enough for real-world scenarios. An overview of underwater structured light methods used in different research work is given in Tab. 3.3.

Strengths and Weaknesses

A structured light system is a cost-effective high-resolution 3D sensor that can be composed of off the shelf components. Its main advantage is its relatively high depth resolution for close ranges, which can be easily controlled by selecting a suitable baseline without adding any additional constraints on the hardware like timing or sensor resolution. However, this makes it unsuitable for longer ranges because to maintain a constant resolution over its working range, the baseline would have to be

Light Coding	Pattern/Code	Authors
Temporal Coding	Binary/Gray Code	Bräuer-Burchardt et al. [2015] Bruno et al. [2011] Bianco et al. [2013]
	Phase Shifting	Bräuer-Burchardt et al. [2015]
Spatial Coding	Line - Pattern	Lucht et al. [2018] Fan et al. [2017] Chi et al. [2016] Smart et al. [2013] Inglis et al. [2012] Brignone et al. [2011] Roman et al. [2010] Liu et al. [2010]
	De Bruijn Patterns	Kuo and Nobuhara [2017]
	Pseudo-random Patterns	Ouyang and Dalgleish [2012]

Table 3.3: Survey: Underwater structured light methods.

quadratically scaled, which would lead to very bulky setups. But, this argument is less prominent for the underwater domain because the system is usually contrast or power limited after $10m - 20m$, and proper operations ranges are generally below $10m$ limiting the maximal useful baseline. In any case, because the camera and the projector are spatially separated, they are also facing perspective distortion, including shadowed areas which either camera rays and or the light pattern cannot reach. The results are holes in the 3D reconstruction, which are more prominent if the baseline is increased with respect to the observed scene depth.

The real strength of structured light systems is that the light pattern minimizes backscattering in comparison to areal illumination resulting in higher contrast. Also, the pattern acts as artificial scene features that can be more accurately be located in the image domain than natural features, usually resulting in overall higher accuracy and working distance than passive vision systems. Also, due to the projected light pattern, featureless image regions can be reconstructed, which often occurs for painted human-made objects and areas where fine dust is covering the seabed.

However, any additional light source, such as the sun or an underwater light, has a negative influence on the performance, which is especially the case for shallow water during the daytime. Also, there is currently no underwater off the shelf structured light system available, which can estimate their sensor pose based on their sensor readings, and an additional navigation system is required when mounted onto a moving platform.

3.4 Time of Flight

Time of flight systems consist of an emitter and a receiver, which is capturing the emitted signal after reflected by an object. By measuring the period, the signal requires to travel from the emitter to the object, and back to the receiver, the distance between the sensor and the object can be determined. Here, depending on how this period is measured, they can be grouped into direct- and indirect-time of flight systems. In the case of direct-time of flight, a timer starts counting after the signal is emitted and stops upon receiving the signal requiring a timing accuracy in the order of picoseconds. To relax this, indirect-time of flight systems measure a quantity which is proportional to the real period but which is easier to be measured with high accuracy.

Scene Depth Estimation

A time of flight system is based on measuring the time a signal takes to travel to an object and back to the sensor system. By knowing this period Δt and the velocity v of the signal, the distance from the sensor to the object can directly be calculated by Eq. 3.8.

$$z = \frac{v\Delta t}{2} \tag{3.8}$$

In the case of light, the velocity of the signal is the speed of light c for the corresponding medium the signal is traveling in. For vacuum, this translates to $c = 299792458m/s$ requiring a timing resolution of a few picoseconds to be able to have a depth resolution in the mm range. These pulse-based, or linear-mode, systems are therefore usually limited by how accurate the timespan Δt can be measured, taking all delays due to electrical components and their signal propagation into account.

The direct-time measurement is therefore often substituted by a measurement via interferometry on modulated light, which is proportional to the distance traveled. This indirect-time measurement is achieved by using the phase of a modulated laser beam to calculate the distance to the object as a fraction of the modulated wavelength of the signal. In the continuous wave mode, the phase difference between the radiated and receiving signal can be calculated from the relation between four electric charge values C_1 to C_4, which are collected with a 90 degree phase delay to each other as

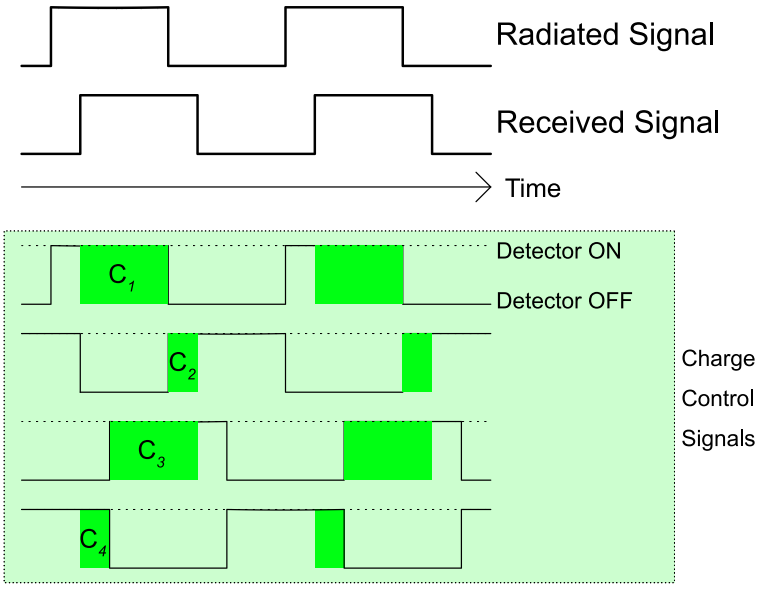


Figure 3.10: Timing for collecting the four electric charge values C_1 to C_2.

shown in Fig. 3.10. Based on these four measured values, the phase can be calculated with the help of Eq. 3.9 (Remondino and Stoppa [2013]).

$$\rho = \arctan\left(\frac{C_3 - C_4}{C_1 - C_2}\right) \tag{3.9}$$

The corresponding distance z can then be calculated, using the speed of light c and the signal frequency f:

$$z = \frac{c}{2f}\frac{\rho}{2\pi} \tag{3.10}$$

Here, because the measurement is based on the phase which wraps around every 2π, the derived measurement will also have an aliasing distance. The distance is called the ambiguity distance and defined by the first term of Eq. 3.10 which is $c/(2f)$.

In addition to the continuous wave mode, the phase can also be estimated using pulsed modulation, which only requires to measure two electric charge values instead of four. Still, the latter method automatically reduces the effect of constant offsets and constants gains due to attenuation or object color.

These phase-shift-based indirect-time of light systems are closely related to pure optical interferometry and share the same principle of operation but use intensity modulation of the light signal at radio frequency to obtain a much larger unambiguous range at the cost of achievable resolution (Heredia Conde [2017]).

Underwater Challenges

In the simplest case where the light source is a laser beam located behind a flat-port, the system will face no geometric distortion due to an orthogonal angle to the different layers. However, the refraction index n will change the speed of light c_{medium} according to:

$$c_{medium} = \frac{c_{vacuum}}{n} \tag{3.11}$$

This effect can easily be compensated in post-processing by a constant distance offset and a scale factor according to the refraction index, assuming a constant refraction index across the penetrated water column. In the case, the incoming or receiving signal is no longer orthogonal to the glass water interface, the measurement will face a geometric distortion, and the X, Y coordinate will be altered while still having a correct Z coordinate.

The primary challenge can be seen in an increase in unexpected interference between two or more optical signals due to multipath effects. Notably, in case of light scattering, due to low-quality optics or high water turbidities, the assumption made for deriving the phase difference between the radiated and the receiving signal with the help of two or more charges no longer holds. The reason for this is that the backscattering of the signal will offset the measured charges in a nonlinear fashion and also might drive the detector into saturation.

In the case of direct time of flight systems, backscattering makes it challenging to determine at which point in time the signal is returned from the object and not from some particle in the water column. Here, the main issue is that a wrong triggering puts the timer into a mode neglecting all further detections for the period of the light pulse. Even in the case, the timer would be able to deal with multiple detections; each detection would disable the detector for a short period until its initial state is restored. In the case of single-photon detectors, this period is usually a few nanoseconds which translate to a dead time of several centimeters during which the sensor can no longer detect a signal (Vornicu et al. [2014]). A solution for this could be gated viewing in an iterative fashion, which would, however, dramatically reduces the acquisition rate.

Another challenge is the electromagnetic absorption of water requiring high power lasers, which can be precisely pulsed. This requirement significantly increases the system costs and is currently the limiting factor for these systems, which are usually power limited rather than contrast limited.

Underwater Applications

In general, scanning time of flight systems have one of the most extended ranges of all commercially available visual underwater systems. Therefore, they are mainly used to measure the pose and dimension of larger underwater installations. Here, the sensor is usually linked to the seafloor, and the scene is statically scanned. After each scan, the position of the system is altered, and subsequent scans are merged using, for example, iterative closest point algorithms or manually selected control points. Other applications are seabed mapping using remote or autonomous underwater vehicles. However, usually due to its limited opening angles, in comparison to cameras, multiple units must be combined to reach comparable coverage. An overview of time of light systems used in the underwater domain is given in Tab. 3.4.

Method	Application	Authors
Range Gated Imaging	Tracking of animals	Dubrovinskaya et al. [2018]
	Dense 3D data	Mariani et al. [2018]
	Image enhancement	Church et al. [2014] Shen et al. [2008]
LiDAR	Dense 3D Data	McLeod et al. [2013]
	Marine life classification	Dalgleish et al. [2017]
	Seafloor mapping	Jaffe et al. [2001] Jaffe [2015]

Table 3.4: Survey: Underwater time of light systems.

Strengths and Weaknesses

For directly measuring the time of flight as well as for measuring the phase-shift using charge counters, high-speed electronics are needed since achieving one-millimeter accuracy in water requires a timing resolution of around ten picoseconds. This level of accuracy is hard to accomplish at room temperature using any silicon technology pointed out by Remondino and Stoppa [2013] limiting the achievable resolution usually to several mm.

Furthermore, several sources of noise exist. Here some of the most important noise sources are the optical shot noise, the thermal noise, flicker noise, and the quantization noise. Therefore, in the case of phase-shift-based systems using a photonic mixing device (PMD), which most of the time of flight cameras do, the precision is limited to around $1cm$ at $20MHz$ modulation frequency (Heredia Conde [2017]).

In the case of marine applications, the wavelength of the system has to lie in the visible spectrum of light, or strong absorption will take place, limiting the possible range to several centimeters (Anwer et al. [2017]). This absorption limits the use of standard commercial systems because most of them are using near-infrared for illumination to take advantage of characteristics of the sunlight spectrum after passing the earth's atmosphere and to be non-disturbing to the human eyesight. Here, the required high power visible laser is one of the most expensive components of an underwater LIDAR system.

However, LIDAR systems use by design a single laser beam minimizing the backscattering while the signal strength is maximized. In combination with gated viewing, this allows for the maximal possible range of visual systems without putting excessive constraints onto the water body. Also, unlike triangulation-based systems, time of flight systems require no spatial separation between the signal emitter and the receiver, and very compact pre-calibrated systems can be achieved. Furthermore, these systems have nearly depth independent depth resolution with minimal shadows, making them especially suited for longer ranges where other optical systems would be bulky or are contrast-limited.

3.5 Comparison Sensor Technologies

This comparison between the different optical sensor technologies shown in Fig. 3.1 is mainly focused on the underwater case. For in-air applications several comparisons through out the literature exists like for example given by Bosch et al. [2001], Jain [2003], Li [2014], Sarbolandi et al. [2015], or Kovacovsky [2017] and which are summarized in Tab. 3.5.

For underwater applications, the environmental conditions reduce in general the performance of every optical system. The reason is that additional noise enters the system while the signal is weakened due to the medium water and potential particles floating in the water column. However, not all sensing technology are impacted the same way, and there are regions where each of the sensor technology has advantages

Consideration	Stereo Vision	Sfm	Structured Light	Indirect ToF	Direct ToF
Software Complex.	Medium	High	Medium	Low	Low
Material Costs	Medium	Low	Medium	Medium	High
Compactness	Low	High	Low	High	High
Response Time	Medium	Low	Medium	High	High
Depth Accuracy	Low	Low	High	Medium	Medium
Low Light Perf.	Low	Low	High	High	High
Bright Light Perf.	High	High	Low	Medium	Medium
Scene Dependency	Medium	High	Low	Low	Low
Range in-air	Low	High	Low	Low	Medium

Table 3.5: Comparison of optical methods for range measurements in air.

in comparison to the others. Also, some of the effects can be compensated using specialized hardware and or software.

The most critical parameters for underwater optical range measurements are similar to the one from in the in-air case. However, due to additional efforts for protecting sensor systems against the medium water and the usage of usually stronger light emitters to counteract the signal loss in the medium the material costs are considerably higher for all regarded systems. Also, external lights have a significantly stronger effect on the performance of active systems, which must have their signal emitter in the visible spectrum to stay in the narrow window of transparency of water. However, the comparison from Tab. 3.5 stays valid when tuned to the visible spectrum of light while Tab. 3.6 extends it with additional considerations for the underwater case, which also includes the dependency on external navigation systems as a static measurement in the water column often cannot be maintained or is too costly.

Consideration	Stereo Vision	Sfm	Structured Light	Indirect ToF	Direct ToF
Calibration Effort	High	Medium	High	Low	Low
Distortion Complexity	High	High	Medium	Low	Low
Dep. on Turbidity	High	High	Medium	High	Low
Dep. on Known Pose	Low	Low	High	Medium	High
Range in Water	Low	Low	Medium	Low	High

Table 3.6: Comparison of optical methods for range measurements in water.

Also, the payload volume, weight, and power available on sensor carriers such as remotely operated vehicles have to be taken into account. But, because in the underwater area, tightly integration solutions are still barely available and considerable weight comes from the underwater housings, the benefits for one or the other technique are less prominent than for in-air applications.

As a summary, Fig. 3.11 gives some guidance where each technology has the highest performance for a specific set of mission constraints. Consequently, no technology

is delivering even close the performance underwater, which is achievable for in-air applications. However, in the absence of alternatives, they still can provide viable information about the surrounding. Also, it is this lack of performance which makes it more and more vital to use multi-modal approaches to maximize the performance and to extend the working area of such sensor systems in challenging conditions.

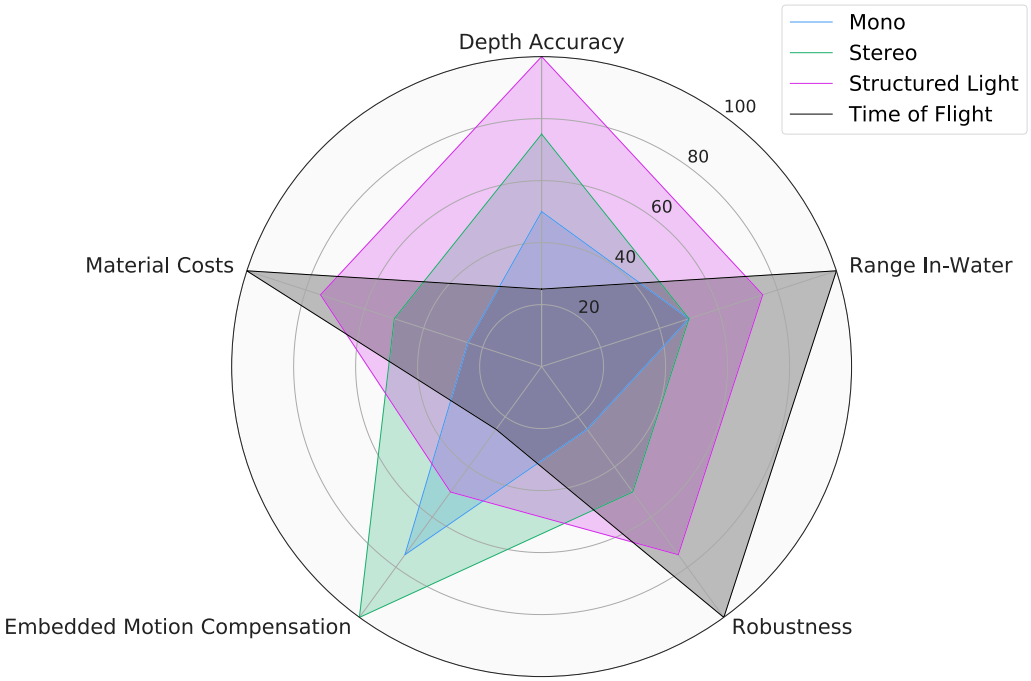


Figure 3.11: Guideline underwater optical sensing.

In case a single technology has to be deployed to accomplish a given task. The following considerations can be used to identify the technology with the best possible performance. For measuring a static object from a distance exceeding $6 - 10m$ the best performance can be usually achieved with time-of-flight systems rigidly mounted on a tripod. Structured light systems typically outperform all other methods for closer static scenes, if the water turbidity does not lead to a contrast limitation of the signal. For seabed/object mapping from closer than around $3m$ distance using a moving platform, mono and stereo systems usually, deliver the best performance as they also nativity estimate the pose of the system as an embedded processing step. Here, many different software suites are available to transform an image sequence into a consistent 3D reconstruction after data acquisition using dome port cameras. In

case an external pose estimation is available, structured light systems usually deliver the best near to mid-range performance. Whereas, time-of-flight systems typically deliver the best far range performance.

Chapter 4

Modeling the Optical Path

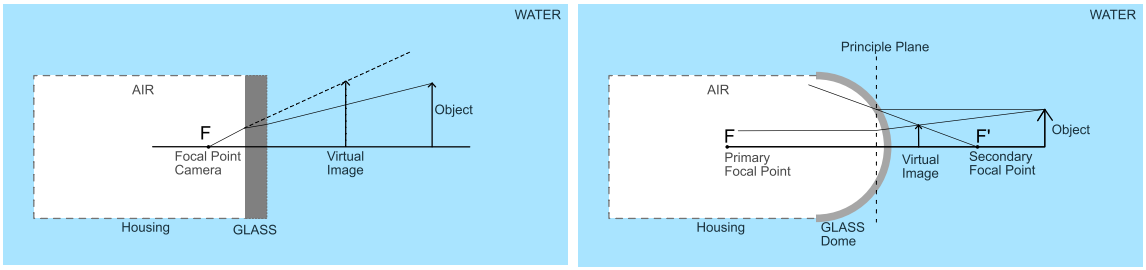
Optical systems are usually designed to operate in-air or vacuum and optimized to be approximated by linear light propagation models. These linear models are the base for projective geometry and, for example, are heavily used by triangulation based systems. Therefore, special care has to be taken if transferred to the underwater domain where these models might be invalidated due to light refraction.

4.1 Sensor and Emitter Housings

From a modeling point of view, there is no difference in which direction the light traverse through an optical system. Therefore, a light projector can also be considered as an inverse camera. However, it has to be differentiated between solving a direct or an inverse problem, which might be even impossible to solve depending on the nature of the problem. Here, tracing the path of light using ray-tracing techniques and projecting a 2D point into 3D space using a refractive camera model is considerably less complicated than calculating the inverse. The reason for this is that calculating this inverse involves estimating which 2D image point belongs to a 3D scene point located behind multiple transparent layers such as glass without explicitly knowing the ray vector in each layer.

Flat-Ports

Underwater flat-port cameras use a flat transparent window in front of the camera lens to seal the camera against the environment. This additional layer introduces a considerable change in the refraction index when a light signal parses the different media. The result is geometric and chromatic distortion depending on the angle of the light ray with respect to the different layer normals. However, flat-ports are



(a) Flat-Port Camera (b) Dome-Port

Figure 4.1: Schematic of an underwater flat-port camera and dome-port.

usually quite inexpensive and can be easily mathematically modeled. Therefore, many underwater systems rely on flat-port cameras, and even without modeling the light refraction, they produce satisfactory results in many scenarios. As a summary, the following characteristics should be regarded when choosing a flat-port. A more in-depth consideration is given in Section 4.2:

- Objects appear around 25% closer in water than in-air.
- Refraction can be minimized by placing the camera axis orthogonal to the glass layer and the focal point as close as possible to it.
- Different wavelength facing different refraction leading to chromatic distortion.
- The image center is less affected than outer image regions.
- Standard lens distortion models work only well if calibrated to a fixed object distance.

Dome-Ports

Hemispherical dome-ports consist of two spherical surfaces that have the same center of curvature. Its thickness is defined by the difference between the two radii of these two spherical surfaces. In the case, the focus point of a camera is precisely located in the center of the dome; all camera rays will enter the dome-port orthogonal to its spherical surfaces and will face no refraction. However, this requires a highly accurate manufacturing process and is more demanding and expensive than the production of flat-ports. Also, due to the concave nature of the dome-port, an underwater object is always imaged as a smaller virtual object in front of the dome glass in several centimeters away while being submerged visualized in Fig. 4.1(b). In-air, a dome-port does not have this effect, and objects appear at the correct distance. Therefore, it is difficult to set the correct focus during assembly, especially for small dome-ports.

It is worth to note that a dome-port also produces a curved field of focus, reducing the effective depth of field. This curved field of focus is especially problematic for

planar structures that are blurred at their edges. To counteract this effect, usually underwater dome-port cameras are using a large F-Stops, but this also reduces the amount of light the sensor can collect. Nonetheless, they are widely used for many marine applications ranging from photogrammetry to video documentation because they can preserve the main geometrical characteristics of standard lenses supported by most software packages. Here, small miss-alignments can usually be compensated using standard lens distortion models, which is not the case for flat-port cameras requiring an explicit refractive camera model (Nocerino et al. [2016]). As a summary, the following characteristics should be regarded when choosing a dome-port:

- The camera focus point should be located in the center of the dome.
- The focus point of a lens is changing when using zoom lenses.
- The focus should be set to macro mode and is depending on the dome radius.
- The field of focus is curved requiring large F-Stops to maintain a sufficient sharpness.
- Small miss-alignments can be modeled by standard lens distortion models.
- Larger dome-ports are less prone to miss-alignments, focus and sharpness issues.

Special Underwater Lenses

There are only a limited number of underwater wet lenses that take the refraction of water into account as their last lens element. They are considered to deliver the sharpest images with minimal distortion. One very famous camera series was the Nikonos from Nikon, supporting special underwater lenses, which only shoot sharp images underwater. However, the series was discontinued in 2001 in favor of cheaper underwater accessories for regular cameras (Camera-wiki [2015]). Nowadays, a standard camera, including a lens, is usually housed in a waterproofed housing with a dome or flat-port. In the case of flat-ports special waterproofed conversion lenses can be added to change the focal length and to improve the underwater performance.

Comparison

Flat-port cameras, as well as dome-port cameras, have advantages for specific applications and are widely used for underwater applications. Although special underwater lenses can deliver sharper images across the whole image, they are less often used due to their price and availability. A comparison between these three lens types is given in the following table.

Considerations	Flat-Port	Dome-Port	Underwater Lens
Work Principle	Flat transparent window for sealing the camera including its lens	Concentric lens acting as additional optical element after the camera lens	Assembly of multiple lenses with the water considered as last element of the assembly
Field of View	Reduced by around 25% in comparison to in-air. Limited to 97° due to total reflection	More than 180° possible	Usually <100°
Focal Length	Increased by around 25%	Same as in-air	Designed for submerged only
Depth of Field	Compression of around 25%	Curved reducing effective Depth of Field	According to lens specification
Low Light Conditions	Small F-Stops are supported as required for low light condition	Large F-Stop is needed due to curved depth of field and good illumination is required	Low F-Stops are supported required for low light condition
Geometric Distortion	High	Low	Low
Chromatic Distortion	High	Medium	Low
Assembly	Camera axis should be orthogonal to the glass layer. The focus point should be as close as possible to the window to minimize distortions.	The camera-lens should be focused to a smaller virtual image in front of the dome port. Focus point of the camera lens should be in the center of the dome.	Must be manufactured by a specialized lens company.
Calibration	Refractive camera model in combination with standard lens distortion model. In the case of fixed object distance standard lens distortion model might be sufficient.	Standard lens distortion model	Standard lens distortion model
Price	Low	Medium	High
Applications	Photogrammetry Structured Light Video / Photo Footage	Photogrammetry Photo mosaique Video / Photo Footage Stereo Vision	Limited availability Video / Photo Footage

Table 4.1: Comparison between different underwater camera ports.

4.2 Camera Models

An optical camera projects a 3D scene onto an image sensor composed of many light-sensitive elements. Each of these elements collects light originated under a specific solid angle. The relation between this solid angle and the location of a sensor element inside the image sensor is established with the help of camera models.

Pinhole Camera Model

One of the widely used camera models is the pinhole camera model. It establishes a projective relationship between the location of light-sensitive sensor elements, also referred to as pixels and the solid angles which get projected onto them. The pinhole camera model is the base for many machine vision algorithms. There are many areas where currently no equivalent algorithms exist for other camera models, such as the five-point algorithm for estimating the relative pose between two camera positions based on five corresponding scene points (Nistér [2004]). Therefore, other camera models, such as omnidirectional camera models (Kannala and Brandt [2006], Micusik and Pajdla [2003]), are often solely used to correct the raw image to be able to use standard pinhole camera pipelines afterward. However, the vision community is currently in the process of generalizing structure from motion algorithm for other types of cameras as well. Here a library for solving calibrated central and non-central geometric vision problems is, for example, the OpenGV library by Kneip and Furgale [2014]. Looking at standard pinhole camera models, they can be characterized as finite projective cameras that have some additional parameters. These parameters augment the most basic pinhole camera model, which maps a 3D point $(X, Y, Z)^T$ to a corresponding 2D point by simply projecting it onto a virtual plane $1m$ away from the origin.

$$(X, Y, Z)^T \mapsto (X/Z, Y/Z) \tag{4.1}$$

The extended model takes also the focal length f_x and f_y and the principal point (image center) c_x and c_y into account.

$$(X, Y, Z)^T \mapsto (f_x X/Z + c_x, f_y Y/Z + c_y) \tag{4.2}$$

This equation can also be rewritten in homogeneous coordinates resulting in:

$$\begin{pmatrix} f_x X + Z c_x \\ f_y Y + Z c_y \\ Z \end{pmatrix} = \begin{bmatrix} f_x & & c_x & 0 \\ & f_y & c_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{4.3}$$

Here, the so called calibration matrix K encapsulates the intrinsic camera parameters and is defined as:

$$K = \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \tag{4.4}$$

In this case, the skew factor s was added for completeness which is usually zero for most modern digital cameras. Based on this, Eq. 4.3 can be rewritten with the homogeneous coordinates for a 3D point X and its projection x as:

$$x = K[I|0]X \tag{4.5}$$

Finally, in the case, also the camera frame orientation R and the camera center C in world-coordinates are given the model extends to:

$$\begin{aligned} x = & \quad KR[I|-C]X \\ x = & \quad PX \end{aligned} \tag{4.6}$$

This linear model, visualized in Fig. 4.2, is the general mapping of a pinhole camera with P referred to as the camera matrix. The 3×4 camera matrix has 11 degrees of freedom, which is identical to a 3×4 matrix defined up to an arbitrary scale. Six of the free parameters encapsulate the pose of the camera and five the intrinsic camera parameters. Based on this camera matrix, any 3D point can be forward projected onto the camera image by simple matrix multiplication. Here, also the backward-projection, mapping a 2D point to its dedicated camera ray $X(\lambda)$, can be written in matrix form as:

$$\begin{aligned} X(\lambda) &= C + \lambda R^{-1} K^{-1} x \\ &= C + \lambda \vec{v} \end{aligned} \tag{4.7}$$

Due to this linearity for forward and backward-projections, the pinhole camera model is widely used in many computer vision applications. Here, its parameterization used in common machine vision applications is given in Tab. 4.2 not taking any lens distortion into account explained in the next section.

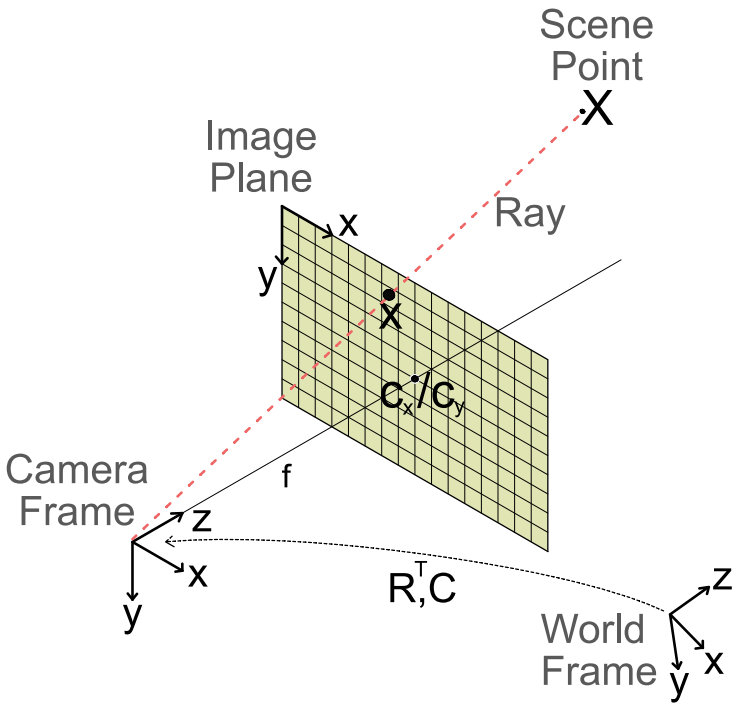


Figure 4.2: Mapping of a scene point X onto its corresponding image point x using a pinhole camera model.

Symbol	Description
f_x, f_y	Focal length
p_x, p_y	Principal point
R	Camera Orientation
C	Camera Center

Table 4.2: Parameters of a standard pinhole camera model.

Lens Distortion Model

Real cameras usually consist of multi-layer none pinhole lenses to increase their light sensitivity. These additional elements in the optical path usually generate a deviation from the ideal projective mapping between 3D points and their corresponding projections. To compensate this deviation a software correction can be applied which is often done using Brown's distortion model (Brown [1971]). Based on this model the 2D image points are re-mapped to new positions which are distortion free. This mapping from a distorted point (x_d, y_d) to its undistorted version (x_u, y_u) as projected by an ideal pinhole camera is defined by the following equation with the principal point (c_x, c_y), the radial distortion coefficients K_n, the tangential distortion coefficients P_n, and the radial distance $r = \sqrt{(x_u - c_x)^2 + (y_u - c_y)^2}$:

$$\begin{aligned} x_d = &\quad x_u + (x_u - c_x)(K_1 r^2 + K_2 r^4) + (P_1(r^2 + 2(x_u - c_x)^2) + 2P_2(x_u - c_x)(y_u - c_y)) \\ y_d = &\quad y_u + (y_u - c_y)(K_1 r^2 + K_2 r^4) + (2P_1(x_u - c_x)(y_u - c_y) + P_2(r^2 + 2(y_u - c_y)^2)) \end{aligned} \tag{4.8}$$

The benefit of this approximation is that it is independent of the actual scene depth and can be pre-computed and efficiently applied to all raw images. After this, subsequent algorithms directly can take advantage of the projective mapping defined by the pinhole camera model. Taking this additional radial and tangential distortion into account, the pinhole model expands to the normally used parameterization listed in Tab. 4.3.

Symbol	Description
f_x, f_y	Focal length
p_x, p_y	Principal point
K_1, K_2	Radial distortion
P_1, P_2	Tangential distortion
R	Camera Orientation
C	Camera Center

Table 4.3: Parameters of a standard pinhole camera model including lens distortion.

Here, also other lens distortion approximations are available, which allow more compact representations, especially for certain types of lenses like wide-angle lenses or lenses mainly subject to radial distortion, for example, addressed in Tang et al. [2017].

Refractive Camera Model

A refractive camera model is a mathematical model taking the refraction of light rays at interface boundaries explicitly into account and can correct deviation from the projective relation between the location of light-sensitive sensor elements and the solid angles which get projected. In the case of underwater cameras, usually, flat-ports are modeled using three layers leading to two defined interfaces between them. Here, the first interface is between the air layer, the camera lens is located, and a flat glass port. Following this, the second interface is between the glass port and the medium water on the outside of the camera system.

The refraction of light at these interfaces due to different refractive indices, ergo, air, glass, and water can be described using Snell's law (Eq. 4.9). For this, the refraction indices of both media as well as the incoming light ray angle with respect to the interface normal must be known. Also, the location where the refraction takes place must be known to not only calculate the direction of the refracted light ray but also the exact path. Assuming an underlying pinhole camera, this is the case when the position, orientation, and thickness of the glass layer is known with respect to the camera center. This consideration leads to the required minimal parameter set of a refractive camera model given in Tab. 4.4 extending the standard pinhole camera model with eight additional parameters. This general refractive model for flat-port cameras supporting thick glass ports with tilted interfaces is also used by Agrawal et al. [2012] and Jordt [2013] and is visualized in Fig. 4.3.

Symbol	Description
f_x, f_y	Focal length
p_x, p_y	Principal point
K_1, K_2	Radial distortion
P_1, P_2	Tangential distortion
$\vec{n}$	Interface normal
n_a, n_g, n_w	Refraction index for air, glass and water
d_a, d_g	Thickness of the air and glass layers
R	Camera Orientation
C	Camera Center

Table 4.4: Parameters of the refractive camera model.

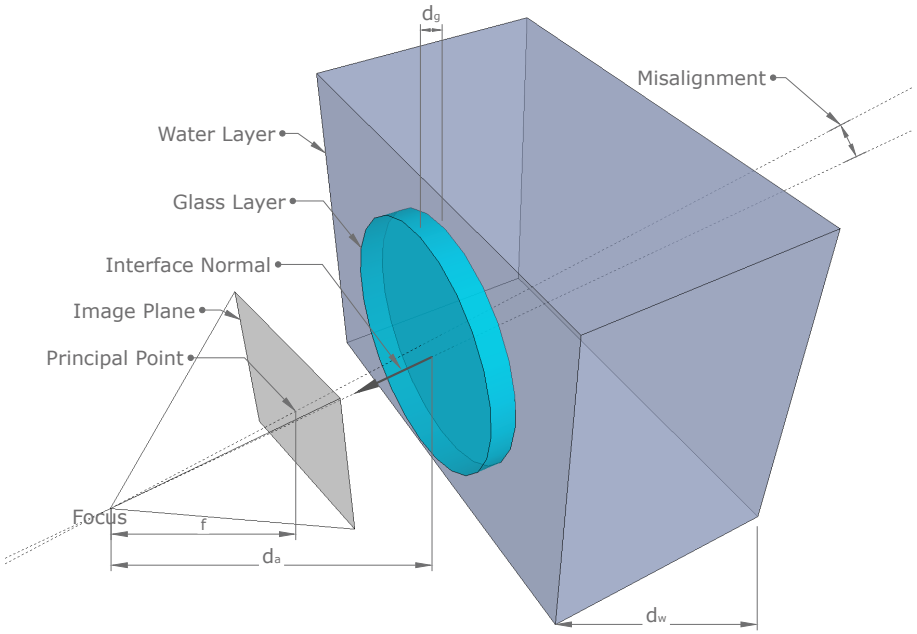


Figure 4.3: The flat-port camera model.

Back-Projection

Based on the parameterization listed in Tab. 4.4 and Snell's law in vector form, back projecting any image point onto its corresponding 3D ray, can be achieved in a straight forward manner. In the first step, the light ray captured by the camera sensor at location p is mapped to its corresponding 3D ray $P(\lambda)$ in-air using the projective relation between it and the sensor location p established by the pinhole camera model. In general, this is referred to as the back-projection of a 2D point to 3D space. Following this, the ray $\vec{v_a}$ is refracted when entering the glass layer. This refracted ray $\vec{v_g}$ in the glass layer can be derived using Snell's law in vector form. Here, n_a, n_g are the refractive indices, and $\vec{n}$ is the plane normal of the glass layer with respect to the camera orientation assuming a planar glass window.

$$v_g = \frac{n_a}{n_g}\left[\vec{n} \times (-\vec{n} \times \vec{v}_a)\right] - \vec{n}\sqrt{1 - \frac{{n_a}^2}{n_g}(\vec{n} \times \vec{v}_a) \cdot (\vec{n} \times \vec{v}_a)} \tag{4.9}$$

Analog to this, the ray vector $\vec{v_w}$ in the medium water can be computed by:

$$v_w = \frac{n_g}{n_w}\left[\vec{n} \times (-\vec{n} \times \vec{v}_g)\right] - \vec{n}\sqrt{1 - \frac{{n_g}^2}{n_w}(\vec{n} \times \vec{v}_g) \cdot (\vec{n} \times \vec{v}_g)} \tag{4.10}$$

However, this only calculates the direction of the ray in each of the layers. To calculate the exact path, also the entry and exit points of the ray into and out of the

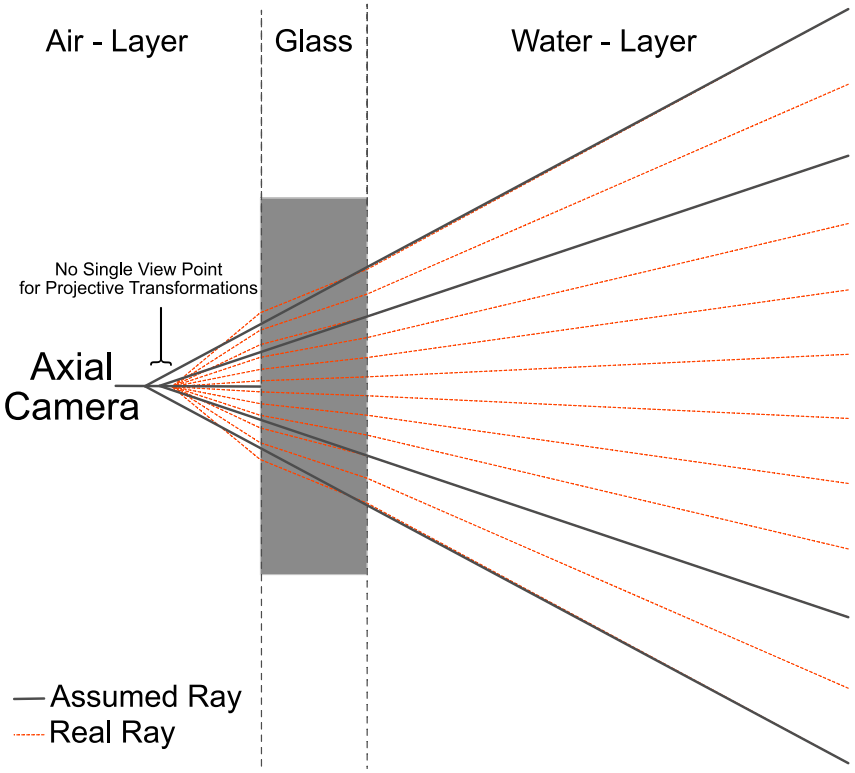


Figure 4.4: A pinhole camera converts to an axial camera due to a flat-port housing.

glass must be calculated. Because the camera center defines the starting point, the entry point can be derived by intersecting $P(\lambda)$ with the plane equation for the first glass surface. Analog to this also, the exit point can be calculated, which gives the line equation for the refracted ray in the medium water. In computer graphics, this is also known as eye-based ray tracing, which is a rendering technique for generating an image by shooting rays from the eye to the light source and simulating the effects of its interactions with virtual objects and media.

Forward-Projection

The forward-projection of 3D points onto the image plane is considerably more complicated than the back-projection because, unlike for the pinhole camera model, there is no projective relation between a 3D point and its projected version subject to refraction. In fact, due to the additional optical elements, a pinhole camera behind a flat-port converts to an axial camera when being submerged. Here, all camera rays no longer intersect in a common point (center of projection) but a common line visualized in Fig. 4.4. This invalidation of the single viewpoint constraint is a direct consequence from the plane of incidence or plane of refraction (POR), which is defined

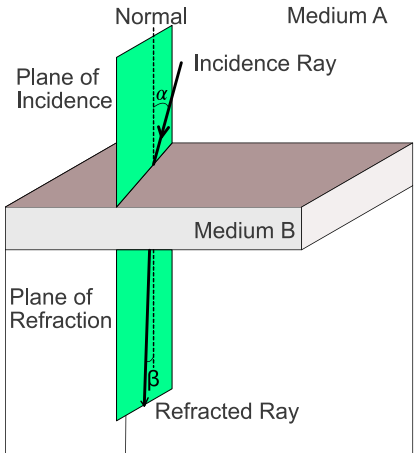


Figure 4.5: An incoming and its corresponding refracted ray always lie in a common plane called plane of incidence or plane of refraction (POR).

by the incoming ray $\vec{v}$ and the surface normal $\vec{n}$ of the interface. Based on Snell's Law, it can be shown that also the refracted ray must lie on this plane visualized in Fig. 4.5. This constraint can further be used to define a coplanarity constraint between each 3D point P and its projection $p = (p_x, p_y)$ which is independent of the actual refraction indices (Agrawal et al. [2012]). This constraint describes the metric distance of a 3D scene point to the POR defined by its projection and the layer normal of the glass layer. It is zero if the light rays intersect with the projection and the 3D scene point at the same time and therefore originated from this point.

$$\vec{n} \times \vec{v} \cdot P = 0 \tag{4.11}$$

$$\vec{v} = \begin{pmatrix} \frac{p_x - c_x}{f_x} \\ \frac{p_y - c_y}{f_y} \\ 1 \end{pmatrix} \tag{4.12}$$

However, the coplanarity constraint cannot directly be applied to global optimization problems because the scale of its residual depends on the scene depth and the location of the projected scene point. Also, it ignores the direction orthogonal to the POR and leaves only one constraint for every 3D point. Therefore, in the following, the re-projection error is approximated using a novel first-order Taylor expansion of the flat refractive forward projection centered around the projection obtained by the pinhole camera model. Here, the measured 2D point can directly serve as the center of the Taylor expansion, making it possible to integrate it seamlessly into non-linear optimization problems such as bundle adjustments. This new approach was already presented in Duda and Gaudig [2016]. Here a new, more compact derivation is given.

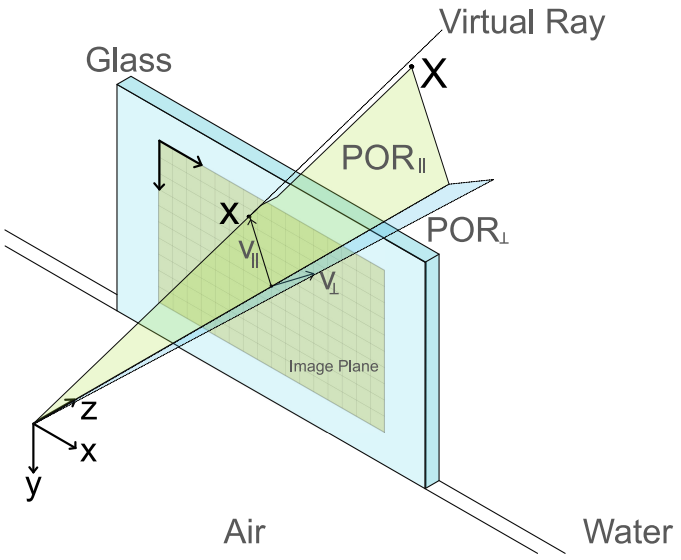


Figure 4.6: The basis change to simplify the refractive calculation.

In general, the flat refractive forward projection can be simplified by converting the three dimensional problem into a two dimensional problem. This simplification is achieved by performing the calculation with respect to the POR containing the whole path of the camera ray, including any refraction. Per definition, the POR contains the layer normal n and the camera ray $\vec{v}$. Therefore, the vector $\vec{v}^{\parallel}$, which is the intersection of the POR with the image plane, is given by:

$$\vec{v}^{\parallel} = \frac{\vec{v} - \begin{pmatrix} \frac{n_x}{n_z} \\ \frac{n_y}{n_z} \\ 1 \end{pmatrix}}{\left| \vec{v} - \begin{pmatrix} \frac{n_x}{n_z} \\ \frac{n_y}{n_z} \\ 1 \end{pmatrix} \right|} \tag{4.13}$$

Any change along this vector will lead to a new associated 3D point, which also lies in the same POR. Therefore, this direction is used as the new x-coordinate in the image domain, and its re-projection error is calculated by only tracking the ray inside the $POR = POR^{\parallel}$. For the y-coordinate in the image domain, a vector $\vec{v}^{\perp}$ orthogonal to $\vec{v}^{\parallel}$ is defined encapsulating deviations from the POR:

$$\vec{v}^{\perp} = \begin{pmatrix} -\vec{v}_y^{\parallel} \\ \vec{v}_x^{\parallel} \\ 0 \end{pmatrix} \tag{4.14}$$

Based on these two vectors, visualized in Fig. 4.6, the basis vectors of the POR

are derived. It is worth mentioning that these basis vectors are only aligned with the image plane if the camera is orthogonal to the glass interface.

$$\begin{aligned} \vec{e}^{\perp} &= \frac{\vec{n} \times \vec{v}^{\parallel}}{|\vec{n} \times \vec{v}^{\parallel}|} \\ \vec{e}^{\parallel} &= \frac{\vec{e}^{\perp} \times \vec{n}}{|\vec{e}^{\perp} \times \vec{n}|} \\ \vec{e}^{n} &= \vec{n} \end{aligned} \tag{4.15}$$

Here, $\vec{e}^{\parallel}$ and $\vec{e}^{n}$ are lying in the POR and $\vec{e}^{\perp}$ and $\vec{e}^{n}$ are lying in a second plane $POR^{\perp}$ which is orthogonal to POR. Based on this, the slope angle α of the camera ray $\vec{v}$ inside the POR is given by:

$$\sin\alpha = \left| \vec{e}^{n} \times \frac{\vec{v}}{|\vec{v}|} \right| \tag{4.16}$$

In the next step, using Snell's Law in standard form, the slope m of the camera ray can be calculated after entering a different medium with n_a being the refraction index of the first and n_b the one of the second medium.

$$\begin{aligned} \sin\beta &= \frac{n_a}{n_b}\sin\alpha \\ \tan\beta &= \frac{\sin\beta}{\sqrt{1-\sin\beta^2}} \\ m(\vec{v}, n_a, n_b) &= \frac{\frac{n_a}{n_b}\left|\vec{e}^{n} \times \frac{\vec{v}}{|\vec{v}|}\right|}{\sqrt{1-\left(\frac{n_a}{n_b}\left|\vec{e}^{n} \times \frac{\vec{v}}{|\vec{v}|}\right|\right)^2}} \end{aligned} \tag{4.17}$$

Using Eq. 4.17 and the refraction indices of the media the slope m of the camera ray in the three different medias air, glass and water is given by:

$$\begin{aligned} m_a(\vec{v}) &= m(\vec{v}, n_a, n_a) \\ m_g(\vec{v}) &= m(\vec{v}, n_a, n_g) \\ m_w(\vec{v}) &= m(\vec{v}, n_a, n_w) \end{aligned} \tag{4.18}$$

Following this, as the real 3D point P must also lie inside the POR it position can be calculated using the distance d_a between the camera center and the glass interface, the thickness d_g of the glass layer and the distance d_w between the 3D point and the glass interface. This circumstance is visualized in Fig. 4.7 and given by the following equation:

$$g(\vec{v}, d_a, d_g, d_w) = d_a m_a(\vec{v}) + d_g m_g(\vec{v}) + d_w m_w(\vec{v}) \tag{4.19}$$

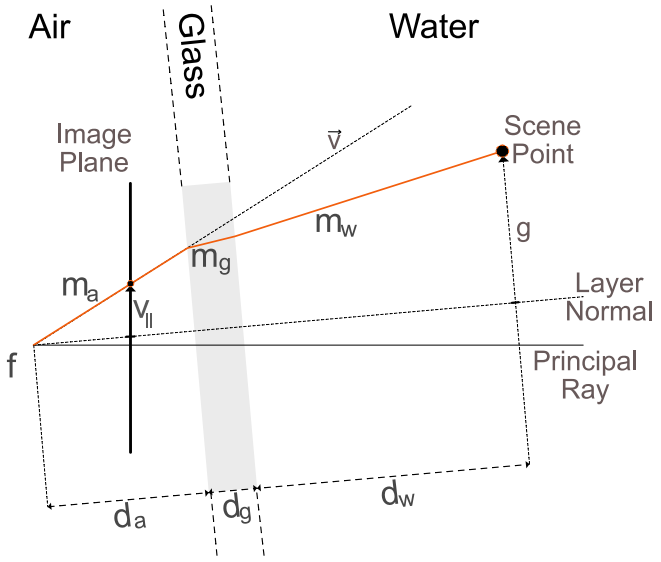


Figure 4.7: Refraction inside the plane of refraction.

By applying a change of basis to transform the 3D point P to the coordinate system of the POR the following constraint can be defined.

$$g(\vec{v}, d_a, d_g, \vec{e}^n \cdot P - d_a - d_g) = \vec{e}^{\|} \cdot P \tag{4.20}$$

Also, per definition, all refracted rays are contained by the POR. Therefore, the following constraint also applies which is, in fact, the normalized coplanarity constraint.

$$g(0, d_a, d_g, \vec{e}^n \cdot P - d_a - d_g) = \vec{e}^{\perp} \cdot P = 0 \tag{4.21}$$

Assuming the re-projection error $\Delta^{\|}$ and $\Delta^{\perp}$ between a projection p_1 of a triangulated 3D point P_1 and its measured image location p_0 is known, the constraint can be rewritten as:

$$\begin{aligned}
\vec{e}^{\|} \cdot P_1 =& d_a m_a(\vec{v_0} + \Delta^{\|}\vec{v}^{\|}) + d_g m_g(\vec{v_0} + \Delta^{\|}\vec{v}^{\|}) + (\vec{e}^n \cdot P_1 - d_a - d_g) m_w(\vec{v_0} + \Delta^{\|}\vec{v}^{\|}) \\
\vec{e}^{\perp} \cdot P_1 =& d_a m_a(\Delta^{\perp}\vec{v}^{\perp}) + d_g m_g(\Delta^{\perp}\vec{v}^{\perp}) + (\vec{e}^n \cdot P_1 - d_a - d_g) m_w(\Delta^{\perp}\vec{v}^{\perp})
\end{aligned} \tag{4.22}$$

This re-projection error is visualized in Fig. 4.8. Here, because the re-projection error is usually unknown, the goal is to solve Eq. 4.22 for $\Delta^{\|}$ and $\Delta^{\perp}$ which is non-trivial in closed form and would lead to a 12th degree equation which was first published by Agrawal et al. [2012]. However, it can also be solved by approximating all slopes m by a Taylor series centered at the projection p_0.

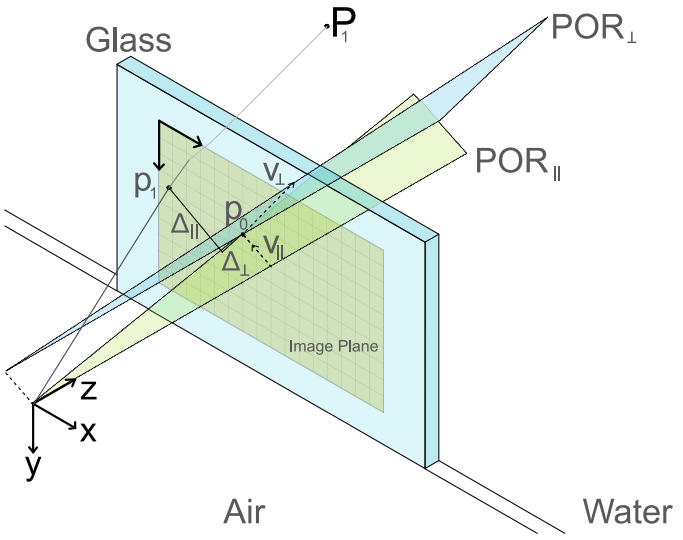


Figure 4.8: Re-projection error of a flat-port camera.

$$\begin{aligned} m(\vec{v_0} + \vec{v}^{\parallel}\Delta^{\parallel}) \approx &\quad m(\vec{v_0}) + m^{'}(\vec{v_0})\Delta^{\parallel} \\ m(\vec{v}^{\perp}\Delta^{\perp}) \approx &\quad m(0) + m^{'}(0)\Delta^{\perp} \end{aligned} \tag{4.23}$$

Plugging Eq. 4.23 into Eq. 4.22 and solving for the re-projection errors while using the fact that due to Snells's Law $m(0) = 0$ results in:

$$\begin{aligned} \Delta^{\parallel} &\approx \frac{\vec{e}^{\parallel}P - m_a(\vec{v}_0)d_a - m_g(\vec{v}_0)d_g - m_w(\vec{v}_0)\cdot(\vec{e}^{n}P - d_a - d_g)}{m_a^{'}(\vec{v}_0)d_a + m_g^{'}(\vec{v}_0)d_g + m_w^{'}(\vec{v}_0)\cdot(\vec{e}^{n}P - d_a - d_g)} \\ \Delta^{\perp} &\approx \frac{\vec{e}^{\perp}P}{m_a^{'}(0)d_a + m_g^{'}(0)d_g + m_w^{'}(0)\cdot(\vec{e}^{n}P - d_a - d_g)} \end{aligned} \tag{4.24}$$

This simplification leads to the final re-projection error in image coordinates given by:

$$\begin{pmatrix} \Delta^x \\ \Delta^y \\ 0 \end{pmatrix} = \Delta^{\parallel}\vec{v}^{\parallel} + \Delta^{\perp}\vec{v}^{\perp} \tag{4.25}$$

However, this equation is an oversimplification because the $POR^{\perp}$ does not go through the real camera center, and a new virtual camera center is generated with a slightly different focal length scaling the re-projection error in the direction of $\vec{v}^{\perp}$. This difference can be compensated by applying the correct scaling factor.

$$\begin{pmatrix} \Delta^x \\ \Delta^y \\ 0 \end{pmatrix} = \Delta^{\parallel}\vec{v}^{\parallel} + \Delta^{\perp}\vec{v}^{\perp}\vec{e_n}\cdot(0,0,1)^{\tau} \tag{4.26}$$

The resulting re-projection error is based on the Taylor series centered at an approximation for the corresponding projection. Therefore, the result diverges from the real value depending on the initial approximation. In addition to this linearization error, there is an additional error resulting from a POR, which does not contain the real 3D point. This deviation happens if there is an error in the direction of $\vec{v}^{\perp}$, which leads to an offset for the estimated basis vectors of the POR, which can be minimized when Eq. 4.26 is used in an iterative fashion.

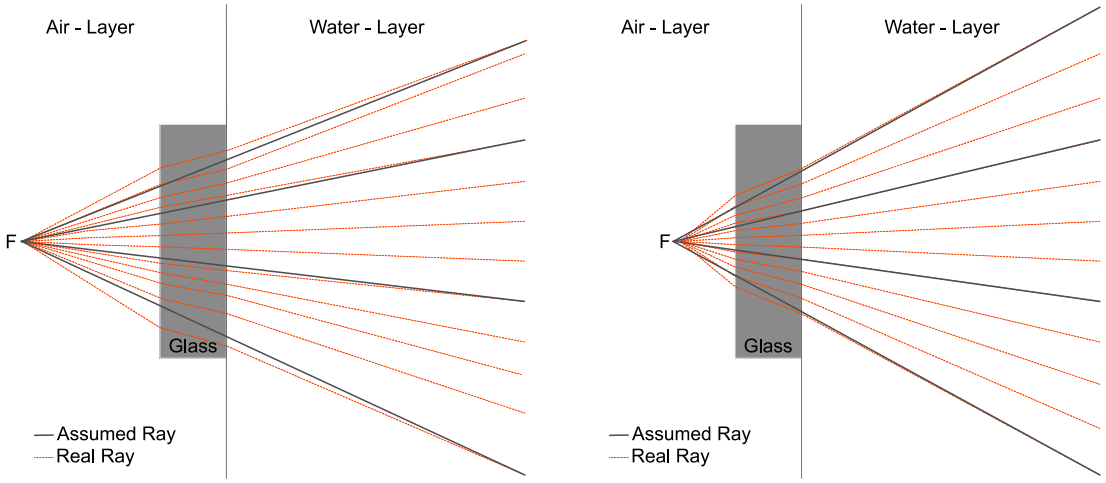
In the following, several experiments verify the accuracy of the proposed approach in dependency on the image location. Assuming a flat-port camera has the intrinsic model parameters displayed in Tab. 4.4 where Flat_1 is an optimal flat-port camera having an image plane that is parallel to the glass interface and an optimal distance to the glass interface which mostly cancels out the effects of refraction. Here, the constraint for cancellation is given below, which can only be fulfilled for a single concrete camera ray vector $\vec{v}$.

$$\begin{aligned} d_a m_a(\vec{v}) + d_g m_g(\vec{v}) &= (d_a + d_g) m_w(\vec{v}) \\ d_a &= d_g \frac{m_w(\vec{v}) - m_g(\vec{v})}{m_a(\vec{v}) - m_w(\vec{v})} \end{aligned} \tag{4.27}$$

However, it should be considered during camera design to minimize the average geometric distortions, as shown in Fig. 4.10 and 4.11. This optimal distance was also identified by Schattschneider [2014]. Here the conclusion is that the camera should be placed as closely as possible in front of the window as the optimal distance is usually below what is physically achievable. A schematic of the optimal placement is visualized in Fig. 4.9 minimizing the overall geometric error.

Following this, Flat_2 is a flat-port camera which is still orthogonal but is shifted away from the glass interface by an additional $5mm$. Finally, Flat_3 is a flat-port camera with a small tilt against the glass layer similar to one which is manually assembled without special equipment for alignment. Analog to this, in Tab. 4.3 the corresponding model parameters for the pinhole camera model are given optimized to minimize the geometric error with respect to the flat-port model assuming a calibration target in $3m$ distance. Here, for the orthogonal cases, the refraction is mainly compensated by changing the focal length and the radial distortion coefficients of the camera. However, in the case, the camera is tilted, this is no longer sufficient, and the image center, focal length, and radial and tangential distortion coefficients must be modified to minimize the error.

Using these model parameters for each pixel of the camera image, the corresponding 3D points are calculated based on refractive back-projection and a virtual plane



(a) Non-Optimal Placement (b) Optimal Placement

Figure 4.9: An optimized placement of the focal point with respect to the glass port minimizes the observed geometric distortion.

Symbol	Flat_1	Flat_2	Flat_3
f_x, f_y	785, 785	785, 785	785, 785
p_x, p_y	640, 480	640, 480	640, 480
$\vec{n}$	[0.0,0.0,1]	[0.0,0,1]	[0.02,-0.01,1]
n_a, n_g, n_w	1,1.77,1.337	1,1.77,1.337	1,1.77,1.337
d_a, d_g	0.005, 0.01	0.010, 0.01	0.005, 0.01

Table 4.5: Parameters of the refractive camera model for different simulated cameras. The first one has an orthogonal camera axis with respect to the glass interface and an optimal distance to the glass interface. The second one is also orthogonal to the glass interface, but its position is shifted by $5mm$ away from the glass interface. The third camera has an optimal distance to the glass interface but has a small tilt angle similar to cameras manually assembled and aligned.

Symbol	$\text{Pin}_1(\text{Flat}_1)$	$\text{Pin}_2(\text{Flat}_2)$	$\text{Pin}_3(\text{Flat}_3)$
f_x, f_y	1047.30, 1047.30	1046.74, 1046.74	1048.69, 1048.57
p_x, p_y	640.00, 480.00	640.00, 480.00	644.95, 477.47
K_1, K_2, K_3	0.395, 0.182, 0.282	0.394, 0.182, 0.276	0.368, 0.358, -0.038
P_1, P_2	-0.578e-6, -0.556e-6	-0.559e-6, -0.532e-6	-0.00444, 0.00914

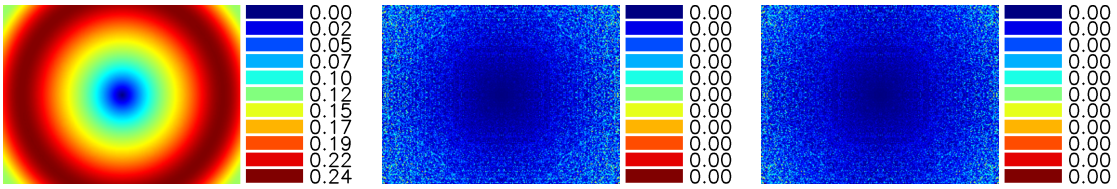
Table 4.6: Parameters of the pinhole camera model optimized to minimize the re-projection error of the corresponding flat-port model given in Tab. 4.4.

in a scene distance of $1m$. After this, each 3D point is forward projected onto the image plane using the pinhole camera model optimized to minimize the geometric error with respect to its corresponding refractive camera. The difference between the original 2D point and the one obtained using the virtual image plane is shown as the pinhole model error in Fig. 4.10(a), 4.11(a) and 4.12(a) for the three different flat-port parameterizations. In the ideal case, the camera plane is parallel to the glass layer, and the un-modeled distortion is less than one pixel using a pinhole camera model. However, this error is range dependent and has its minimum for ranges similar to the one used for optimizing the pinhole model. This range dependency is, for example, problematic for bundle adjustments as it will favor 3D point locations at a specific distance to the camera. In the case, the camera plane is slightly tilted against the glass layer, or the camera has a none-optimal distance d_a to the glass-port; the error of the pinhole model significantly increases as the pinhole model can no longer compensate the error.

For the proposed refractive forward projection, an estimate for the 2D point location is required. The pinhole camera model can obtain this initial location despite its deviation to the real value. This is done fo Fig. 4.10(b), 4.11(b) and 4.12(b) and the remaining errors in pixels are visualized for the proposed flat projective forward projection. Following this, for Fig. 4.10(c), 4.11(c) and 4.12(c) the new improved locations of the projections are used to re-center the Taylor Series and the results of the second iteration are visualized. As the deviation of the starting points to the real locations of the projections increase with larger tilt angles of the camera, also the remaining errors of the proposed refractive forward projection increase. However, in real case scenarios the camera is usually only slightly tilted against the glass layer similar to Flat_3, where the pinhole model can deliver reasonable close approximations to obtain a final accuracy close to $\frac{1}{10}th$ of a pixel after one iteration. This accuracy is sufficient for most applications as it is close to the accuracy of sub-pixel peak detectors for localizing natural feature points in the image domain.

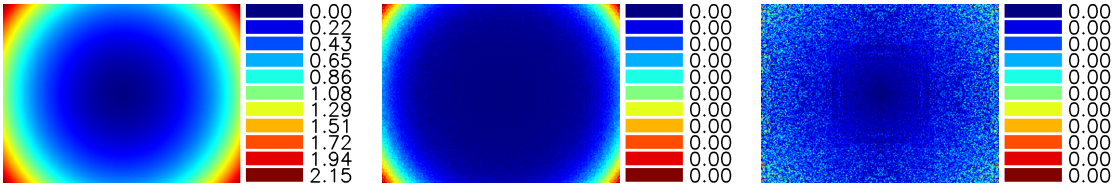
For validating the proposed flat refractive forward projection is stable for cases the point for centering the Taylor Series is very far away from the real projection of the given 3D point P_0, a 2D image point p_0 at location 200/200 is chosen, and its corresponding 3D point is calculated. Following this, for each image point p, the proposed flat projective forward projection is calculated using the point location itself as a starting point for projecting the 3D point P_0 onto the image plane.

In Fig. 4.13 the relative remaining error after one iteration is visualized. Here, for all image locations, the proposed flat refractive forward projection can considerably



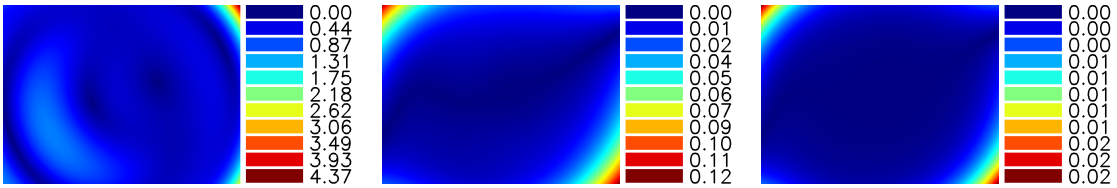
(a) Pin_1 (b) Flat_1 1. iteration (c) Flat_1 2. iteration

Figure 4.10: Re-projection error in pixel introduced by a pinhole camera model when used instead of a flat-port model. a) error of Pin_1; b) error after the first iteration of the proposed refractive forward projection centered at the projections obtained by Pin_1; c) error after the second iteration centered at the projections obtained by the first iteration.



(a) Pin_2 (b) Flat_2 1. iteration (c) Flat_2 2. iteration

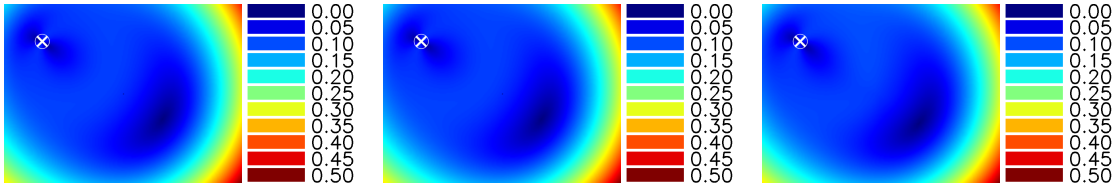
Figure 4.11: Re-projection error in pixel introduced by a pinhole camera model when used instead of a flat-port model. a) error of Pin_2; b) error after the first iteration of the proposed refractive forward projection centered at the projections obtained by Pin_2; c) error after the second iteration centered at the projections obtained by the first iteration.



(a) Pin_3 (b) Flat_3 1. iteration (c) Flat_3 2. iteration

Figure 4.12: Re-projection error in pixel introduced by a pinhole camera model when used instead of a flat-port model. a) error of Pin_3; b) error after the first iteration of the proposed refractive forward projection centered at the projections obtained by Pin_3; c) error after the second iteration centered at the projections obtained by the first iteration.

reduce the remaining deviation to the real projection. Therefore, for the whole image, the proposed approximation always converges to the true value without getting stuck in a local minimum. However, the convergence speed greatly depends on how close the point p is located to the point p_0. In the worst case, the deviation is only reduced by a factor of two after the first iteration. However, in case the distance to the true value is close, the deviation can be reduced by a factor of more than 10 using a single iteration.



(a) Flat_1 (b) Flat_2 (c) Flat_3

Figure 4.13: Relative error of the approximated refractive forward projection depending on the point p used for centering the Taylor Series. Here, the true projection is (x=200,y=200) shown as cross. The color indicates the remaining error after one iteration relative to the error at the beginning for each possible image point.

These considerations show that for many applications, a standard pinhole camera model can correct most of the distortion introduced by the flat-port housing if the camera is designed with care and is orthogonal to the glass interface. However, small deviations from the optimal camera placement will introduce considerable additional distortions that cannot be compensated by the pinhole camera model. Also, demanding applications like structured light systems able to track artificial features with an accuracy of $\frac{1}{100}$ of a pixel, bundle adjustments, or high-resolution cameras will face a reduced performance if the refraction is not explicitly taken into account during processing. Also, having access to the underlying flat-port model improves the calibration process as in-air calibration is now possible. This allows estimating the model parameters for the underwater case taking the refraction index of different water bodies (fresh versus saltwater) into account. Here, an in-depth consideration for calibrating cameras is given in the next sections.

4.3 Line Projector Model

A line laser projector can be handled as an inverse line camera, which only has a single coordinate in the former 2D image domain. Therefore, in the case of the flat glass interface, the flat-port model can also be used for the line laser projector. Here,

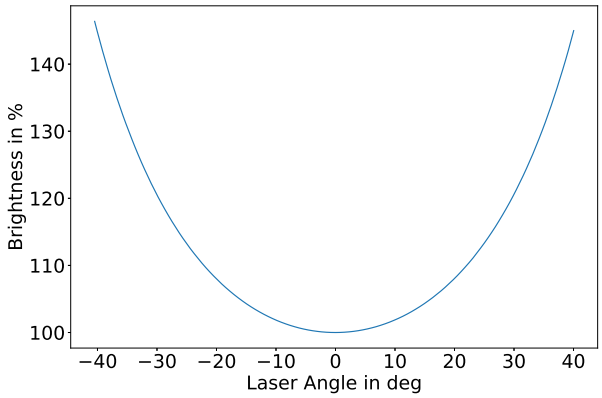


Figure 4.14: Brightness distribution of a former homogeneous laser line due to refraction compressing the line at its edges.

the glass port has mainly two effects onto the laser sheet while being submerged. Assuming the laser projector is perfectly orthogonal to the glass port, all refraction will take place in the laser plane itself. The result is that a former homogenous bright laser line gets transformed into a line that is brighter on its edges, like visualized in Fig. 4.14. However, the absolute brightness of the laser is also altered by the medium water, and the texture of the object reflecting the laser line to the camera is usually not used for depth estimation.

The second effect is more critical and takes place when the laser line is no longer orthogonal to the glass interface. In this case, the refraction no longer lies inside the laser sheet, and the laser line gets geometrically distorted, as visualized in Fig. 4.15 for different tilt angles. This distortion leads to an imprecise triangulation of laser points if not taken into account and is usually ignored by underwater line structured light systems assuming the orthogonal case.

However, the distortion can be approximated by shifting the principal point of an inverse pinhole camera and by applying tangential distortion. The rationale behind this new parametrization is that a flat-port model is, in fact, an axial camera where all camera rays intersect in a common line, which has the same direction as the normal of the flat-port. As long as this line is part of the ordinal laser sheet, no geometric distortion will be observed. However, in the case, the laser sheet is non-orthogonal to the flat-port interface, the virtual location of the light emitter is pushed away from the original light-sheet depending on the radial distance of the point leading to a symmetrical distortion of the laser line. Therefore, a sufficient projector model parameterization is listed in Tab. 4.7. Here, the focal length and the principal point

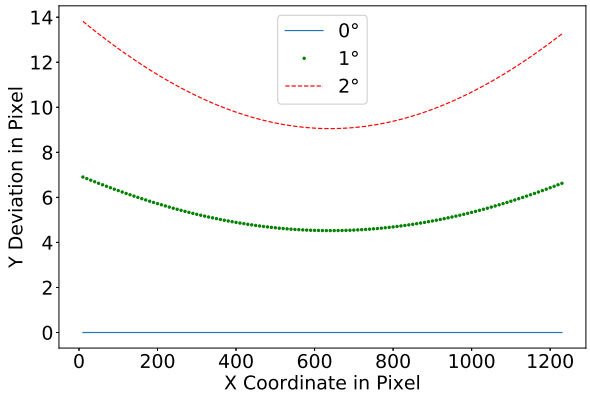


Figure 4.15: Deviation of the laser line due to a tilted glass port. The deviation is visualized as observed from an ideal camera (Pin1) having the same pose as the line projector.

Symbol	Description
f	Focal length (canonical value)
p	Principal point (canonical value)
P_1	Tangential distortion
R	Orientation (plane normal)
d	Baseline (distance to origin in m)

Table 4.7: Parameters of a flat-port line projector model.

can be set to the canonical values because no real measurements can be taken in the image domain of the projector. The accuracy which can be reached by this model for different use cases is visualized in Fig. 4.16.

This proposed line projector parameterization is a simplification of the flat-port model. It is only valid for small tilt angles against the glass port, which usually is in the order of one degree when manually assembled without specialized equipment. The benefit of using this model is that the laser sheet is approximated by a parabola back-projected into 3D space. This allows simplifying the triangulation of 3D points by calculating the intersection between the refracted camera rays and the back-projected parabola.

4.4 Calibration

Every real camera has a specific characteristic due to small imperfections and tolerances during assembly. These characteristics can usually be described by a suitable

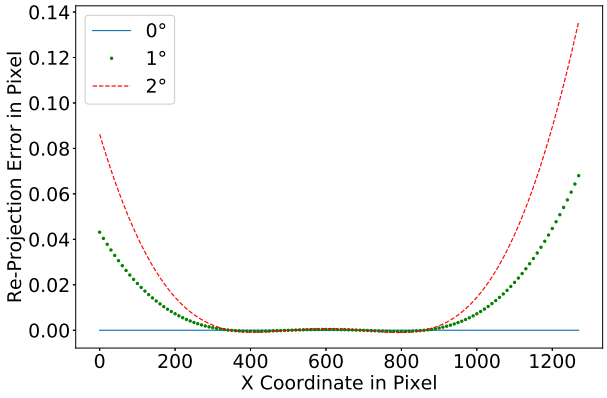


Figure 4.16: Re-projection error between the laser model and the refractive forward projection using a given tilt angle to modify the parameterization of the flat-port Flat1 from Tab. 4.5.

camera model resulting in remaining re-projection errors in the order of a fraction of a pixel. However, the correct model parameters have to be determined for each camera, which is usually referred to as camera calibration or finding the intrinsic camera parameters.

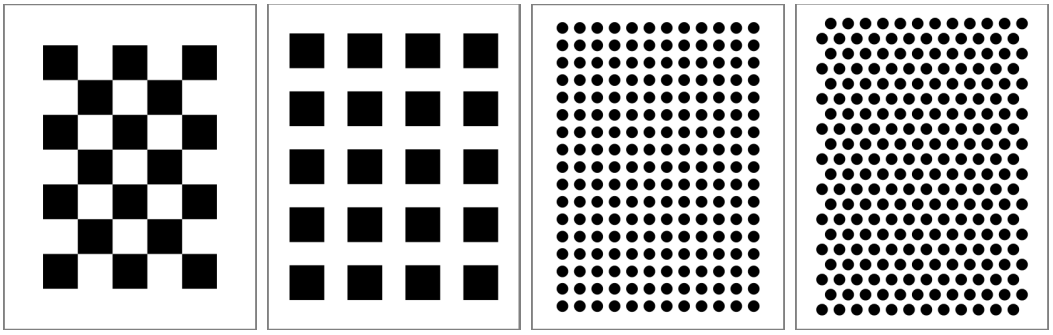
Finding the intrinsic camera parameters usually involves the detection of a known object in the camera image, which signals multiple object-points. Those relationships with each other are precisely known in 3D space. Based on these object points and their measured projections, the model parameters can be approximated and iteratively refined by minimizing the re-projection error between measured image points and projected ones.

This calibration process is usually done for most machine vision cameras by collecting a set of images showing a calibration target from different positions (Duda and Frese [2018]). In general, this involves the following steps:

- Setup suitable light conditions.
- Record multiple images of a calibration target from different positions.
- Detect calibration target in the image domain.
- Filter out distorted images due to motion blur, difficult light conditions or difficult target poses.
- Estimate camera model parameters as starting point for iterative refinement.
- Refine model parameters.
- Filter out images subject to high re-projection errors above a given threshold.
- Repeat refinement and filtering until convergence.

Calibration Targets

Checkerboards are by far the most popular calibration targets in the vision community (Fig. 4.17). They have strong gradients in all directions, and more importantly, the exact localization of their crossings is mostly invariant to lens distortion. However, it is difficult to get measurements close to the image borders required for an accurate calibration as four full fields are required to signal a single point. Here, calibration targets using squared/circle grids are less prone to this limitation. Still, it is more difficult to detect them in the image domain with a comparable precision like checkerboard patterns (Abeles [2018]). One reason for this is that their geometric center is not invariant under perspective distortion, and extra care must be taken in the case of larger circles. They are also more sensitive to image noise (Luhmann et al. [2013]), whereas a robust and accurate calibration can be achieved using checkerboards (Duda and Frese [2018]).



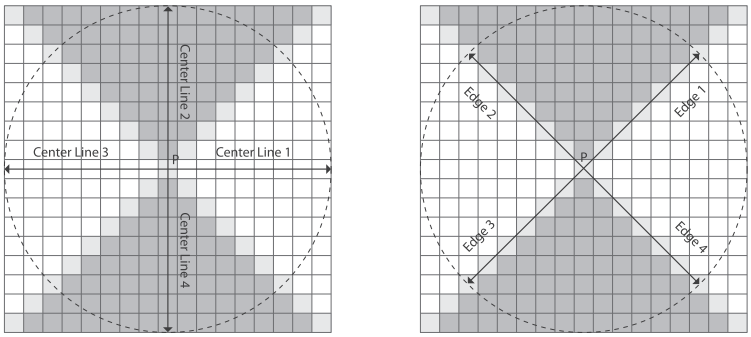
(a) Checkerboard (b) Squared Grid (c) Circle Hex. Grid (d) Circle Reg. Grid

Figure 4.17: Standard targets for camera calibration Abeles [2018].

Robust Checkerboard Detection

The central part of this work has been already published in Duda and Frese [2018], which was mainly motivated by the demanding requirements of a flat-port camera calibration due to its additional model parameters in comparison to standard models. Here a summary is given as the presented approach is used for calibrating all flat-port cameras part of conducted experiments.

Checkerboard detection has mainly two aspects: The first is to detect the checkerboard as a whole and distinguish it from other image content. This step determines, in particular, in difficult conditions, whether the image can be used in the calibration or not. The second is finding accurate corner locations. This localization determines the accuracy and precision of the calibration. In the case of checkerboard targets,



(a) Centerline based junction model (b) Edge based junction model

Figure 4.18: Junction models in the case of no projective distortion.

the Harris operator by Harris and Stephens [1988] is often used for detection and combined with the Förstner Operator published by Forstner and Gulch [1987] for sub-pixel refinement. Here, the Förstner Operator exploits the observation that the image gradient $\nabla f[x, y]$ at a point $\binom{x}{y}$ is perpendicular to the line from $\binom{x}{y}$ to the corner $\binom{c_x}{c_y}$. Implementations, based on these two operators, can be found, for example, in the OpenCV's camera calibration pipeline (Bradski [2000]), or work published by Douskos et al. [2008], and Rufli et al. [2008].

Following Sinzinger Sinzinger [2008], there are basically two models for a checkerboard corner: Black and white sectors with centerlines (Fig. 4.18(a)) and a cross of edges (Fig. 4.18(b)). An edge-based view leads to algorithms based on gradients that are susceptible to image noise in comparison to an algorithm that takes a sector-based view. Such a sector-based algorithm is developed in the following exploiting the point symmetry of checkerboard corners with a localized Radon transform approximated by box filters. This approach is far less affected by image noise than gradient-based methods and considerably improves the calibration results.

The underlying concept of the new corner detector is that all neighboring regions are point symmetric to the desired point (Fig. 4.18(a)), unlike in most "natural" corners. Therefore, centerlines from opposite regions are viewed as a common centerline. Following this, the sum of all pixel values along the centerline through the bright areas is higher than the sum of pixel values along the other centerline through the dark areas. In fact, if the two paths used for summation/integration are the centerlines of a checkerboard corner, the difference between both integrals will be maximized. In all other cases, the value will be reduced by the amount the paths are touching neighboring regions and is zero for homogeneous image regions.

Assuming a function $f_c[x, y]$ exists, which calculates the integral for all possible centerlines around a given image point $f[x, y]$ and returns the square difference of the maximal integral value and the minimal integral value. Checkerboard corners can be identified with a min-max search by transforming the image into a response map using this function. This approach is similar to the Harris detector response map (Harris and Stephens [1988]). However, instead of shifting window patches in x and y-direction of the image plane they are rotated around the selected point. After local maximums are identified in the response map subpixel position of corresponding corners can be estimated by subpixel peak algorithms. A comparison between different algorithms is given in Fisher and Naidu [1996] and Kiger [2010]. In this case, a **Gaussian peak fit** is used.

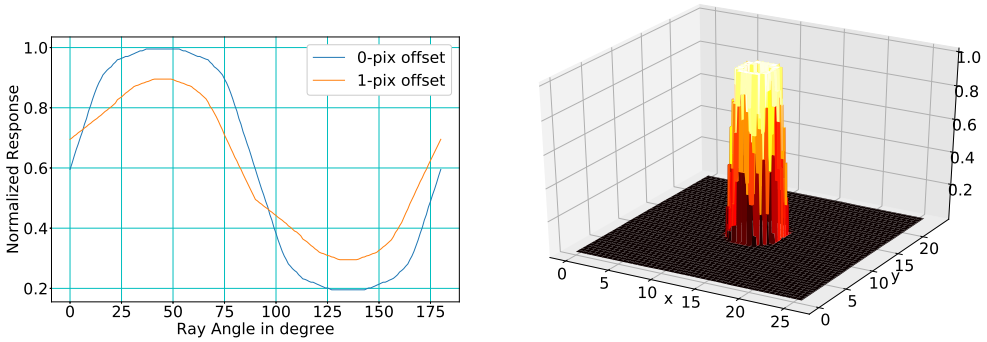
A closer examination of the requested function $f_c[x, y]$ reveals that it is in fact similar to the Radon transform $Rf[r, \alpha]$ (Ginkel et al. [2004]) localized to each image point $f[x, y]$ while setting r to zero. This localized Radon transform $Rf_{local}[x, y, \alpha]$ provides the integral for all possible centerlines or rays for a given point coordinate x, y. Here, only the squared difference between the strongest ray and the weakest ray is returned by $f_c[x, y]$ with m specifying how large the corner is expected to be in the image.

$$\begin{aligned} Rf_{local}[x, y, \alpha] &= \sum_{i=-m}^{i=m} f\big[x + i\cos(\alpha), y + i\sin(\alpha)\big] \\ f_c[x, y] &= [\max_{\alpha\in[0,\pi]} \big(Rf_{local}[x, y, \alpha]\big) - \min_{\alpha\in[0,\pi]} (Rf_{local}[x, y, \alpha])]^2 \end{aligned} \tag{4.28}$$

However, calculating the localized Radon transform for each image point would be very time-consuming. Therefore, max and min are approximated by considering only $\alpha \in \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}\}$ as angles.

This approximation is possible due to the point symmetry of checkerboard corners generating a function similar to a sine or cosine after the localized Radon transform is applied (Fig. 4.19).

This function curve allows to efficiently approximate the transformation of the input image into a response map by convoluting the image with box filters using a kernel size matching the considered window patch around a corner. But instead of rotating the filter kernels by an angle α (steerable filter), the image is rotated counter-wise, convoluted with two 1D kernels, and turned back. This approach is motivated by the work of Maire [2009] using the same method to speed up contour detection. However, in this case, box filter kernels are used to calculate the integrals of a localized Radon transform for pre-defined discrete ray angles.



(a) Localized Radon Response (b) Derived Repsonse Map

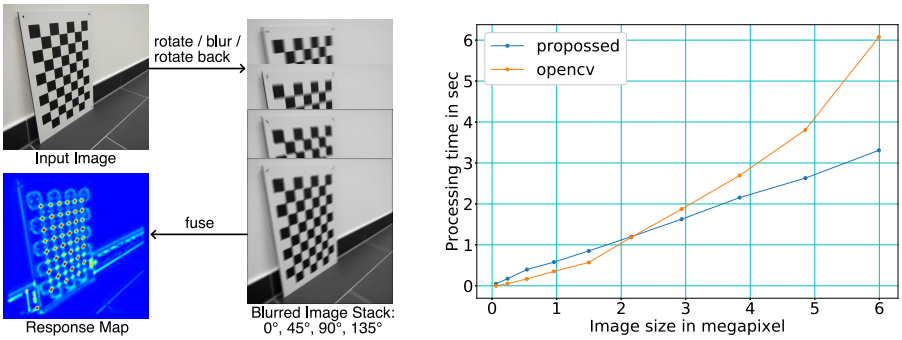
Figure 4.19: Responses for a checkerboard corner.

$$
\begin{array}{lcl}
f_{blur}[x,y] & = & \frac{1}{2k+1}\sum_{i=-m}^{i=m} f[x+i,y] \\
f_{rot}[x,y,\alpha] & = & f[x\cos(\alpha)-y\sin(\alpha), x\sin(\alpha)+y\cos(\alpha)] \\
f_r[x,y,\alpha] & = & f_{rot}\big(f_{blur}(f_{rot}(f[x,y],-\alpha)),\alpha\big) \propto Rf_{local}[x,y,\alpha] \\
f_c[x,y] & \sim & [\max\big(f_r[x,y,0], f_r[x,y,\frac{\pi}{4}], f_r[x,y,\frac{2\pi}{4}, f_r[x,y,\frac{3\pi}{4}]\big)- \\
 & & \min\big(f_r[x,y,0], f_r[x,y,\frac{\pi}{4}], f_r[x,y,\frac{2\pi}{4}, f_r[x,y,\frac{3\pi}{4}]\big)]^2
\end{array}
\tag{4.29}
$$

Each rotation combined with a convolution generates one new copy of the input image blurred in a discrete direction. Therefore, a blur kernel size of $n \times m$ is used corresponding to the size of the window of the detector and is usually in the range between 1×3 and 1×9. By changing this window size in both dimensions it can easily be generalized to detect corners at a specific scale. In the next step, these blurred images are joined to a response map, according to Eq. 4.29.

A complete procedure of the detection is given below:

- Convert input image into a greyscale image.
- Supersample greyscale image by a factor of two to improve anti-aliasing.
- Rotate greyscale image around its center using the angles 0 and $\frac{\pi}{4}$.
- Blur rotated images each with an $n \times m$ and $m \times n$ box filter resulting in twice as many images as before and inflating the angle interval to $[0, \pi[$.
- Rotate blurred images back.
- Join images which are rotated back to their original orientation using Eq. 4.29.
- Blur resulting response map with a $k \times k$ box filter to account for the discretization errors.
- Locate corners by searching for local maxima in the blurred response map.
- Use thresholding and nonmaximal suppression to filter out weak corners.



(a) (b)

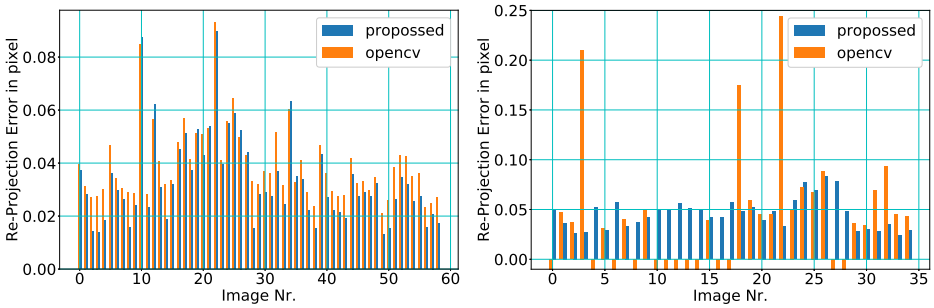
Figure 4.20: Detection of checkerboards using the proposed method: (a) response map calculation; (b) processing time for a checkerboard detection including subpixel estimation.

This method has some additional advantages useful for a checkerboard detector such as supporting scale space (Crowley and Parker [1984], Lindeberg [1991]) or providing the angle of each corner by fusing the blurred input images using a subpixel peak detector estimating the angle of both centerlines. Also, it can be directly applied to determine the subpixel location of each corner. Based on this, a checkerboard detector is implemented, which uses cross ratios to grow an initial $3x3$ checkerboard by adding more and more local maxima from the response map stored in a k-d tree. The initial board is generated by drawing a random point from an initial point set and by using the provided angles to search for neighboring maximums in the response map.

Here, the described method is also mostly linear in the number of processed pixels and usually outperforms the checkerboard detector implemented in OpenCV with respect to localization error and processing time. The processing time is displayed in Fig. 4.20(b) for the detection of the checkerboard visualized in Fig. 4.20(a) using the described and the OpenCV method.

Furthermore, the results for two real image sets using two different camera setups are displayed in Fig. 4.21. Here, the first image set Fig. 4.21(a) was taken indoors in a controlled environment with a static checkerboard while the second Fig. 4.21(b) was taken outdoors under difficult illumination conditions and a quasi-static checkerboard setup introducing small amounts of motion blur.

For the checkerboard detection, the described method is compared to OpenCV's implementation ($9x9$ window size). The result is passed through the same code to estimate camera parameters and to calculate the remaining re-projection error of each checkerboard corner. These errors are plotted in Fig. 4.21 as mean error for each



(a) (b)

Figure 4.21: Re-projection error of each detected checkerboard after calibration: (a) calibration in controlled environments; (b) outdoor calibration - negative values indicate that no checkerboard was detected.

	Proposed	OpenCV
Focal Length	990.52 / 990.94	990.31 / 990.66
Principle Point	642.94 / 552.31	642.37 / 551.81
Radial Distortion	-0.12172 / 0.12032	-0.11951 / 0.11917
Tangential Distortion	0.00316 / -0.00101	0.00316 / -0.00101
Mean Residual in pixel	0.0332	0.0391

Table 4.8: Camera Parameters - Indoors

	Proposed	OpenCV
Focal Length	746.49 / 745.86	746.79 / 745.48
Principle Point	514.66 / 512.73	514.12 / 514.12
Radial Distortion	0.01752 / -0.01585	0.02114 / -0.01993
Tangential Distortion	0.00062 / 0.002016	0.00147 / 0.000798
Mean Residual in pixel	0.0459	0.0738

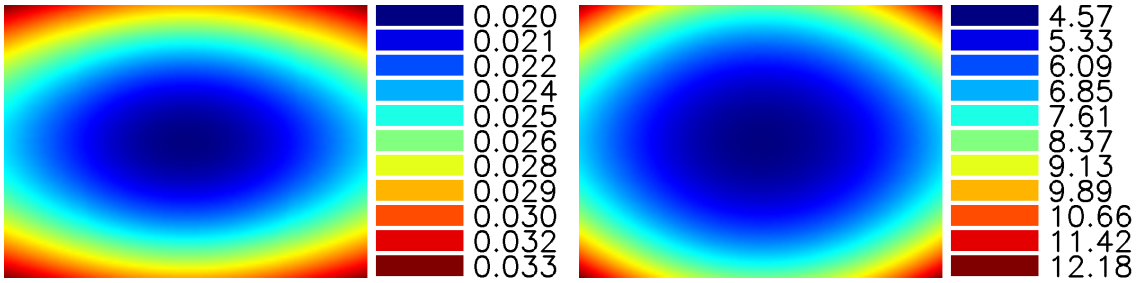
Table 4.9: Camera Parameters - Outdoors

image of the set. In case, the checkerboard was not detected, or the residual error was above 0.5 pixels, the image was excluded and is marked with a negative re-projection error in the plots.

Here, the proposed method is outperforming the OpenCV implementation (OpenCV v3.3 - cv::findChessboardCorners) under all tested conditions and results in considerably higher detection rates and a smaller re-projection error. The resulting intrinsic camera parameters are displayed in Tab. 4.8 and 4.9. In particular, the outdoor case shows a problematic increase in the re-projection errors for the baseline method, while the proposed method can maintain a relatively small re-projection error. A more in-depth analysis is given in Duda and Frese [2018].

Camera Model Refinement

The ability to detect and track calibration targets unbiased and with high accuracy in the image domain, discussed in the previous section, is especially crucial for calibrating complex camera models like, for example, the flat-port camera model. The more noise is present in the detected point coordinates, the more the convergence of the optimization step is affected during model parameter optimization. Here, if the noise is too high, the optimization problem might converge only to a local minimum leading to poor performance when used for applications like motion tracking or 3D scene reconstruction.



(a) Error in-Air (b) Error in-Water

Figure 4.22: Re-projection error of a flat-port camera (Flat1) in pixels due to an one degree tilted $10mm$ glass layer with respect to an orthogonal case.

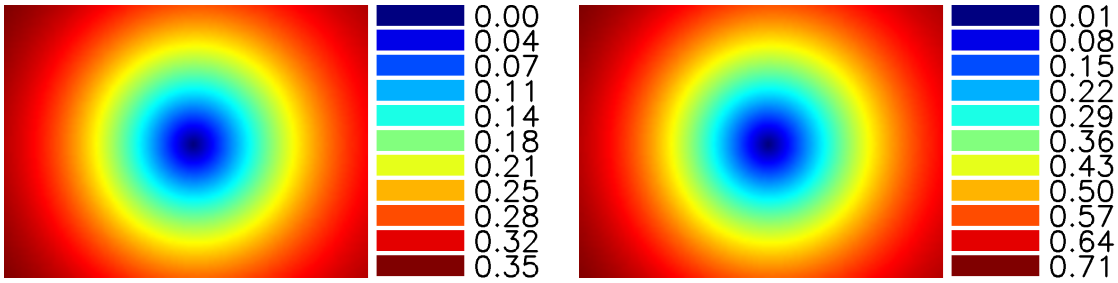
For accurately calibrating a flat-port camera, the previously discussed flat-port model in combination with the new checkerboard detector is used. In the first step, a pinhole camera model is assumed, and the camera is calibrated in-air using the method described in Zhang [2000] and implemented in Bradski [2000]. Following this, these values are used as an initial guess for the flat-port model. Here, the additional flat-port model parameters are set to their default values assuming all refraction indices and the thickness of the glass layer are known. Using these initial values in combination with the poses of the checkerboards found during the pinhole camera calibration, a bundle adjustment is performed. Here, as residuals, the forward projection of the flat-port model is used utilizing the measured 2D point locations of the checkerboard corners to center the Taylor series.

During the iterative refinement, the normal vector of the glass layer is set to be constant as the effect of a tilted glass layer in-air is close to the theoretical limit of sub-pixel peak detectors in the image domain. As an example of a one-degree tilt, the re-projection error is visualized in Fig. 4.22. According to this, the glass layer normal can only be calibrated when the camera is being submerged. However, without loss of generality, for small tilt angles, the camera can be assumed to be orthogonal to the glass layer for the in-air case.

The same is true for the thickness of the air layer because the media on both sides of the glass are near identical for an in-air calibration of a flat-port camera. Therefore, it is impossible to find the exact location of the glass layer between the object point and its projection while the camera is not being submerged. For the in-air case, any value smaller than the distance to the real object can be used. The following table summarizes the initial values of the in-air calibration, which is assuming n_g and d_g are known based on the mechanical design of the flat-port camera.

Symbol	Description	Initial Value	Attribute
f_x,f_y	Focal length	Pinhole Camera Calibration	variable
p_x,p_y	Principle Point	Pinhole Camera Calibration	variable
K_1,K_2	Radial Distortion	Pinhole Camera Calibration	variable
P_1,P_2	Tangential Distortion	Pinhole Camera Calibration	variable
$\vec{n}$	Interface normal	[0,0,1]	const
n_a, n_g, n_w	Refraction indices	1.0,n_g,1.0	const
d_a, d_g	Layer thickness	[0.01,d_g]	const
R	Camera Orientation	Pinhole Camera Calibration	variable
C	Camera Center	Pinhole Camera Calibration	variable

Table 4.10: Flat-port model parameters for iterative refinement in-air.



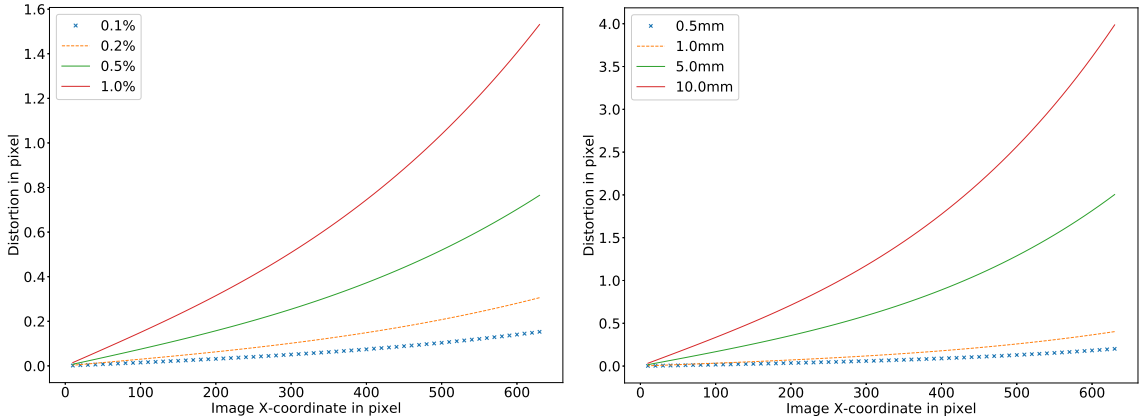
(a) 5mm Glass Layer (b) 10mm Glass Layer

Figure 4.23: Influence of the glass layer thickness onto projected scene points in pixels using the parameterization of Flat1.

Comparing the calibration results between a pure pinhole camera model and the flat-port camera model for the in-air case, the small effect of the additional glass layer can be noticed. Its influence on the re-projection result is visualized in Fig. 4.23 for different glass thicknesses requiring an exact sub-pixel target location for in-air calibration capability.

In addition, Tab. 4.11 shows the change in the camera parameters for a real-world calibration due to the additional flat-port refinement step while being submerged. For the calibration and the further refinement, around 50 checkerboard images with random checkerboard poses were collected averaging errors due to Bayer pattern, motion blur, in-homogeneous illumination, image noise, etc.

A small tilt of the camera against the glass layer is noticeable when comparing the free parameters, optimized while the camera is being submerged, with the one from the in-air calibration listed in Tab. 4.11. This tilt is due to the manufacturing process of the camera and hardly avoidable as the glass port is pressed against a sealing that gets compressed. This compression is also the reason why the camera



(a) Change in the salinitly level

(b) Change in d_a

Figure 4.24: Image distortion due to an change in the salinity level of the water body or a distance change of the focal point with respect to the flat-port.

Symbol	Flat-port in-air Calibration	Flat-port in-water Refinement
f_x,f_y	777.73,776.93	777.73,776.93
p_x,p_y	648.49,489.61	648.49,489.61
K_1,K_2, K_3	-0.329023, 0.121897,-0.022262	-0.351943, 0.168047 -0.044378
P_1,P_2	-0.000556, -0.000227	0.000294, -0.000974
$\vec{n}$	[0,0,1]	[0.004942, -0.006180, -0.999969]
n_a, n_g, n_w	1.0,1.72,1.0	1.0,1.72,1.337
d_a, d_g	[0.01,0.01]	[0.0195,0.01]
Reprojection Error	0.07478	0.05157

Table 4.11: Flat-port model parameters after in-air calibration and after iterative refinement of the distortion parameters, layer normal and air layer thickness while being submerged.

parameters are subject to small changes while diving. However, these changes are small in comparison to a change in the refraction index due to different salinities. Therefore, in the case, high precision is required, an on-site refinement step should be considered.

Line Projector Refinement

The line projector model outlined in Tab. 4.7 has five degrees of freedom, which is analog to the equation of a plane in 3D space augmented with an additional parameter for modeling a bending of the line. Therefore, by ignoring this extra parameter, the initial projector parameters can be found from multiple 3D points lying on a virtual plane, which is projected onto the line of the projector. Here, at least three points are required that are not collinear. Using multiple 3D points allows using least square fitting techniques finding the best plane parameters by minimizing the sum of the squares of the offsets of the 3D points from the plane (Eberly [2018]). In the case of outliers, this can also be combined with the Random sample consensus (RANSAC) algorithm, which acts as an outlier detection method (Fischler and Bolles [1981]).

After the initial values are calculated, the model is refined by using the distance d of projected 3D points x to the distorted image of the plane as residuum to optimize the model parameters. Here, the residuum is defined by the following equation where X are the homogenous coordinates of a 3D point, R the rotation matrix of the projector, and C the location of the projector. This equation is analog to the pinhole camera model with K set to identity and the additional constraint that all projected points must have no y component after they are un-distorted.

$$\begin{aligned} \begin{bmatrix} x_x \\ x_y \\ x_z \end{bmatrix} &= R[I|-C]X \\ d &= \frac{x_y}{x_z} - P_1 \frac{x_x^2}{x_z^2} \end{aligned} \tag{4.30}$$

Assuming a camera with a fixed baseline to the projector is already calibrated, the projector can be co-calibrated by projecting its line onto a planar target whose pose is known with respect to the camera. Here, as a planar target, standard calibration targets such as checkerboards can be deployed. Based on this, for each detected line point x on the planar target, its corresponding 3D point X can be calculated by intersecting the plane equation of the target with the camera ray belonging to the detected point x in the image domain. Because, at least three 3D points, used for calibration, must not be coplanar, the planar target must be positioned at least at

two different distances to the camera. For each distance, multiple 3D points must be collected. Here, also other methods exist to access 3D points part of the laser sheet. For example, in the work of Roman et al. [2010], a pre-calibrated stereo system was used to recover the 3D points underwater to calibration the line projector on-site without the need for particular calibration targets.

Conclusion

The medium water considerably alters the path of optical signals, and care must be taken to minimize these effects onto underwater sensing using special underwater lenses or by model the refraction of light. Here, flat-port housings invalidate the single viewpoint constraint of the pinhole camera model leading to significant errors depending on the application. In general, a pinhole camera behind a flat-port converts to an axial-camera where all rays meet in a common line instead of a common point. To account for this, refractive backward and forward projection must be taken into account, which is challenging to solve in a closed-form required for camera calibration and bundle adjustments. Therefore, a new approximation for the refractive forward projection based on Taylor Series is proposed to broaden the usage of general flat-port cameras. As an alternative, if the position of the camera with respect to the flat-port can be precisely controlled during manufacturing, the light refraction can be eventually mostly absorbed by the standard lens distortion model. In an ideal case, it might be even able to push re-projection errors into the sub-pixel range, which is sufficient for many marine applications. However, this is usually only possible for thick flat-ports or small lenses otherwise the lens cannot be placed close enough to the window.

For the calibration of underwater cameras and projectors, similar approaches like for the in-air case can be used. Here, most parameters of the refractive camera model can be already estimated in-air. And only some of its parameters require additional underwater calibration images as these parameters have only a measurable influence on the image formation process while the camera is submerged. For this extra refinement step, accurate detection and precise localization of known calibration targets in the image domain are required. Therefore, a new checkerboard detector is developed able to localize checkerboard corners with higher precision in comparison to standard methods. This higher precision improves the accuracy of flat-port camera calibrations and allowing to reduce the amount of required underwater calibration images considerably.

Chapter 5

Combined Active-Passive Vision System

Active visual sensing can be seen as orthogonal to passive visual sensing. Here, on one side, passive visual sensing has its strength when the scene is strongly textured, allowing to detect and match scene features between multiple images taken from different locations. However, in the absence of scene texture, the passive vision system is prone to fail as no scene features can be detected. On the other side, active systems project their own light signal onto the scene used for triangulation or time of flight measurements and work best when there is no scene texture or environmental light present. Nowadays, there are many systems available that combine both methods and deliver range data in addition to RGB images, also known as RGB-D cameras, TOF systems, or LIDAR systems fused with color cameras. However, these systems have one common design approach. The depth sensor is usually shifted to the infrared spectrum and does not interfere with the passive vision system. This separation allows us to separate both sensing modalities and simultaneously measure scene features and associated depth readings resulting in a well constraint optimization problem (Yang et al. [2010]).

In the underwater domain, only the visual spectrum of light is available for sensing, as discussed in Chapter 2. All other wavelengths are strongly absorbed by the water body and are unsuitable for optical sensing. Therefore, in the case passive and active vision are combined to collect data from the same underwater scene simultaneously, they have to use overlapping wavelength or at least wavelengths which are very close to each other leading to interferences on the sensing side. Therefore, classical fusion approaches for RGB-D systems are subject to high noise as they usually fuse sensor data after pose and or range data were separately estimated, not taking any crosstalk into account. Here, a promising solution is to simultaneously track natural scene

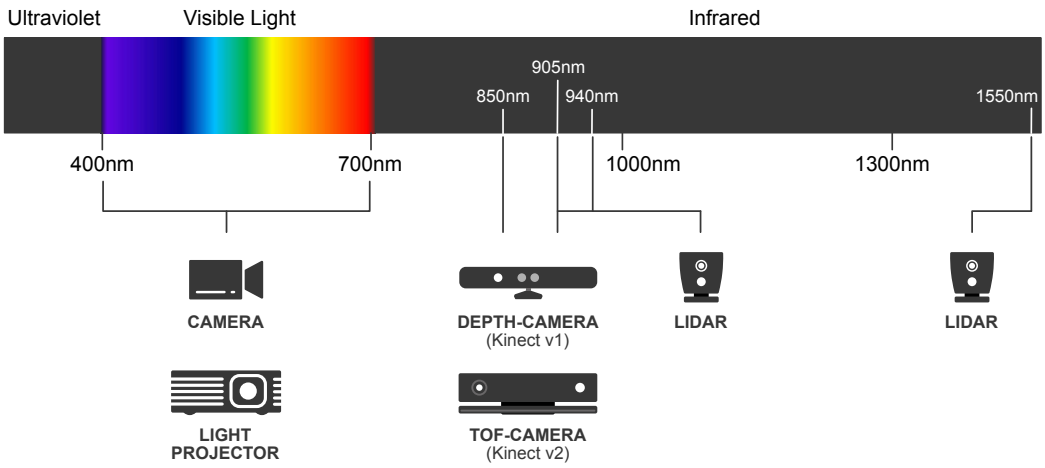
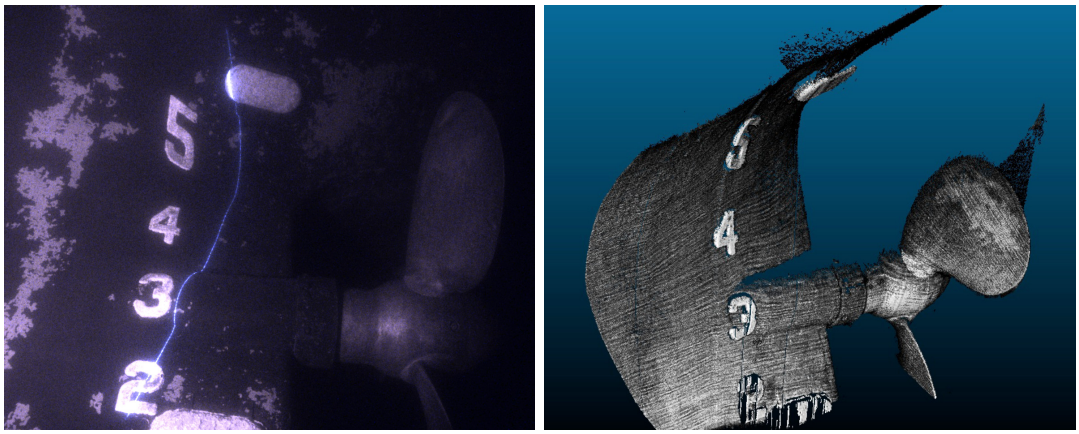


Figure 5.1: Wavelengths used for optical 3D range sensing.

features (passive-features) and features originating from the projected light pattern (active-features) with the help of the same camera sensor instead of using separate sensor elements. This approach has several advantages for cases requiring overlapping wavelengths between passive and active sensing, such as for underwater applications:

- Co-calibration between multiple sensors and time synchronization is not required, as the same physical camera is used for active and passive sensing.
- The resolution and characteristics are similar for both sensing modalities.
- The detection of active and passive-features in the same image domain reduces the complexity to label image features into one or the other reducing the noise both systems are facing when used simultaneously.
- The fused system can gradually fall back to passive or active sensing for image regions favoring one or the other modality as shown in Fig. 5.2.

However, the requirements on the sensing element depend on the sensing method, and, as discussed in Chapter 3, time of flight systems have very different requirements than triangulation based approaches. The first one requires a good time versus the second one a good spatial resolution. Therefore, using currently available sensing technologies, it is only feasible to fuse passive with active triangulation based methods using the same physical sensing element. But even here, competing requirements are present, and it is usually infeasible to measure the scene texture at image regions showing the projected light pattern. In the extreme, a light pattern covering the



(a) Image with projected light pattern. (b) 3D reconstruction of a ship hull.

Figure 5.2: Underwater scan of a ship hull with a line structured light system. Here, large regions have no scene texture, making it extremely difficult to be reconstructed with passive vision systems.

whole image and, therefore, masking all scene features, it is no longer possible to perform robust scene feature tracking simultaneously. Thus, in the following, a line structure light system is used, which is the ideal compromise between both extrema. This circumstance allows to fuse it with a structure from motion system and to utilize the remaining regions of the image for motion tracking. Here, the scale and dense 3D range profiles are recovered by the line structured light system while the structure from motion system performs the pose estimation. This approach was already published in Duda et al. [2016], and a more in-depth consideration is given below, which might also be extended to other light patterns than a laser line. However, line pattern prove very practical for many marine applications like discussed, for example, in the work of Narasimhan et al. [2005], Roman et al. [2010], and Bleier and Nüchter [2017].

5.1 Overview

The proposed passive-active system consists of a color camera and a visible line laser with a fixed baseline between both components, which is visualized in Fig. 5.3. From a hardware point of view, this is analog to a standard line structured light system where a Bayer filter replaces the usually used narrow optical bandpass filter.

Here, in each RGB image, the laser line is detected by the structured light system, and simultaneously a structure from motion system identifies and tracks passive-features. To reduce the impact of the projected light pattern on the detection of

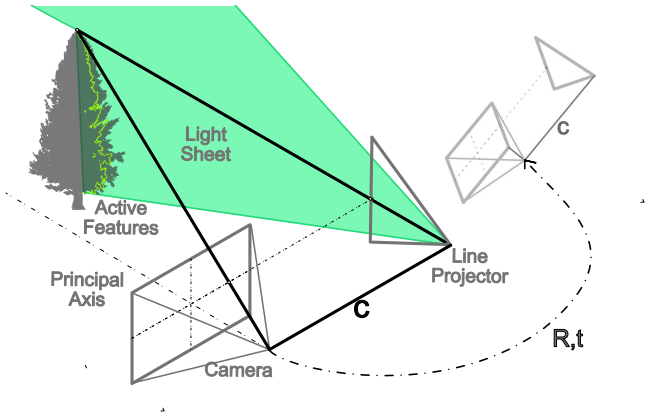


Figure 5.3: Overview of an passive-active line structured light system.

passive-features, all possible active-features are masked before passive-features are extracted. In addition, active-features are validated after the pose of the camera was recovered using epipolar constraints. An overview of the algorithm is outlined in Fig. 5.4 and a detailed description of each of the components is given in the following sections.

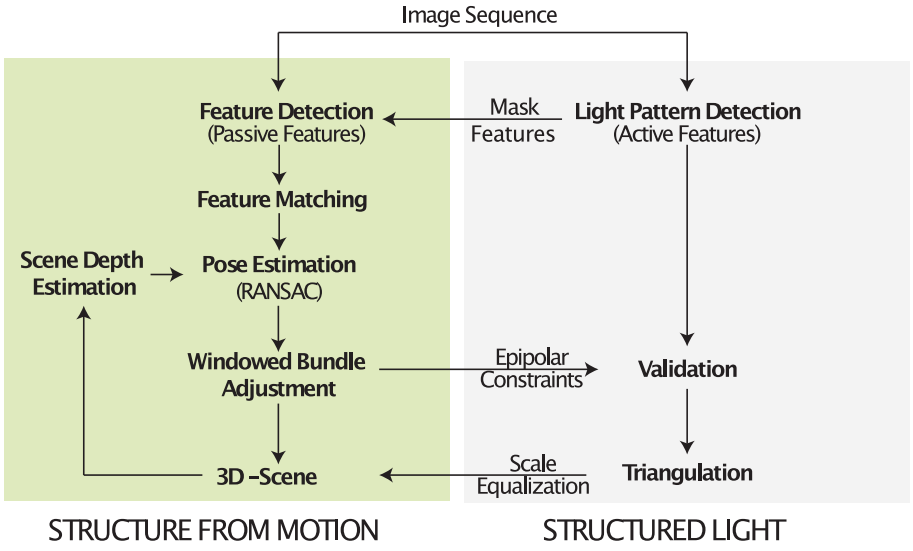


Figure 5.4: Overview fusion structure from motion with structured light.

5.2 Structure from Motion

The structure from motion component estimates the poses of a calibrated camera and the corresponding sparse scene structure by tracking visual scene features through an image sequence and by adjusting the camera poses in such a way that the detected

and back-projected features intersect with their corresponding 3D points. Here, each image contributes one bundle of rays with known angles between them according to the identified features in the image domain. By adjusting these bundles with respect to each other, the re-projection errors of 3D points projected onto each image are minimized, and the camera poses and the 3D scene structure are recovered. This circumstance is also reflected by the naming schema of this approach labeled as structure from motion in combination with bundle adjustment techniques for final refinement.

In general, there are direct, indirect, or semi-direct approaches available for simultaneously estimating the scene structure and sensor pose based on image sequences. Direct approaches skip the detection and matching step of visual features and directly minimize a photometric error in the image domain. Whereas, indirect approaches detect features in the image domain, match them with detected features from subsequent images, and finally minimizes a geometric error. Also, a combination of both approaches exists, usually referred to as semi-direct structure from motion. In the following, only an indirect structure from motion system is considered. However, similar methods can be used for other structure from motion flavors, which mainly differ in their robustness against certain types of noise and their relative error with respect to the traveled distance. An overview of different structure from motion systems is, for example, given in the work from Schonberger and Frahm [2016].

Indirect structure from motion systems usually consist of the following subsequent steps, often referred to as structure from motion pipeline because the images are entering the pipeline on one side and are processed by following components until the pose and 3D scene structure are recovered.

- Detection of features with high gradients in the image domain.
- Matching of features between subsequent images.
- Camera pose estimate based on the feature matches.
- Depth initialization of the 3D points based on the feature matches and camera poses.
- Bundle adjustment to refine camera poses and 3D point coordinates.

Here, many different algorithms and flavors exist for implementing each of these components, and there is no universal strategy for ensuring the best performance for a particular application. Some people refer to the design of a structure from motion pipeline still as black magic as the designer is confronted with an overwhelming

amount of possible combinations with unknown performance and possible side effects due to other components in the pipeline. In the following, the structure from motion pipeline is briefly described used for implementing an underwater passive-active vision system. However, it can be easily substituted with any other structure from motion pipeline as long as additional cues from the structured light system can be embedded.

Feature Detection & Matching

Each captured image is divided into n rectangular segments, and for each segment, the strongest m Harris corners are stored. These are used to calculate a sparse optical flow based on the iterative Lucas-Kanade method with pyramids (Bouguet [2001]). Following this, the flow is used to generate feature matches between consecutive images and to track features through images sequences.

Camera Pose Estimation

After the initial pose and scene structure is recovered using the five-point algorithm derived by Nistér [2004], new poses are calculated based on the P3P algorithm (Gao and Hou [2003]) to speed up the calculation. Here, if the current features have enough parallax to the last keyframe, the current image is regarded as a new keyframe. In the case of a new keyframe, its corresponding pose is recalculated based on the trifocal tensor of the last three keyframes using the five-point and P3P algorithm in a RANSAC framework similar to the work presented by Nistér et al. [2004]). This approach allows to maintain an arbitrary but constant overall scale of the reconstruction and to recover from wrongly added 3D scene points. Following this, all inliers are used to calculate an optimized pose by minimizing the re-projection error based on the Levenberg-Marquardt algorithm.

3D Depth Initialization

After the pose of a new keyframe is recovered, new 3D points are triangulated and added to a global map using the camera views with the largest baseline. Here, the optimal triangulation method described by Hartley and Zisserman [2004] is used, which minimizes the geometric distance between the projections of a 3D point and the epipolar lines defined by their corresponding feature matches and the camera centers.

Windowed Bundle Adjustment

The complexity for long-lasting image sequences is bounded by a windowed bundle adjustment, which regards only the last n keyframes to minimize the re-projection error of 3D points seen from the selected keyframes. The optimization itself is done by the open-source Ceres library developed by Agarwal and Mierle [2018]. Here, the flat refractive forward projection described in Section 4.2 is used as a residuum to support underwater flat port cameras. A comparison of 3D underwater scene reconstructions performed with different camera models is given in Chapter 6.

5.3 Structured Light

The structured light component detects the known light pattern in the captured images and triangulates the scene depth along this pattern based on the known fixed baseline and orientation between the camera and the light projector. In general, a structured light system performs the following processing steps:

- Detection of known light patterns in the image domain.
- Association of the detected light patterns to corresponding projector angles.
- Triangulation of 3D scene points along the detected light patterns.

In the case of marine applications, the most common light pattern is a laser line projected by a laser diode in combination with a line optic such as a Powell lens. The reason for this preference is that high power multi-mode laser diodes can deliver up to several watt optical power on the line, which is not easily achievable with other projection techniques. Also, the laser line minimizes the common volume between the camera sensor and projector, leading to less backscattering in comparison to areal illumination schemas. Although a range-gated point source would be the optimum, the laser line gives the advantage to allow for connectivity constraints along the line to filter out noise such as marine snow appearing as bright spots in the camera image. Therefore, in the following, a laser line projector is used but could be replaced by any other light pattern projector working in the visible spectrum of light and is covering only a sub-part of the image to allow for the detection of passive-features simultaneously.

Light Pattern Detection

Camera images usually show the projected light pattern in conjunction with some background textures illuminated by an unknown external light source like visualized in Fig. 5.5. This background noise, is also present if a narrow optical bandpass filter is used to block most of the environmental light not belonging to the laser line projector.

For the proposed system, the common optical narrow bandpass filter of a line structured light system is replaced by a Bayer filter to improve the capturing of passive-features by the structure from motion system. This replacement increases the effective passband from around $10nm$ to $150nm$ for the structured light system. Therefore, in the presence of other light sources, it reduces the signal-to-noise-ratio of the structure light system as more photons from the different light sources can reach the sensing element. Also, the Bayer filter reduces the spatial resolution of the system because usually, not all pixels can be used for sub-pixel peak detection of the light pattern. However, it improves the rejection of false positives as the scene is not only spatially sampled but also over its spectrum.

Also, the sub-pixel peak detection is not only limited by the sensor resolution but also by the light pattern projected back to the camera. Here, the highest accuracy can be achieved if the light pattern has a size of two pixels pointed out by Kiger [2010]. The reason for this is that smaller features can no longer be resolved, and larger ones increase the random noise counteracting accurate sub-pixel peak detection using regression analysis. Therefore, in the case of a standard Quad-VGA color camera, the light pattern should have an optimal thickness of around $8mm$ in $3m$ distance to have the highest accuracy assuming a flocal lenght of 1112. This requirement is close to what can be achieved with underwater lasers even in relatively clear water bodies. Therefore, even for relatively low-resolution cameras, the Bayer pattern is usually not the limiting factor of the system.

For the detection of a light pattern in RGB images, and to suppress unwanted background textures, a simple difference image approach can be used assuming a static scene. Here an image is taken while the light projector is turned on and another one while it is turned off. By subtracting both images from each other, the background can be successfully removed (Mertz and Koppal [2012]). This is also often referred to as ambient light filtering. However, this assumption does not hold if the system is in motion because a slight difference in the camera pose will produce artifacts similar to the light pattern. This circumstance also applies to methods using temporal coding

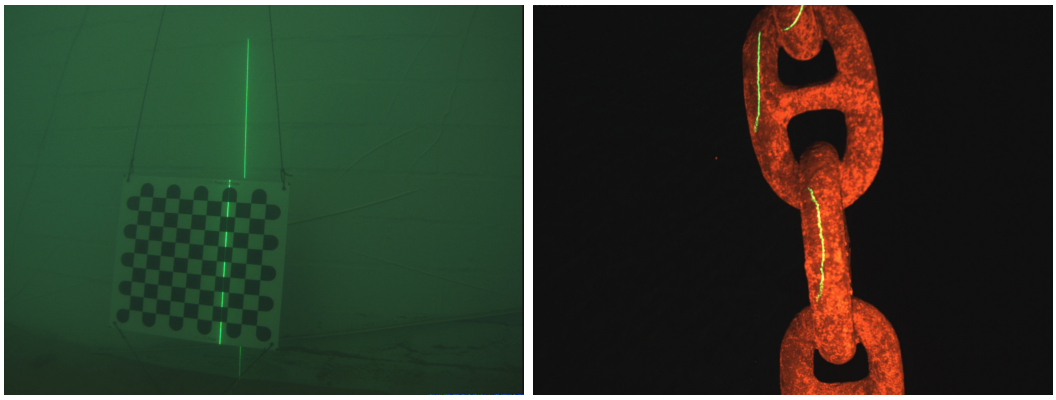


Figure 5.5: Raw images captured by the camera showing the projected light pattern.

requiring multiple measurements for each pixel. Therefore, to overcome this, the following method for laser line pattern extraction is proposed:

The image is logarithmically scaled to normalize the laser linewidth across the whole image (Duda and Albiez [2013]). This scaling allows to use a subsequent 1D FIR bandpass filter with a constant bandwidth applied vertically in respect to the laser plane assuming the target signal is symmetric across its two sides.

Following this, the signal-to-noise-ratio is calculated for each filtered image pixel assuming the noise is represented by the mean value of a 1D window around each pixel. Based on this, the laser peaks are finally detected using non-maximum suppression. After possible laser points are identified, they are segmented into groups fulfilling connectivity constraints and are rated based on their group size and signal strength to filter out background noise. Here, for the connectivity constraints the distance to the next detected laser point is regarded and the score s is calculated according to the following formula with $b(n)$ being the measured delta brightness value of the n^{th} point of the segment and m the number of points belonging to the segment.

$$s = m + 10^4 \sum_{n=0}^{m} b(n) \tag{5.1}$$

This score is analog to the sum of all measured brightness values of a segment reduced by the estimated noise floor. Here, the introduced scale ensures that for standard camera resolutions, the segment size is only dominant if multiple segments have the same brightness sum. Other scores such as the signal-to-noise-ratio, yield less accurate results as the noise is only estimated by the surrounding pixels. Here errors in the noise estimation have a significant impact on the segment score, which would effectively suppress blurred laser projections due to, for example, water turbidity.

After the score for each segment is calculated, the laser line is extracted by combining the strongest non-overlapping segments. Finally, its sub-pixel location is estimated by a Gaussian peak fit using raw measurements around each detected peak value.

Pattern Association

In the case of a line laser, the pattern association or also called the decoding step is relatively simple as every camera ray intersects only at one discrete location in 3D space with a laser sheet/plane defined by the projector position with respect to the camera. Therefore, for each laser detection, only its corresponding camera ray must be associated to find the corresponding position in the projected light pattern. For other patterns such as pseudo-random point patterns, additional cross-ratio constraints between detected features would have to be used to associate projector rays to detect light pattern points in the image domain.

Triangulation

The location of the light projector is known with respect to the calibrated camera. Therefore the 3D location X of a detected light pattern can be estimated by calculating the intersection of the corresponding camera ray $\vec{v}$ with the associated projector ray. This step is analog to the triangulation of 3D scene points based on two cameras, but here one camera is replaced by a laser line projector (inverse camera) where each epipolar line is, in fact, a discrete point in the projector image domain and no feature matching step is required. Based on this circumstance, the 3D location of a detected active-feature can be triangulated with the help of the following equation with p_0 being the origin of the projector, $\vec{n}$ the laser sheet normal with respect to the coordinate system of the camera, $\vec{v_w}$ the camera ray leaving the flat port and entering the water layer and g the location where the camera ray enters the water layer.

$$X = \vec{v_w} \cdot \frac{(p_o - g) \cdot \vec{n}}{\vec{v} \cdot \vec{n}} + g \tag{5.2}$$

It is worth mentioning that no forward projection is required for calculating the 3D scene points based on a structured light system, and efficient closed-form solutions are available, including underwater flat port cameras.

5.4 Fusion

One of the main challenges for fusing structure from motion with structured light using light patterns in the visible spectrum of light is an increased number of outliers encountered by both systems. Here, the structure from motion system is profoundly disturbed by the projected light pattern in the feature detection and matching step, while the structured light system is affected by the environmental light required for the structure from motion system to be able to track passive-features. In the case a single camera is used for both modalities, optical band-pass filters cannot be applied to reduce the effects of environmental light onto the structured light system because this would also reduce the number of passive-features observable by the structure from motion system. This controversy is a substantial disadvantage in comparison to multi-sensor / dual-camera systems allowing a better sampling of the visible spectrum. However, the structure from motion system has to deal with interferences in either way. Therefore, using the same sensor for both modalities can be seen as re-using the interferences to get more knowledge about the scene, which leads to a single sensor passive-active vision system fusing structured light with structure from motion on a software level. Adding more sensing elements certainly improves the overall performance as more measurements and or constraints are available. But, the single-sensor system can be seen as a generalization of a multi-sensor system. It has considerably fewer measurements, fewer constraints, and less short cuts than multi-sensor systems but with the ability to be scaled up and constrained at a later stage using the same software components.

The most critical step for fusing structured light with structure from motion is to distinguish between active and passive scene features. Here, based on the working principle of triangulation based systems, a signal must be located in the image domain respectively spatial domain to derive its bearing with respect to the sensor. Therefore, a pure time encoded signal is not possible for triangulation-based systems, unlike it is the case for a time of flight based system. However, mixed encoding is possible. In the simplest case, the light pattern is interleaved in an on-off like fashion, and the light pattern detection is performed on the difference image between two consecutive images. More complex time encodings are also possible, which could also involve a blur in the spatial domain to increase the accuracy in the time domain. However, underwater systems are usually power-limited. Therefore, to achieve the most reliable signal in the spatial domain, a focused light pattern is preferred. Here, the most straightforward fusion strategy includes the following steps:

- Detection of active and passive-features in the image domain.
- Classification of the detected features.
- Pose and 3D scene structure recovery.
- 3D scale equalization between both modalities.

Each of these basic building blocks is discussed in the following subsections. More complex strategies are derived later in this chapter.

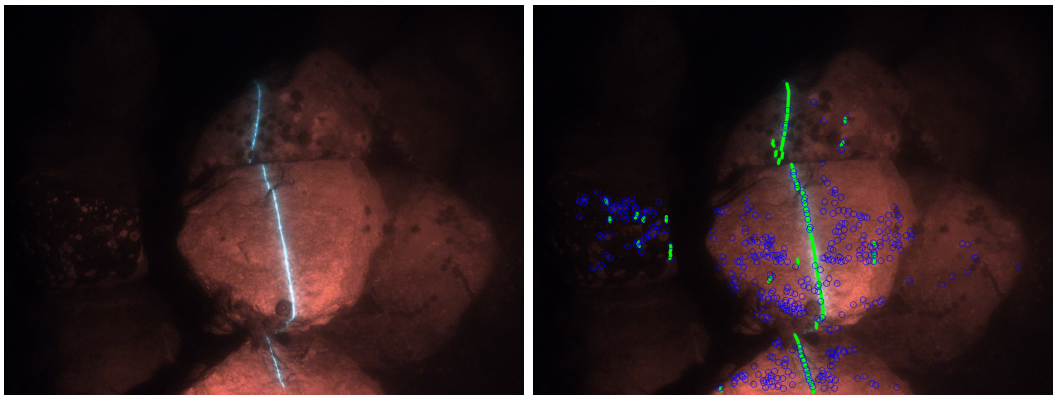
Feature detection

All primary information for active and passive triangulation based vision systems are encoded in the image domain. Here, most information is carried by features with strong gradients in both image directions, as this allows to localize them in the image uniquely. In the case of projected light patterns, the baseline between camera and projector is usually known to imply additional epipolar constraints allowing to use light patterns having a strong gradient only orthogonal to epipolar lines like, for example, a simple laser line. Here, a structure from motion system combined with such a line structured light system requires to detect passive-features with strong gradients in both image direction in addition to the line pattern projected by the line projector. This detection can be achieved by using standard feature detectors like SIFT, ORB, SURF, etc. in addition to a laser line detector searching row/column wise for a local maximum to detect possible locations of the projected light pattern analog to the approach discussed in Section 5.3. As both feature detectors are executed on the same raw image, overlapping feature sets will be generated because the projected laser line might also have gradients in both image directions, and the scene might have line-like elements depending on the scene texture. Therefore, these sets must be filtered before used for 3D depth estimation. Here, an example of overlapping feature sets is visualized in Fig. 5.6.

Feature classification

Separating detected features into passive and active-features belonging to the projected light pattern is a base requirement to use the same physical camera for structured light and structure from motion applications. Here, the challenges faced by the classification step are:

- The shape of natural scene features is unknown.



(a) Raw Image (b) Feature Overlay

Figure 5.6: Active-features (green) and passive-features (blue) detected in a raw camera image.

- The 2D shape of projected light patterns is only known when leaving the projector and is subject to an unknown scene depth dependant projection.
- The brightness distribution of projected light patterns is altered by the medium water, the scene texture and unknown environmental lighting.
- Light patterns move depending on an unknown scene depth change preventing a matching between consecutive images.
- The light pattern is subject to occlusion depending on the scene structure and the projector camera geometry.

The goal is to find mutual information of the projected light pattern which is untouched by projecting the pattern onto a scene and imaged by a camera. In the case of a laser line pattern, the cross-section of the line usually follows a Gaussian distribution, which is defined by the laser diode and the line optic. In the case of no abrupt scene texture changes in the scale of the line width, the Gaussian distribution can be used as such mutual information for line feature detection and classification. Also, by limiting the allowed frequency of the scene depth variations, connectivity constraints can be defined between detected light pattern features, including a global constraint that each row/column can only have a single valid light pattern feature. In the case of more complex light patterns, more connectivity constraints can be added but at the cost of increased complexity to associate a projector angle to each detected light pattern, which might also require tighter restrictions on the scene structure (Zeng and Zhang [2012]).

For passive-features, usually, more constraints are available as they are associated with 3D scene points, which are fixed with respect to the global coordinate system. This association implies that passive-features and their surroundings are only subject to global illumination changes and perspective transformations, allowing for tracking the 3D scene points across image sequences by matching their corresponding projections imaged as passive-features. These passive-features tracks must be in accordance with epipolar constraints implied by the motion of the camera, allowing to filter out wrong feature tracks. In general, the classification of valid passive-features is analog to any structure from motion system and is usually referred to as an outlier rejection schema. However, in the case of a projected light pattern, the feature distribution gets biased as many features from the light pattern might be picked up, which all move in accordance to the unknown scene depth change. Therefore, the standard outlier rejection is prone to be less effective, leading to wrong pose estimations in challenging environments. To avoid that many active-features are entering the outlier rejection step of the structure from motion system, the following steps are proposed:

- Mask all possible active-features based on the assumption that the cross section of the projected laser line follows a gaussian distribution.
- Match remaining features with features from previously images.
- Reverse matching and match features from the previous image with the current one and filter out all matches which are not in agreement.
- Use remaining feature matches in combination with a standard Random Sample and Conses Framework (RANSAC) to find a set of minimal features explaining the motion of most of the remaining features enforced by epipolar constraints.
- Remove all features not in agreement with the RANSAC step.

This approach allows to separate detected features into, active-, passive-features, and outliers. Here, additional outliers can be detected after the motion of the system is recovered.

Pose and 3D scene recovery

Based on the found passive feature matches, the poses of the camera can be recovered using, for example, the five-point algorithm to find the initial pose, followed by an iterative refinement to geometrically minimizing the re-projection error of all

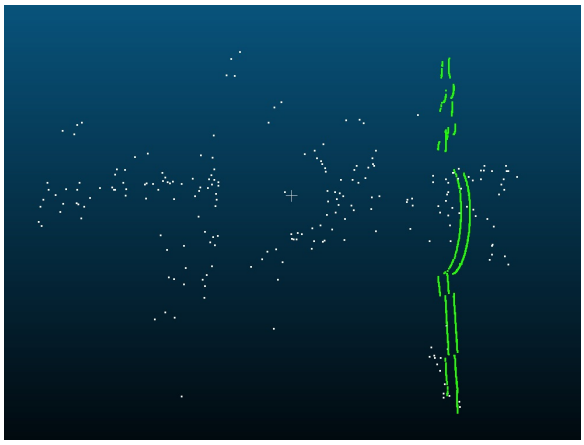


Figure 5.7: Sparse 3D points recovered by a structure from motion system based on two images in addition to the two line profiles simultaneously recovered by a structured light system using the same input images.

triangulated points explained in Section 5.2. This approach allows performing a 3D scene reconstruction up to scale as the amplitude of the motion scales with the size of the observed object and vice versa. This correlation is similar to an image showing a house where it is impossible to tell if its a dollhouse or a real house if no other information is available even when imaged from multiple locations.

In addition to the passive-features, each image also contributes 3D measurements based on the structure light system, which are metrically correct. However, they are only loosely coupled to the 3D scene structure estimated by the structure from motion system as it is usually not possible to track passive-features through image regions where active-features are present. This circumstance is visualized in Fig. 5.7.

3D scale equalization

The structured light system only delivers 3D scene measurements without reliable information about the scene texture because the light pattern dramatically changes the illumination at these locations. Therefore, the only reliable information which can be used for matching is their 3D coordinate. To match these 3D points with the sparse scene structure recovered by the structure from motion system, the scene structure is re-projected onto each image, taking the camera pose estimations into account. This re-projection cancels out the arbitrary scale of the scene structure estimated by the structure from motion system. Here, for each re-projected point that intersects with an active feature a range constraint is added to the iterative refinement of the last n poses, also referred to as windowed bundle adjustment. The underlying assumption

is that the camera moves through the scene so that at some point, formerly tracked features intersect with the light pattern of the structured light system. However, because the light pattern usually prevents the tracking of passive-features in the same region, the pose estimation of the camera must be used to transfer features from previous images to the regions covered by the light pattern. Although the scale of the reconstruction does not matter at this point, as the re-projection removes the scale dependency, any pose error will eventually introduce wrong associations between the structured light and the structure from motion system. To counteract wrong associations, a robust loss function is used, which reduces their influence in cases of large residuums s. Here, for example, the Cauchy Loss function ρ can be used to damp large errors defined as:

$$\rho(s) = log(1 + s) \tag{5.3}$$

In addition, to increase their influence even further, constraints having large residuums are automatically removed after a windowed bundle adjustment was performed and before new constraints are added to the optimization problem.

5.5 Extended Fusion

The fusion explained in the previous section is algorithmically complex but based on an established vision pipeline for motion tracking. Under suitable conditions, such a system can perform surprisingly well and produces dense metrically correct 3D reconstructions with a sub-millimeter resolution like shown in Chapter 6. However, the structure from motion system suffers from the projected light pattern as it reduces the image area available for feature tracking and prevents continuous tracking of passive-features throughout the image sequence. This reduction removes many constraints used by the motion estimation and reduces its overall robustness to outliers and noise. To counteract this, several improvements to the software and hardware level are addressed in the following sections.

Software Improvements

A critical limitation for fusing active with passive vision sharing the same sensor element is that each image region can only contribute either active or passive-features but not both. Therefore, 3D points from the structured light system cannot directly be linked to measurements of the structure from motion system. As a direct consequence, feature tracks are cut into at least two segments. The first segment consists

of the projections of a 3D point before and the second one of projections after the 3D point was hit by the light pattern. Therefore, the 3D scale equalization step can be further improved by not only associating feature tracks to active-features but also other passive features. This association is possible if they intersect on the image plane during re-projection, and a distance measurement was associated, indicating the track was interrupted by the projected light pattern.

In theory, the distance constraints can also be used to rate the RANSAC hypothesis from the structure from motion system. This rating can be performed, by validating the depth of triangulated 3D points using their cross-ratio and compare it with the cross-ratio of associated distance constraints canceling out the arbitrary scale. However, experiments have shown that usually not enough distance constraints are available to improve the pose estimation significantly. The reason for this is that usually around 100 up to 10000 passive-features are tracked, covering not more than one percent of the total pixel area. Therefore, the likelihood for a light pattern intersecting with this sparse point set is very low, which leads to only a few distance constraints on a frame per-frame basis. A statistic about how many distance constraints can be generated for different scenes is given in Section 6.4. Here, only a semi-dense reconstruction would yield enough constraints to not only fix the scale of the structure from motion system but also to significantly contribute constraints to the motion estimation. However, a dense reconstruction, based on a single camera, is usually very computationally demanding. It also uses the photometric error between different patches directly, like discussed by Ummenhofer and Brox [2012]. Therefore, it is even more sensitive to artifacts due to the projected light pattern than features with sharp gradients and is usually not used for marine applications.

In the case the system is in motion, active-features usually move differently to passive-features. The reason for this is that passive-features move according to the scene structure, and the sensor motion, whereas active-features only depend on the scene depth change. These different movements can be used to improve the detection of active-features over a sequence of images. But there always exists a critical motion vector which will lead to a situation where active and passive-features move the same way for a particular scene. For example, assuming no rotation and that all scene points are lying on a horizontal plane with respect to the camera. The following equation would define the critical motion vector with C being the baseline between camera and projector and $X = (X_1, X_2, X_3)^T$ the 3D coordinate of passive-feature which moves like an active-feature on the plane.

$$\vec{v}_{critical} = \begin{pmatrix} 1 \\ 0 \\ \frac{X_3}{X_1 - C} \end{pmatrix} \tag{5.4}$$

Therefore, this constraint can only be used to score light pattern features rather than filtering them out. Here, the assumption is: If an active-feature moves differently to its surroundings in the image domain, the likelihood is higher. It belongs to the light pattern and not to the scene texture than the other way around. Following this, after the camera motion is recovered up to scale, the following method can be used to score active-features:

- Match detected active-features with previously detected ones using epipolar constraints implied by the recovered sensor motion. This also supports matching of active-features lying on a line orthogonal to the sensor motion.
- Transfer matched features to a third image using a trifocal tensor.
- Calculate the geometric distance to the closest active feature in the third image and use its norm for rating the feature.

Here, in the case, an active-feature moves like a passive-feature. The distance to the predicted location is close to zero as the sensor motion model can explain it. This circumstance significantly improves the noise removable from a structured light system by filtering out all segments that norm is smaller than $X\%$ of the camera motion vector. In case, this step suppresses more than $Y\%$ of the active-features, they are restored. This approach avoids scenarios where the sensor is wrongly reporting free space for the entire workspace due to a critical sensor motion. But at the same time, it filters out scene textures with a similar appearance to the projected light pattern.

However, like discussed, due to the inability to simultaneously measure passive and active-features for the same scene region like, for example, possible with RGB-D sensors, no direct constraints can be defined, improving the motion estimation step. For this, either additional assumptions constraining the scene-depth change must be made, or further hardware improvements are required.

Hardware Improvements

A single-camera combined with a constant light projector, having a fixed baseline with respect to the camera, is the simplest hardware setup possible to perform active-passive sensing. But, it also generates only very few cues that can be used for fusing both modalities. More advanced hardware setups allow to increase the number of cues and to increase the robustness of such systems as more measurements can be taken into account, suppressing outliers and noise. Here, mainly three options are available which all use the basic building blocks explained above but utilize more constraints due to an increase in the number of measurements or an improved separation in the time, spectral or spatial domain:

- Multi camera setup increasing the number of measurements.
- Colored light patterns in combination with colored illumination.
- High speed camera in combination with modulated light patterns.

A multi-camera setup allows to directly validate all active-features as the depth estimation by a multi-camera system must be in agreement with the structured light system. This circumstance is similar to a trifocal system where one camera is replaced by an inverse camera (projector). Also, it allows triangulating 2D features without having to estimate the pose change of the system simultaneously. Therefore, adding at least one additional camera solves the classification of features directly into active and passive-features. It also considerably improves the robustness of the pose estimation by adding other range measurements and the ability to generate dense 3D measurements for each captured stereo image pair. However, adding additional cameras comes at the cost of at least doubling the computational complexity and usually a reduced field of view as only a subset of the volume imaged by one camera is also imaged by the other cameras. Also, like shown in Section 3.1, the epipolar planes between two flat port cameras are no longer imaged as straight lines but as curved ones making it more demanding to perform feature matching using epipolar constraints.

In the case of modern color cameras, the sensor is usually covered with an optical color filter array composed of the three band-pass filters green, red, and blue. This arrangement is generally called a Bayer filter and adds the ability to sample the spectral domain at the cost of a reduced spatial resolution. When using such a sensor, the tracking of features can be performed in one color channel while the structured

light system is using a different channel. However, although the performance increases in certain situations, there is still considerable cross-talk as the color channels overlap, and strong light patterns leak into the other color channels. Also, experiments have shown, that usually not all color channels are equally practical and must be chosen depending on the object range and water quality to find a possible combination. Here, in good conditions, accurate 3D reconstructions are possible without substantial cross-talk between color channels, as shown in the next chapter using the fusion described above. The only difference is that for feature classification, multiple color channels are evaluated, and passive-features are only masked if an active feature was found in the same color channel. Here, in the case the raw Bayer image is used as input for light pattern detection, all channels can be detected simultaneously, including none-maximum suppression across the spectral domain.

An improved separation between structured light and structure from motion can also be achieved in the time domain. Here, the light pattern is modulated in time in a predefined manner, and range measurements are assigned to feature tracks only tracked in images with no light pattern present. This approach increases the pressure onto the feature tracker as it has to match features that are further apart. But in case, state of the art low light cameras are used frame rates higher than 140 frames per second can be achieved. Here, according to the work of Handa et al. [2012] there is a considerable gain between feature tracking with 40 and 100 frames per second due to less motion blur. However, the frame-rate cannot be arbitrarily pushed up due to image degradation. Therefore, the optimal frame rate minimizes the motion blur and feature movements on one side while maximizing the signal-to-noise-ratio of the signal, on the other hand, requiring longer exposures. In the case of interleaved feature tracking, the frame rate is halved but without introducing additional motion blur. Therefore, the impact is less prominent as laid out in the research work, not decoupling exposure time from framerate. However, the effect onto the structured light system is directly noticeable as reducing the amount of frames with an active light pattern reduces the number of available range measurements. To compensate for a drop in range measurements, more complex strobe patterns can be used. One possible schema is, for example, to alternate two light patterns imaged at different image regions. Using linear optical flow constraints for groups of three images, the features can be tracked throughout these regions without requiring an accurate sensor motion estimation. Also, the background signal for these regions can be measured while the light pattern is not present, which considerably improves the detection of light patterns.

5.6 Conclusion

In the simplest case, a line laser is fixed with respect to a single camera, and the whole system is moved through the scene to perform dense metrical correct 3D reconstruction with sub-millimeter resolution. This approach works considerably well for situations where the visual motion tracking has enough features to track to compensate for the loss of features due to the active light pattern masking parts of the scene. Here, although the scale can be recovered with the help of the active-system, no generalized constraints can be added on a software level to restrict the motion of the system. Adding additional restrictions is either possible if scene texture and range measurements can be performed for the same image region at the same time. Or if the light pattern is covering more extensive areas of the image, producing overlapping depth measurements which can be utilized for motion estimation. However, this is usually not advised in the underwater domain as even line lasers are already contrast limited (Jaffe [2010]), and using more complex light patterns would decrease the contrast even further. On a software level, the structured light system benefits from the additional noise removal, which can be performed due to the knowledge about the sensor motion. However, the structure from motion system cannot efficiently utilize 3D range measurements produced by the structured light system. Here, the sparse feature set of the structure from motion system allows matching only a few correspondences between the active and passive system on a frame by frame basis. These correspondences are usually just enough to fix the scale of the structure from motion system with respect to the structured light system.

By using more advanced hardware setups, it is possible to separate the motion tracking from the structured light in the time spectral or spatial domain without reducing the density of the final reconstruction. Here, the most promising solution is to use a time-coded structured light system that allows tracking scene features with at least half the image framerate. Here the performance of the motion estimation is close to a purely passive one as the only difference is that passive-features covered by the light pattern must be interpolated between images, which are one frame apparat. Further improvements can be expected by adding a camera for generating more constraints without enforcing a strict stereo setup. This approach can be seen analog to the work of Solà et al. [2007] stating, "Two times mono is more than stereo".

By ignoring the reduced robustness against noise or environmental illuminations, a simple setup consisting of a camera and a line laser can achieve impressive results

in real-time, which cannot be achieved using a single camera or a standard line structured light system. The reason for this is that on one side, a camera system allows accurate motion tracking but has an inaccurate depth estimation, especially in regions with no scene texture. On the other hand, a line structured light system can accurately estimate the depth along the projected line but fails to reference them with respect to each other over time. Combining both modalities results in consistent 3D reconstruction with sub-millimeter resolution. As the system uses only the visual spectrum of light and minimizes the backscattering by utilizing line patterns, the system is well suited for underwater applications. Here, using a passive-active vision system allows tracking the motion in the same sensor domain as the active depth measurements are performed, improving the accuracy of the 3D reconstruction. Although the envelope under which conditions such a system can be operated is rather small and many factors must be considered to achieve a successful deployment. To increase this envelope of operation, more advanced hardware setup can be used without changing the underlying working principle. Here, a comparison between different configurations and the achieved accuracy is discussed in the next chapter.

Chapter 6

Experiments

The previous chapters described several methods for underwater visual scene reconstruction using passive-active image features. Here, mainly analytic approaches are used, including simulated sensor data, to evaluate the sensor models. Therefore, to verify the software implementation and the alignment of simulated with real sensor data, several experiments are described in this chapter designed to answer the following two questions:

- Do-real world experiments support the theoretical consideration or are relevant effects present, which were neglected?
- How do different parametrizations and system configurations affect the overall accuracy of the system?

The conducted experiments are ordered based on their complexity and start with single components such as the flat-port camera model or the line projector model before they are aggregated to more complex setups. Here, for each experiment, the objectives, experimental setup, and calibration procedure are described, and the results are discussed.

6.1 Flat-Port Camera Model

A pinhole camera behind a flat-port converts to an axial camera when being submerged as discussed in Section 4.2. This conversion leads to additional achromatic and geometric distortions. Here, the achromatic distortion cannot be completely compensated in software as each color channel has a certain bandwidth of around $150nm$ which will lead to an additional image blur for each color channel at the image edges. Monochrome cameras are even more affected as each image pixel receives

photon from the full visible spectrum of light. However, the geometric distortion can be compensated if the model parameters of a flat-port camera are precisely known. Therefore, to verify the flat-port model is in agreement with the physical one, the following experiment is carried out.

Objectives

Estimate the flat-port camera model parameters for different camera setups based on calibration images and compare them with manual measurements. Here, the flat-port camera model should be able to correctly estimate the pose of the camera with respect to a flat-port utilizing light refraction.

Experimental Setup

The model parameters of a real underwater flat-port camera are difficult to change without opening its watertight housing. Therefore, the setup is simplified by using an aquarium instead of a submerged flat-port housing shown in Fig. 6.1. The aquarium is filled with fresh water and acts analog to a flat-port housing for a camera on the outside of the aquarium. On the inside, a planar calibration target is placed, which can be moved around. By adjusting the position of the camera on a rail system, the model parameters of the mocked flat-port camera can be modified in a controlled manner allowing to test if the camera model can track these changes using images of the calibration target at different positions as measurements.

(a) Side view (b) Top view

Figure 6.1: Camera in front of an aquarium window mocking an underwater flat-port housing.

In general, a flat-port camera model has seven additional degrees of freedom in comparison to a standard pinhole camera model, whereas, to improve convergence, usually four of them are fixed as they are known or can be independently measured with high accuracy including:

- The refraction index of the medium insight the flat-port housing which is usually air having a refraction index close to one.
- The refraction index of the flat-port which is known for most materials usually ranging from 1.52 for ordinary glass to 1.77 for sapphire glass.
- The refraction index of the water body which can be independently measured using a refractometer.
- The thickness of the flat-port according to the datasheet of the interface.

The remaining three model parameters are the flat-port normal with respect to the principal axis of the camera and the distance of the focal point to the flat-port. During the experiment, these parameters are changed in the following manner:

- The distance between the lens of the camera, and aquarium window is adjusted to $1mm$, $10mm$ and $20mm$.
- The camera axis is set to $0°$ and $4°$ with respect to the normal of the aquarium window.

System Calibration

The camera is calibrated in-air without the additional aquarium window as its effect onto the geometric distortion is smaller than the noise of the system, as shown, for example, in Fig. 4.22. Based on the estimated intrinsic model parameters of a standard pinhole camera, including lens distortion, labeled in Tab. 6.2 as Pin-Air, the remaining flat-port model parameters are initialized for the orthogonal case and refined using images of the visual target located inside the water-filled aquarium. For the refinement, the re-projection error, outlined in Section 4.2, is used as a residuum to optimize the model parameters after each physical modification of the arrangement. Here, as measurement, ten different visual target positions are recorded for each modification. The resulting refined camera model parameters are labeled as Flat-Mixed-X, where X encodes the actual position of the camera with respect to the window of the aquarium. In addition, the same image sequences are also used to

calibrate a pinhole camera model for each pose implicitly minimizing the refraction error. These refined pinhole camera model parameters are labeled analog to the flat-port camera as Pin-X.

Results

In general, it is difficult to directly measure the exact location of the focal point of a camera lens, which is usually insight the lens housing. Therefore, the camera is placed $1mm$ in front of the aquarium window, and the estimated distance between the focal point and window of $4mm$ is used as an offset for the other measurements. Here, the relative displacement of the camera along its axis is correctly tracked with an accuracy better than $1mm$ for all measurements reaching the measurement limit of the physical setup and is given in the Tab. 6.2 as the first value of the layer thickness for each flat-port camera. Whereas, the second value of the layer thickness encodes the thickness of the aquarium window, which is constant during the whole experiment.

For a distance of $20mm$ between lens and window also the angle of the camera to the normal of the glass layer is changed from 0° to 4°, which is estimated as 4.2° by the model. This estimate also lies in the tolerance of the physical setup and shows that the model correctly tracks the physical camera parameters, which is vital for accurate calibration.

Although the pinhole camera model does not explicitly model the physical arrangement, it can be used to approximate a flat-port camera to a certain extent. In this case, it uses all its parameters to minimize the un-modeled light refraction. Therefore, the best performance can be expected for the orthogonal case when the focal point has an optimal distance to the flat-port. This expectation is supported by the residuum of the pinhole model listed in Tab. 6.1. Here, the pinhole camera for $1mm$ distance to the window has a surprisingly small model error, which is even smaller than the flat-port model. However, the pinhole model was fully optimized based on the underwater image sequence, while for the flat-port model, some of its parameters were transferred from the in-air case. Also, the used aquarium has only a maximal working range of $0.4m$, which a pinhole camera model can well approximate. This circumstance is usually not given for larger working ranges as the light refraction introduces a scene depth-dependant distortion term. Therefore, as discussed in Section 4.2, there are many underwater applications which can be carried out with standard pinhole camera models. However, their performance mainly depends on the physical placement of the camera with respect to the flat-port, the working range, and

good in-situ calibration. Whereas flat-port camera models can be pre-calibrated and adjusted to the actual water body, have a range independent distortion, and support arbitrary camera placements with respect to the flat-port.

Parameter	**Pin-Air**	**Pin-Water-1mm**
Focal Length	1112.70 / 1112.58	1496.55 / 1493.91
Principle Point	625.64 / 443.93	624.47 / 432.97
Distortion K1,K2	-0.0489 / 0.1037	-0.007 / 1.6207
Distortion K2,K3	-0.4781 / 0.6206	-14.5986 / 41.3812
Mean Residual in pixel	0.1394	0.1106
Parameter	**Pin-Water-10mm**	**Pin-Water-20mm-4deg**
Focal Length	1490.83 / 1488.66	1493.26 / 1488.06
Principle Point	628.78 / 430.42	546.08 / 431.00
Distortion K1,K2	-0.0178 / 1.1952	0.0171 / -0.4683
Distortion K2,K3	-9.8510 / 26.885	3.0338/ -5.9621
Mean Residual in pixel	0.145	0.218

Table 6.1: Estimated parameters for the pinhole camera model in relation to the physical placement of the camera.

6.2 Line Structured Light

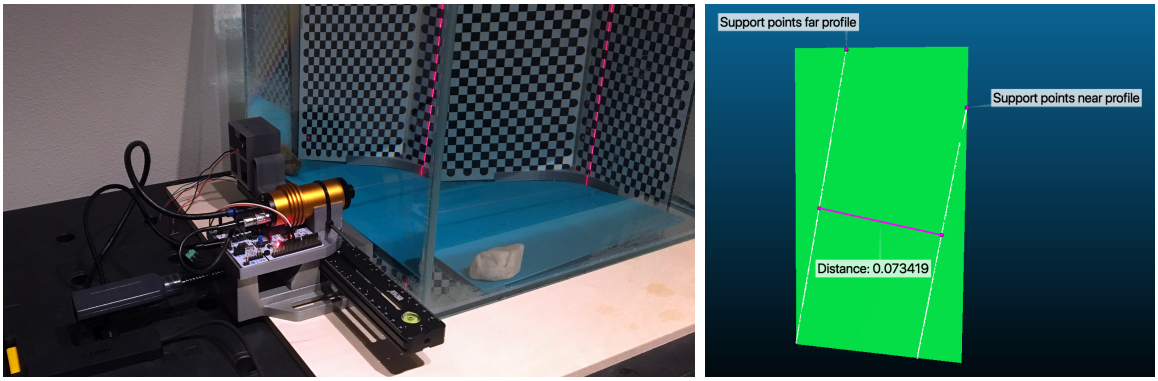
A line structured light system consists of a camera and a line projector, which can be interpreted as an inverse camera. Therefore, both optical systems are subject to refraction if placed behind a flat-port and being submerged. Here, like discussed in Section 4.3, in the case the projector is placed orthogonal to the flat-port, the plane of refraction is identical with the light sheet generated by the projector. Therefore, in this particular case, the straightness of the line pattern is conserved by the flat-port, and only the brightness distribution along the line is slightly changed. However, in the case the line projector is non-orthogonal to the flat-port, the light sheet is geometrically distorted and appears as a curved surface as the plane of refraction no longer coincides with the light sheet visualized in Fig. 6.4.

Objectives

In Section 4.3 it is concluded that the distortion of the light sheet due to a non-orthogonal orientation of the projector with respect to the flat-port can be modeled with the help of first-order tangential distortion for small angles. Here, the objective is to measure the model error for several laser camera orientations and to verify that

Parameter	**Flat-Mixed-1mm**	**Flat-Mixed-10mm**
Focal Length	1112.70 / 1112.58	1112.70 / 1112.58
Principle Point	625.64 / 443.93	625.64 / 443.93
Distortion K1,K2	-0.0489 / 0.1037	-0.0489 / 0.1037
Distortion K2,K3	-0.4781 / 0.6206	-0.4781 / 0.6206
Interface Normal	[0.0020, -0.0077, -1.0]	[0.0004, -0.0110, -1.0]
Refraction Index	1.0,1.52,1.330	1.0,1.52,1.330
Layer Thickness	0.004, 0.006	0.013, 0.006
Mean Residual in pixel	0.1183	0.127
Parameter	**Flat-Mixed-20mm**	**Flat-Mixed-20mm-4deg**
Focal Length	1112.70 / 1112.58	1112.70 / 1112.58
Principle Point	625.64 / 443.93	625.64 / 443.93
Distortion K1,K2	-0.0489 / 0.1037	-0.0489 / 0.1037
Distortion K2,K3	-0.4781 / 0.6206	-0.4781 / 0.6206
Interface Normal	[-0.0015, -0.0117, -1.0]	[-0.073, -0.016, -0.997]
Refraction Index	1.0,1.52,1.330	1.0,1.52,1.330
Layer Thickness	0.023, 0.006	0.023,0.06
Mean Residual in pixel	0.125	0.127

Table 6.2: Estimated parameters for the flat-port camera model in relation to the physical placement of the camera. Parameters in **bold** are refined underwater while the other ones are borrowed from Pin-Air.



(a) Side-view of the experimental setup

(b) Laser sheet estimation based on two measured profiles

Figure 6.2: Setup for measuring the distortion of a laser line due to non-orthogonal orientation of the laser projector with respect to the glass interface.

the proposed model error is outside of the accuracy of sub-pixel peak detectors of the camera.

Experimental Setup

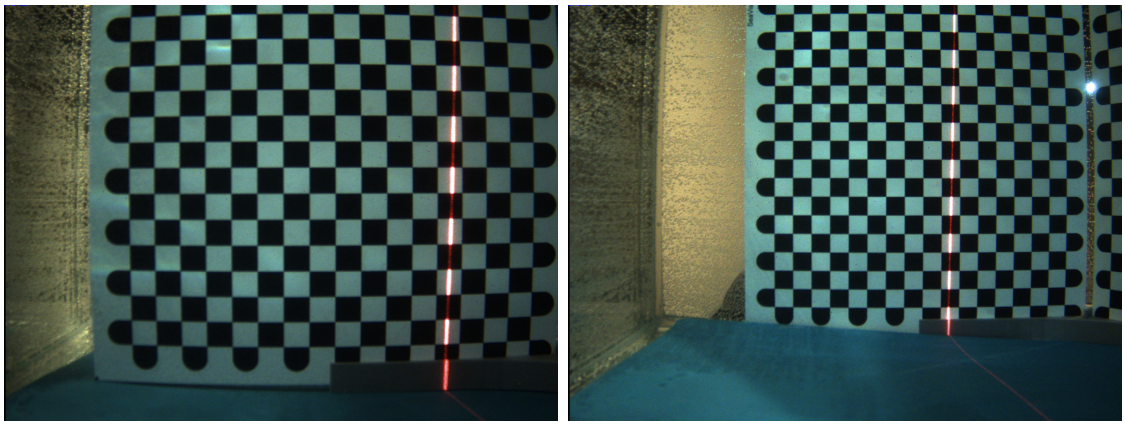
The underwater housing for the camera and laser are similar to the previous experiment, replaced by an aquarium filled with water acting as a flat-port. On the inside of the aquarium, a planar target is attached, reflecting the line laser to the camera. On the outside, the camera and the laser are placed on a rail system equipped with a stepper motor allowing to precisely control the rotation of the laser sheet with respect to the window without changing the physical baseline between camera and projector. An overview of the setup is shown in Fig. 6.2.

System Calibration

For the camera, the flat-port model parameters from the previous experiment ($1mm$ focal distance, orthogonal) are reused listed in Tab. 6.2. In addition to this, the projector parameters are estimated by projecting the laser line onto the visual target at two different target distances for each selected orientation of the laser line. Based on these measurements, two 3D laser profiles for each laser orientation are calculated by intersecting associated camera rays with the plane of the visual target. Following this, these two 3D profiles must also intersect with the distorted laser sheet and therefore describe its 3D shape. Here, as an example, the calibration images for the laser angle 0° are visualized in Fig. 6.3 and the corresponding 3D laser profiles with a fitted laser sheet are displayed in Fig. 6.2(b).

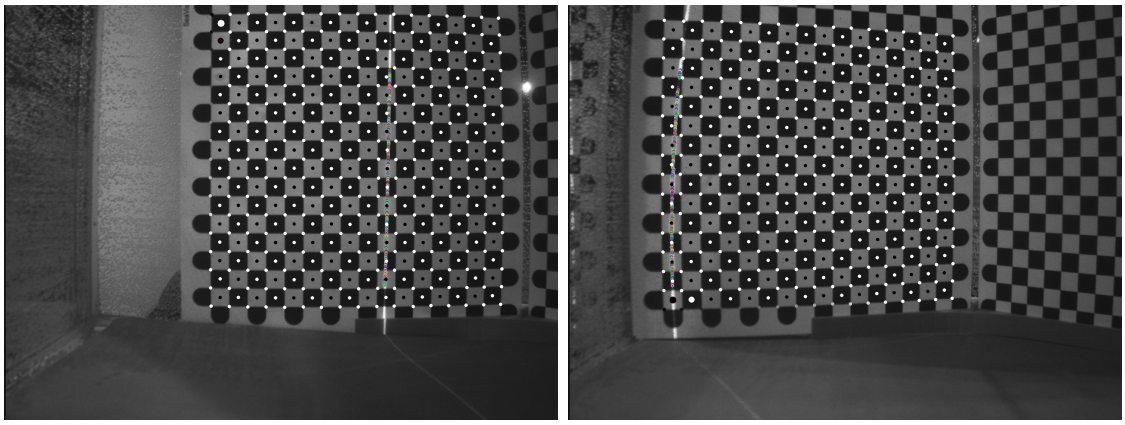
Results

The estimated projector parameters are compared with simulated ones using a flat-port model to back-project a virtual line onto two virtual planes at two distances. Following this, the back-projected points are used to calibrate a virtual line projector displayed as a reference in Fig. 6.4. Here, the effective baseline is defined as the minimal distance of the laser sheet in water to the focal point of the camera. Therefore, the effective baseline is not only dependant on the actual laser angle but also on the amount of refraction and where the interface is located with respect to the focal point of the projector. This further reduces the effective baseline between both systems like predicted and visualized in Fig. 6.5.



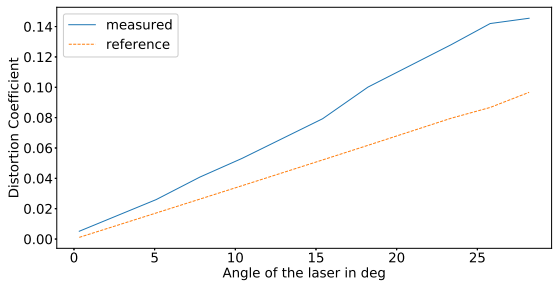
(a) Visual target at close range (b) Visual target at far range

Figure 6.3: Underwater calibration images for estimating the projector model parameters for the laser angle 0°.

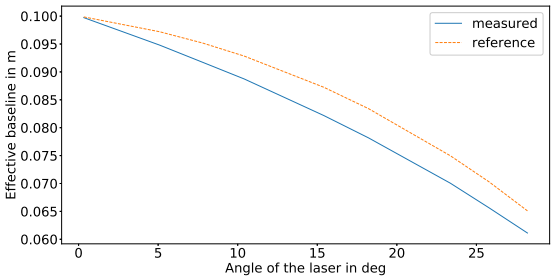


(a) Orthogonal (b) Non-orthogonal

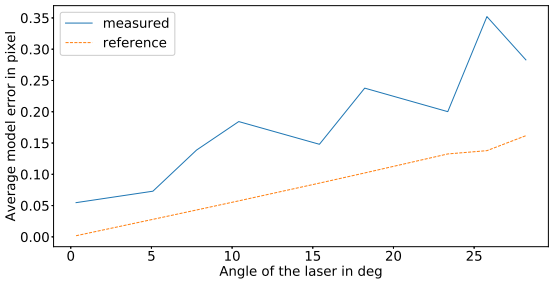
Figure 6.4: Distortion of the laser line due to non-orthogonal orientation of the laser line projector to the glass interface. Detected laser line points used for model refinement are marked in color, and black and white circles indicate detected points on the visual target used to calculate the pose of the target.



(a) Distortion Coefficient



(b) Effective Baseline



(c) Model Error

Figure 6.5: Distortion coefficient, effective laser baseline, and model error in relation to the angle of the line laser with respect to the glass interface.

Furthermore, the model error increases with an increasing angle between the laser and interface normal, like simulated in Section 4.3. Therefore, for values above five degrees, the model error is significant in comparison to sub-pixel peak detectors, and the simplified projector model should no longer be used. However, for standard underwater housings, the model gives enough margin to compensate for small misalignments between line projector and its flat-port. Here, even when manually mounted, usually values better than 1° can be reached.

6.3 Visual Odometry

One of the key aspects of an accurate 3D reconstruction is the ability to track the sensor motion and to fuse 3D measurements over time into one consistent representation. Here, the proposed passive-active visual system uses natural scene features for performing motion compensation while scanning. Therefore, one of its sub-components is a monocular visual odometry using passive image features to track the motion of the sensor. It is based on a windowed bundle adjustment, and its implementation is explained in Section 5.2. The component is embedded in the laser processing pipeline. However, it can also run standalone tracking the motion of the camera up to scale. To test its performance and to identify any systematic errors, it is tested against the

odometry benchmark KITTI released by Geiger et al. [2013]. The benchmark is an automotive vision benchmark with accurate ground truth and was selected to get a comparison with the broader vision community. In addition, ignoring refractive distortion, the challenges for in-air and underwater visual odometry are quite similar if difficult sequences like tunnels or heavy traffic are included. The dataset itself consists of eleven sequences with ground truth trajectories of a camera in different automotive scenarios for training and an additional eleven sequences without public available ground truth for evaluation. After an algorithm recovers the camera trajectories of the eleven sequences without ground truth data, they can be uploaded to a website to participate in an online ranking. Here, the scale of the trajectories must also be recovered, which is non-trivial as usually single-camera systems can only recover the camera motion and scene structure up to scale. Therefore, for this test, the estimated scale was injected into the windowed bundle adjustment by finding scene points on the street in combination with the known mounting height of the camera pose with respect to the street.

Figure 6.6: Image from the KITTI Benchmark - Sequence 13.

Objectives

Test the robustness and accuracy of the implemented visual odometry on a well-known vision benchmark.

Experimental Setup

Two-color and two mono cameras are mounted in a dual stereo setup on top of an automotive vehicle together with a GPS and LIDAR system for ground truth data. The cameras are synchronized at 10 Hz with respect to the laser scanner, and their data are provided as lossless compressed and rectified png sequences. Here, for the monocular data set, only the data from one camera are used with a dynamic exposure

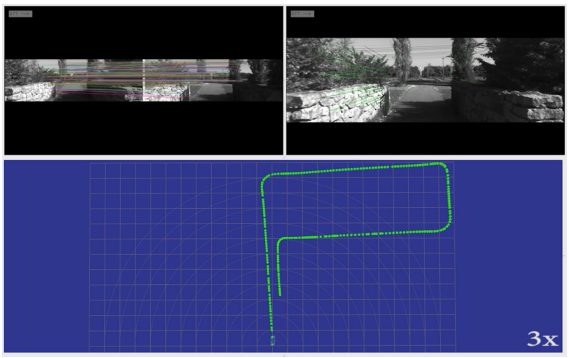
time, which is not allowed to exceed $2ms$. Further information about the raw dataset can be found in work from Geiger et al. [2013].

System Calibration

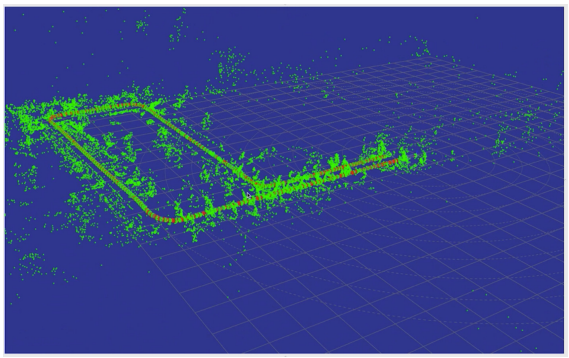
The camera images are cropped to a size of 1382 x 512 pixels and rectified by the authors of the dataset. Therefore, only the focal length, the principal point, and the location of the camera with respect to the road are required, which are provided with the dataset.

Results

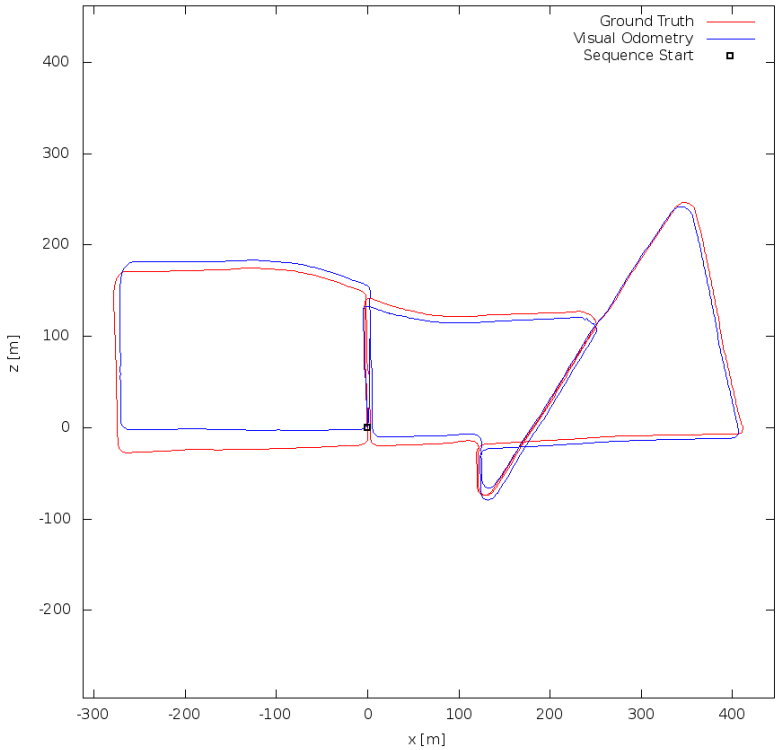
The implemented visual odometry can process all 22 sequences of the dataset consisting of 43552 mono images. The estimated camera trajectory, together with the ground truth, are displayed in Fig. 6.7 for sequence 13 and 14. The average translation error for all trajectories submitted to the KITTI web server for evaluation is 2.16% per traveled meter, which is an excellent result for a monocular vision system and proved that the implementation is robust to many different environmental conditions. However, monocular vision systems are still an ongoing research topic, and the current state of the art algorithms for automotive scenarios report errors as small as 0.8% per traveled distance. This level of accuracy would directly benefit the combined active-passive vision system as any pose error also appears as a 3D structure error in the final scene reconstruction.



(a) Estimated vehicle path for seqence 14



(b) 3D reconsturction for sequence 14



(c) Estimatead vehicle path and ground truth data for sequence 13

Figure 6.7: Estimated camera/vehicle trajectory and the resulting 3D reconstruction showing all tracked and triangulated scene features.

6.4 Combined Active-Passive Vision System

The most straightforward combined active-passive vision system consists of a camera and a line laser, which is identical to a classical line structured light system. However, instead of only using the active light pattern of the laser line, also passive features are used to estimate the pose of the camera while scanning.

Objectives

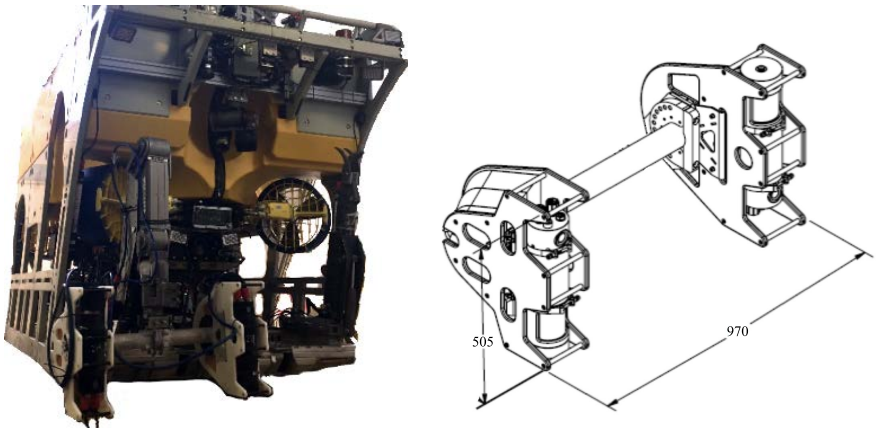
Estimate the accuracy of the proposed passive, active system in a real-world underwater scenario and generate a statistic of the available features.

Experimental Setup

The laser and the camera system are mounted on a remotely operated vehicle to scan the seabed with a line laser, which is mounted on a servo allowing to rotate the laser around its axis while the camera is keeping its orientation. This setup is identical to a classical line structured light system with the addition that the angle between the camera and the laser can be dynamically adjusted. This modification allows scanning whole scenes without requiring to move the sensor setup physically. However, to be able to scan objects in the water column or to avoid attaching sensor setups to the seabed, it is usually desired to scan objectives from moving sensor platforms such as remotely operated vehicles. Therefore, to simulate these scenarios, a visual target is attached to the seabed and scanned from a remotely operated vehicle while hovering introducing micromotion, which must be compensated to allow an accurate 3D reconstruction. For the compensation, the scene is illuminated with a red LED light while a green laser is used for scanning the scene. This separation reduces the crosstalk between active and passive visual features and increases the signal to noise ratio of the reconstruction. For the scan itself, 2300 images with 144 frames per second and an exposure time of $6.94ms$ are recorded with a Quad-VGA camera while the laser is rotated anti-clockwise.

System Calibration

The camera system is calibrated in-air and refined underwater using images of a checkerboard. Here, the refraction index is estimated based on the light refraction at a spot near the location where the experiment is conducted. In addition, the baseline between the laser and the camera system is estimated in-air using the calibration



(a) Remotely operated vehilce with attached sensor frame

(b) Sensor frame in mm

Figure 6.8: The sensor system consisting of two identical tubes mounted $970mm$ apart on a remotely operated vehicle by attaching it to a hydraulic arm. Each tube embeds a PC, a camera, a RGB LED, and a laser mounted on an internal servo which is electrically linked to the camera of the other tube. This linkage allows adapting the baseline between the camera and the laser of the system according to the mission requirements.

routine described in Section 4.4. The estimated and refined model parameters for the laser and the camera are given in Tab. 6.3.

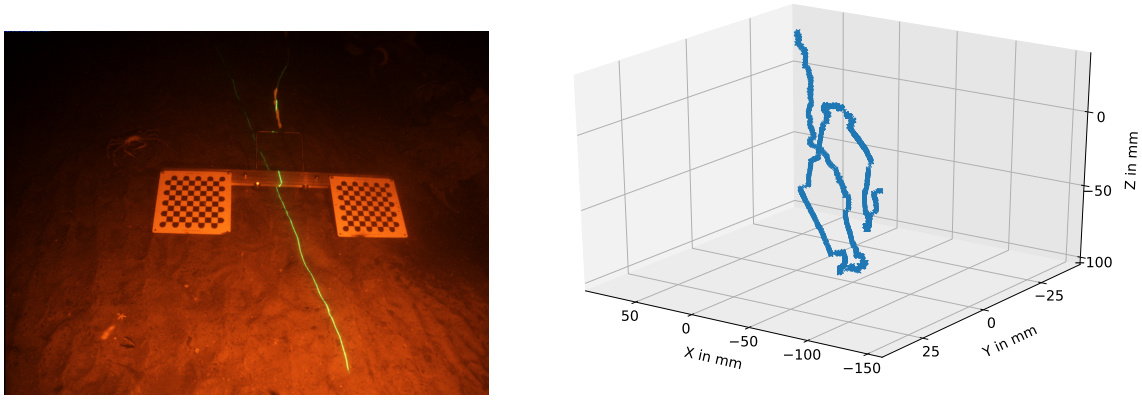
Results

The estimated camera poses for the recorded image sequence based on passive and active image features are plotted in Fig. 6.9 and Fig. 6.10. Here, no direct ground truth exists as underwater navigation systems usually cannot deliver position data in the mm range required for the 3D reconstruction. However, because the shape of the visual target is known, the deviation of the reconstructed target from the ideal one can be used as a precise error measurement. Here, the standard deviation to an ideal planar target is $5.15mm$ in case of an uncompensated scan and $1.86mm$ after motion compensation for an object distance of $2m$. Also, 3D points of the visual target are shifted insight the target plane resulting in imprecise distance measurements leading to an error of around 5% in the uncompensated scan and 0.8% after motion compensation. The resulting 3D reconstructions of the target, the signed distance to an ideal planar target, and distance measurements along the target are visualized in Fig. 6.13. Here, the motion compensation is effectively reducing the reconstruction error by a factor of around five. This reduction allows an accurate 3D reconstruction in

	Flat-Mixed	Projector
Focal Length	782.49 / 782.58	1.0 / 1.0
Principle Point	648.01 / 484.82	0 / 0
Distortion K1,K2	-0.037 / 0.0002	-0.02034 / -
Distortion K2,K3	-0.00045 / 0.002	- / -
Interface Normal	[-0.0004,0005,-1]	-
Refraction Index	1.0,1.77,1.337	-
Layer Thickness	0.007, 0.01	-
rvec	[0,0,0]	[-0.129, -0.104, 1.568]
tvec	[0,0,0]	[0.794, -0.001, 0.009]

Table 6.3: Flat-port camera and line projector parameters for mixed calibration (in-air + in-water refinement).

the millimeter range, which is usually not possible using floating sensor carriers opening interesting new applications such as 3D reconstructions of ship-hulls or mooring chains with millimeter accuracy.



(a) Camera Image

(b) Estimated camera poses

Figure 6.9: a) Demosaicked camera image part of an image sequence used to scan the visual target. b) The estimated motion of the camera while scanning using passive features for motion tracking and active features for scale enforcement.

For demonstrating the geometric error introduced by a physical flat-port, the same scene is reconstructed using a standard pinhole camera model. Here, the calibration of the pinhole is based on the underwater checkerboard images, which are also used to refine the flat-port model. In this complex scenario, including motion compensation, the pinhole camera performs well in the working distance similar to the distance of the checkerboard used for calibration. However, a deviation of up to $10cm$ can be noticed in the far-field due to un-modeled light refraction visualized in Fig. 6.14.

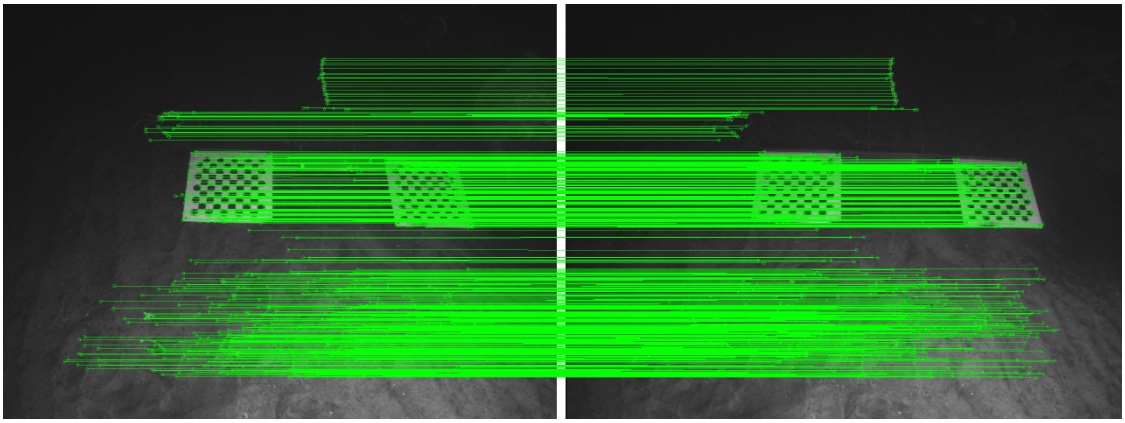


Figure 6.10: Feature matches between two consecutive camera images used for motion compensation while scanning.

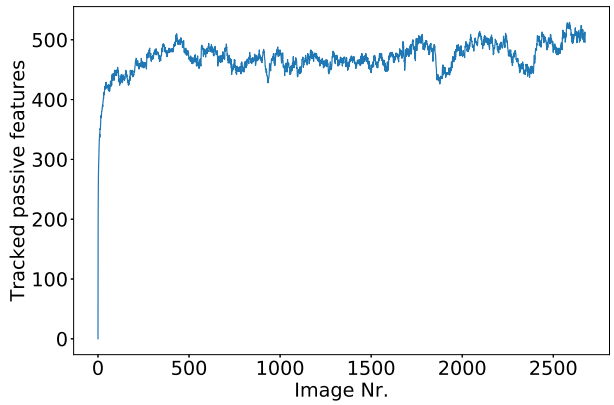


Figure 6.11: Number of currently tracked passive features over the the image sequence.

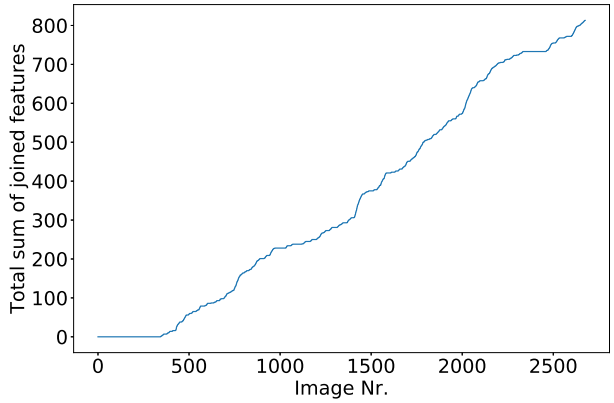
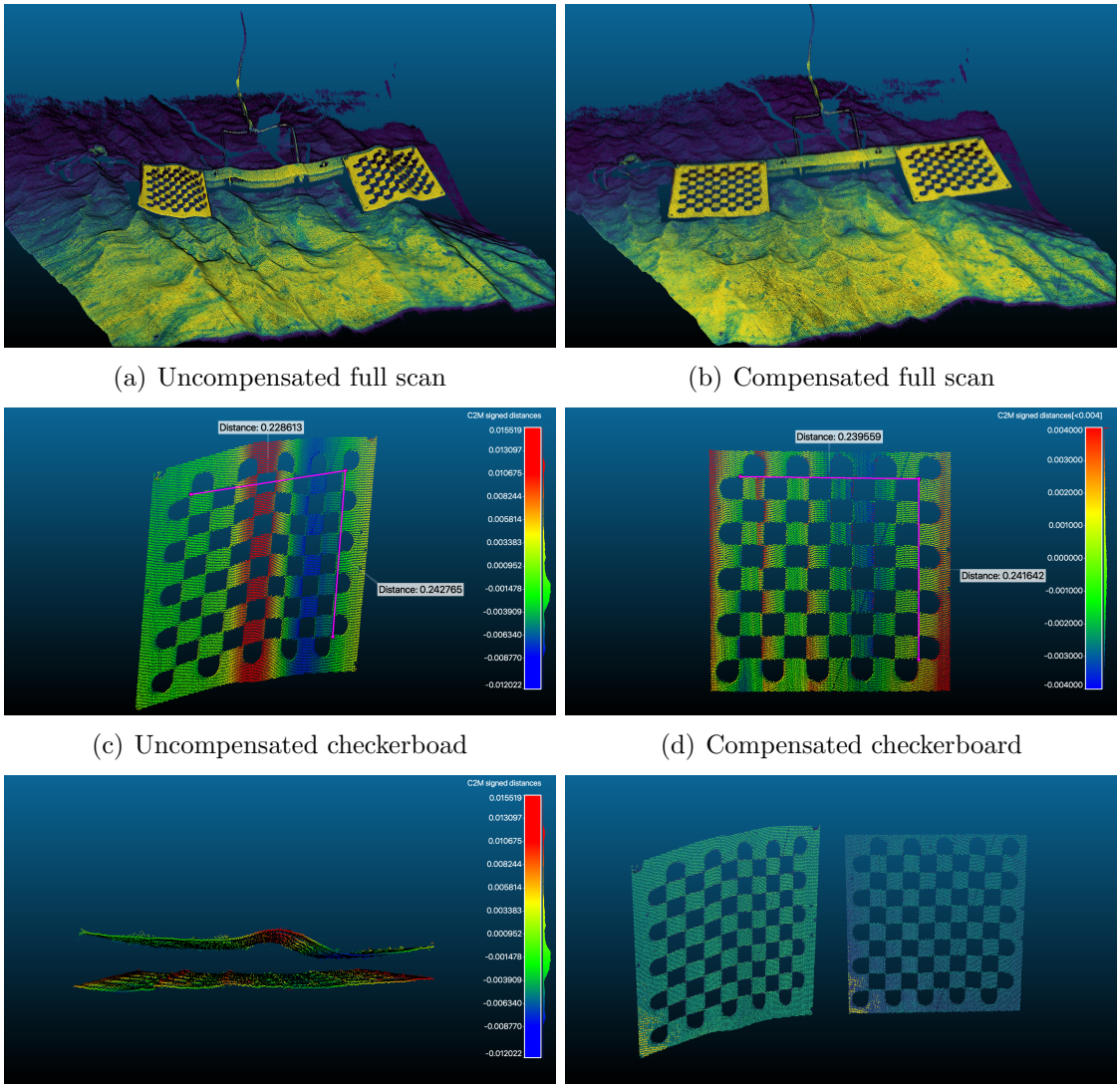


Figure 6.12: Total sum of active features joined with passive features over the image sequence.



(a) Uncompensated full scan

(b) Compensated full scan

(c) Uncompensated checkerboad

(d) Compensated checkerboard

(e) Comparision side view

(f) Comparison top view

Figure 6.13: Dense 3D reconstruction of a checkerboard with $3cm \times 3cm$ field sizes based on passive and active features. a) Raw scan without motion compensation. b) Scan with motion compensation c) Uncompensated checkerboard with signed distances to a fitted reference plane. d) Compensated checkerboard with signed distances to a fitted reference plane. e) Side by side comparison of the uncompensated and compensated checkerboard. f) Side by side comparison of the uncompensated and compensated checkerboard.

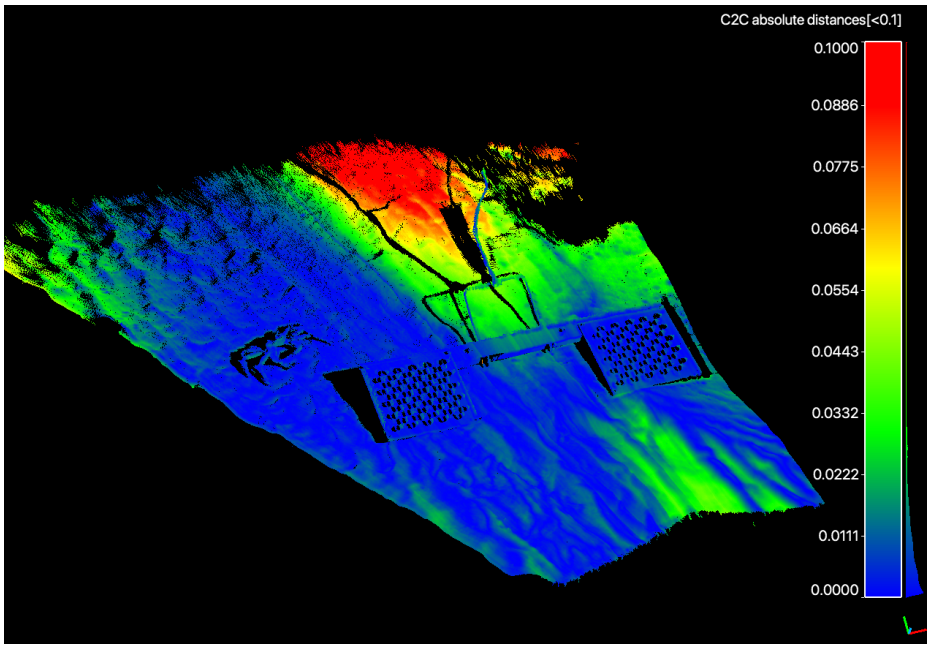


Figure 6.14: Absolute distance in meter between a reconstruction based on the pinhole camera model and a reconstruction based on the flat-port camera model using the same image sequence.

6.5 Ship Hull Inspection

The inspection of a floating object such as a ship hull or a pier is a challenging target for underwater inspection techniques. Not only does the object move relative to an observer, but it also influences state of the art navigation solutions due to magnetic distortions and multi-path effects. Here, as a technology demonstration, a ship hull is inspected with the help of a remotely operated underwater vehicle equipped with the proposed active-passive visual system.

Objectives

Test the proposed combined active-passive underwater system for the application of underwater ship hull inspections and identify its potential for further developments.

Experimental Setup

A modified Falcon from Saab SeaEye is used to inspect a medium-sized ship hull in the North Atlantic near St. John's in a harbor during nighttime to reduce the influence of sunlight on the measurements. The modified Falcon, shown in Fig. 6.15, is equipped with a high-grade inertial navigation system (Phins Compact C3), which

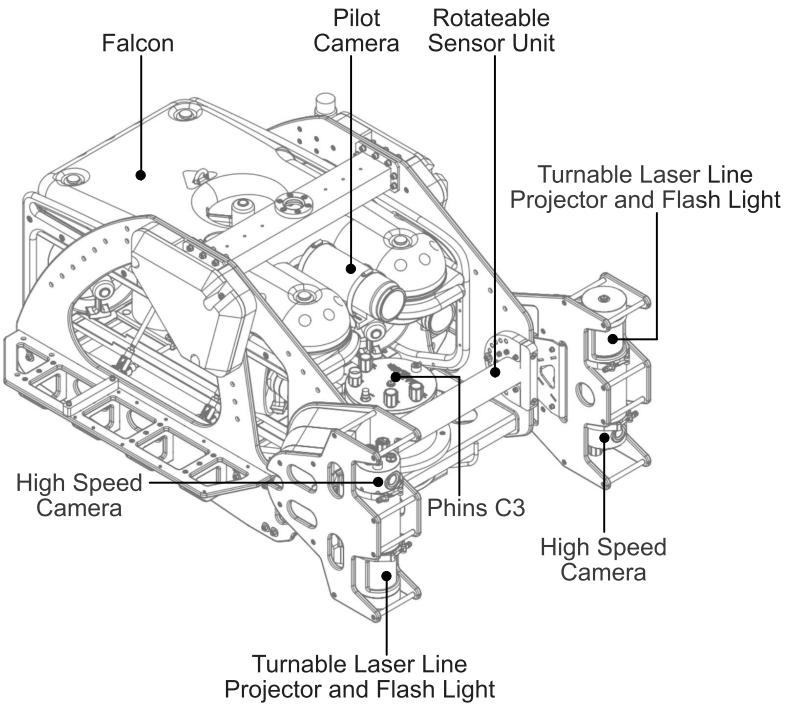
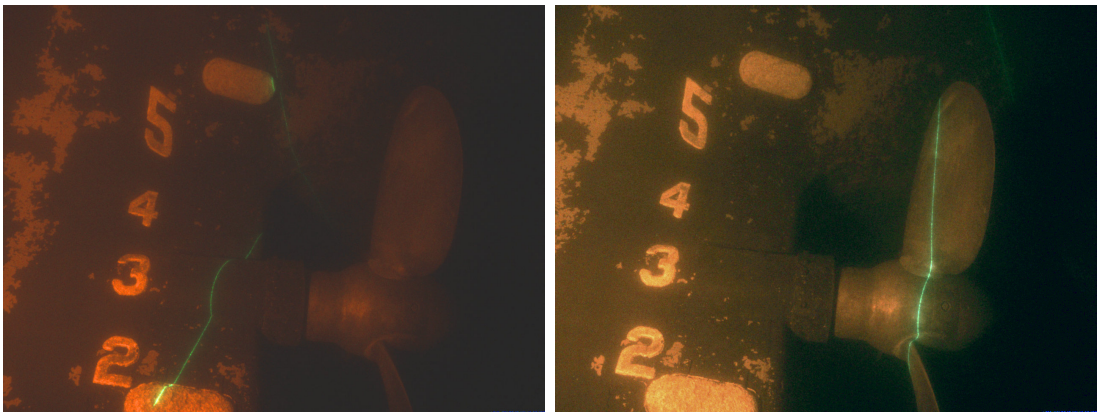


Figure 6.15: A Falcon from Saab SeaEye equipped with an additional inertial system (INS) and a structured light system.

uses a fiber optic gyro for measuring orientation and angular velocity. In addition, two identical pods, embedding each a camera and a steerable laser, are mounted at the front of the vehicle and electronically paired. This pairing allows the camera of the left pod to use the line projector from the right pod and vice versa. All systems are hardware time synced using a pulse per second signal (PPS) emitted by an additional GPS module.

With the help of the modified Falcon, several parts of the ship hull are scanned while hovering in front of the hull using auto depth. The auto heading is turned off because the compass of the Falcon is affected by the ship hull, and the Phins C3 is not fed back into the control system of the vehicle. In addition, to minimize the effect of the ROV light on the measurement, the target was scanned with a green laser while illuminated only with a red LED light. This illumination schema gives the pilot sufficient visual feedback to safely operate the vehicle near the ship hull.

For each scan, the laser line is swept across the scene from the left side ($-10°$) to the right side $+60°$ with an angular velocity of $5°$ using the servo the laser is mounted to. While sweeping, the camera is collecting raw Bayer images with $140Hz$, and the



(a) Camera image while scanning at t_1 (b) Camera image while scanning at t_2

Figure 6.16: Camera image while scanning showing the green laser and the red light for feature tracking.

navigation data from the Phins C3 are synchronously logged with $50Hz$. This setup results in around 2000 raw images for each of the scans with a scan duration of about 14 seconds. Here, two raw images of one of the scans are displayed in Fig. 6.16 after a De-Bayering step to get a color representation for the images. These images are also subsequently used for laser line extraction and triangulation, as well as for visual pose estimation.

System Calibration

The camera system is calibrated in-air and refined underwater using images of a checkerboard similar to the previous experiment. However, for the baseline estimation between the camera and the laser, a single scan is utilized using the servo to change the angle between the laser and the camera. This is possible because, due to the rotation of the laser, the planar calibration target appears at different ranges. By rotating these measurements back around the servo axis according to the servo angle, this generates multiple 3D profiles, which can be used for iterative refinement of the projector parameters without changing the location of the calibration target during the calibration process. Here, the calibration parameters for the camera and the projector used for the following 3D scans is given in Tab. 6.4.

Results

In the case of a ship hull inspection, the relative pose between the sensor and the ship must be precisely known to avoid errors in the resulting 3D data. However, this is

Parameter	Flat-Mixed	Projector
Focal Length	781.19 / 781.27	1.0 / 1.0
Principle Point	648.22 / 485.13	0 / 0
Distortion K1,K2	-0.040 / 0.008	-0.0093 / -
Distortion K2,K3	-0.0093 / 0.00567	- / -
Interface Normal	[-0.001,0006,-1]	-
Refraction Index	1.0,1.77,1.336	-
Layer Thickness	0.007, 0.01	-
rvec	[0,0,0]	[-0.106, -0.112, 1.590]
tvec	[0,0,0]	[0.810, 0.0174, 0.0113]

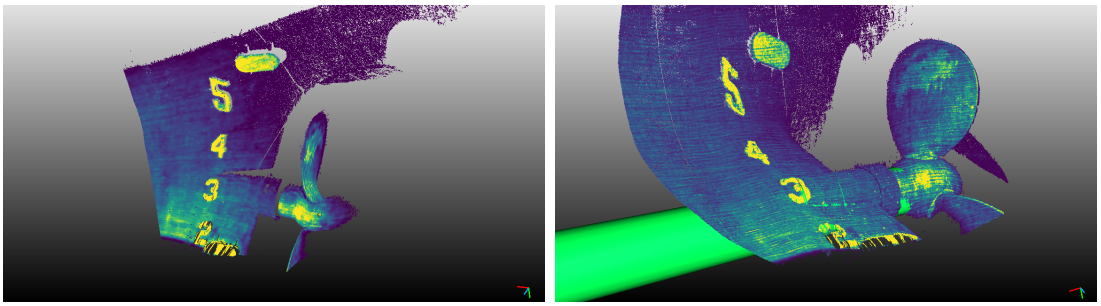
Table 6.4: Flat-port camera and line projector parameters for mixed calibration.

a difficult task, as underwater navigation solutions usually do not provide millimeter accuracy. Therefore, to compare the embedded commercial navigation solution with the pose estimation of the passive-active system, straight features of the ship hull are used as a reference value. Here, in the case of perfect motion compensation, straight features would also appear straight in the final 3D reconstruction. Therefore, any deviation from it is due to an in-precise pose estimation assuming the range measurement error on a per image bases is negligible.

As a first linear feature, the propeller shaft of the ship is scanned with the green laser using the servo to rotate the laser. While scanning, the ship is illuminated with the help of a red LED for visual motion tracking visualized in Fig. 6.16.

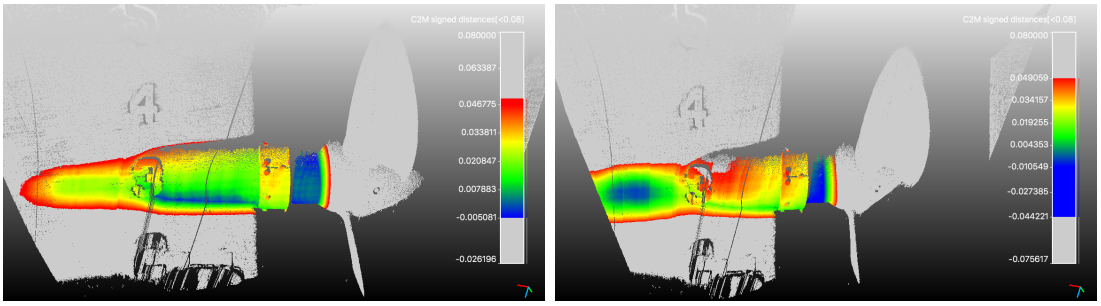
Following this, a cylinder is fitted to the resulting 3D data as a reference value. After this, the signed distance between the cylinder and the measured 3D point cloud is measured for motion compensation with the help of the inertial system combined with a doppler velocity log and for the embedded visual odometry. The result for both cases is displayed in Fig. 6.17 and 6.18. Here, the average error between the reference value and the measured one can be reduced by a factor of 2 in case the visual odometry is used instead of the inertial system.

As a second feature, the front of the ship hull is scanned with the same technique, including a chime, which is regarded as a slightly curved linear feature displayed in Fig. 6.19. Here, the sensor motion introduces a dent in the front section of the 3D reconstruction, which can be compensated by the visual odometry but not by the INS (Fig. 6.20). The deviation between both reconstructions in this section is above $4cm$ (Fig. 6.21). Here, assuming the structure can be approximated by a curved plane, the point cloud compensated by the visual odometry has an around five times smaller signed distance error than the one compensated by the INS (Fig. 6.22).



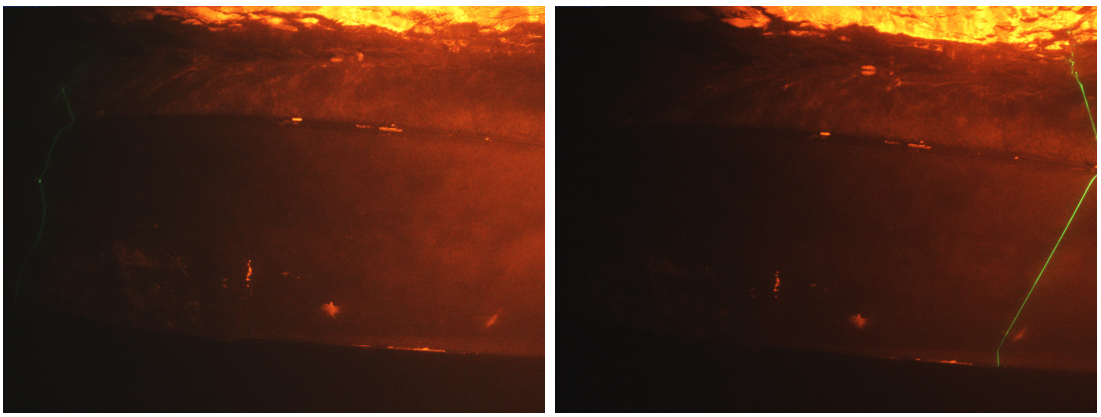
(a) 3D Point Cloud (b) Propeller shaft fitted to a cylinder

Figure 6.17: 3D reconstruction of a propeller using visual odometry for stabilization.



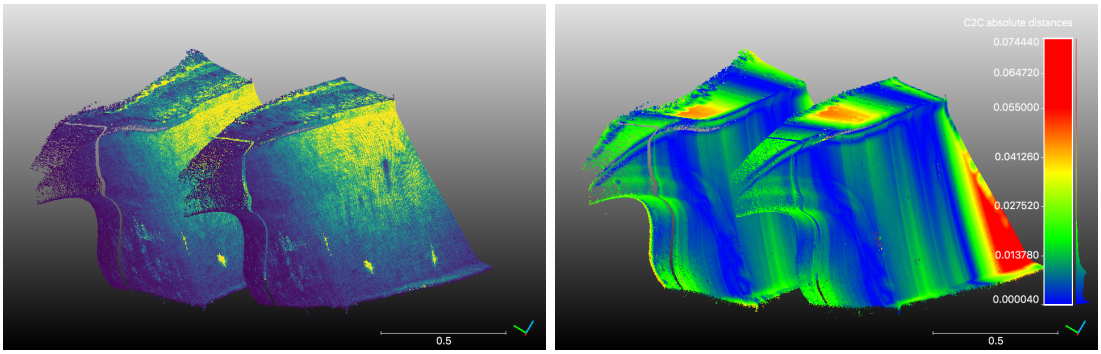
(a) Visual Odometry (b) Inertial Navigation System

Figure 6.18: Difference between a fitted shaft and a measured point cloud motion compensated by a visual odometry and an inertial navigation system.



(a) Camera image while scanning at t_1 (b) Camera image while scanning at t_2

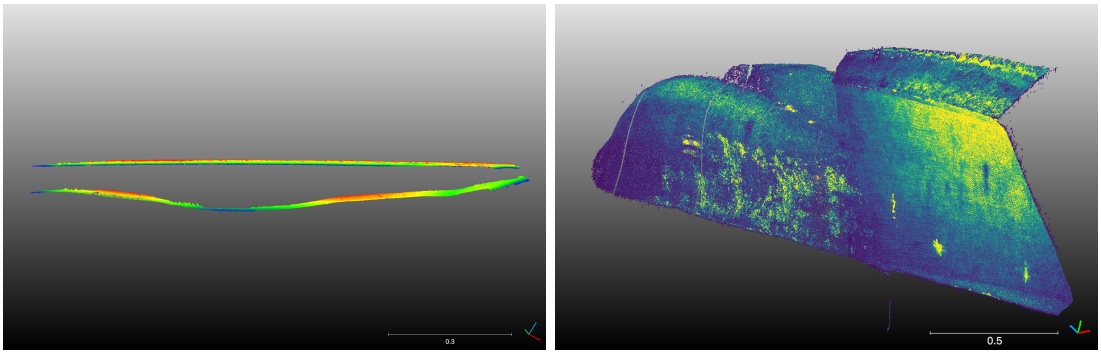
Figure 6.19: Camera image while scanning a ship hull section showing the green laser and the red light for feature tracking.



(a) Laser Remissions: left(INS); right(VO)

(b) Deviation between INS and VO

Figure 6.20: Hull section scanned with a line structured light system which is compensating with an inertial navigation system (INS) left object and with a visual odometry (VO) right object.



(a) Side view of two surfaces reconstructed with VO (top) and INS (bottom)

(b) Reconstruction of the front part of the ship using VO

Figure 6.21: Motion compensated hull section scanned with a line structured light system.

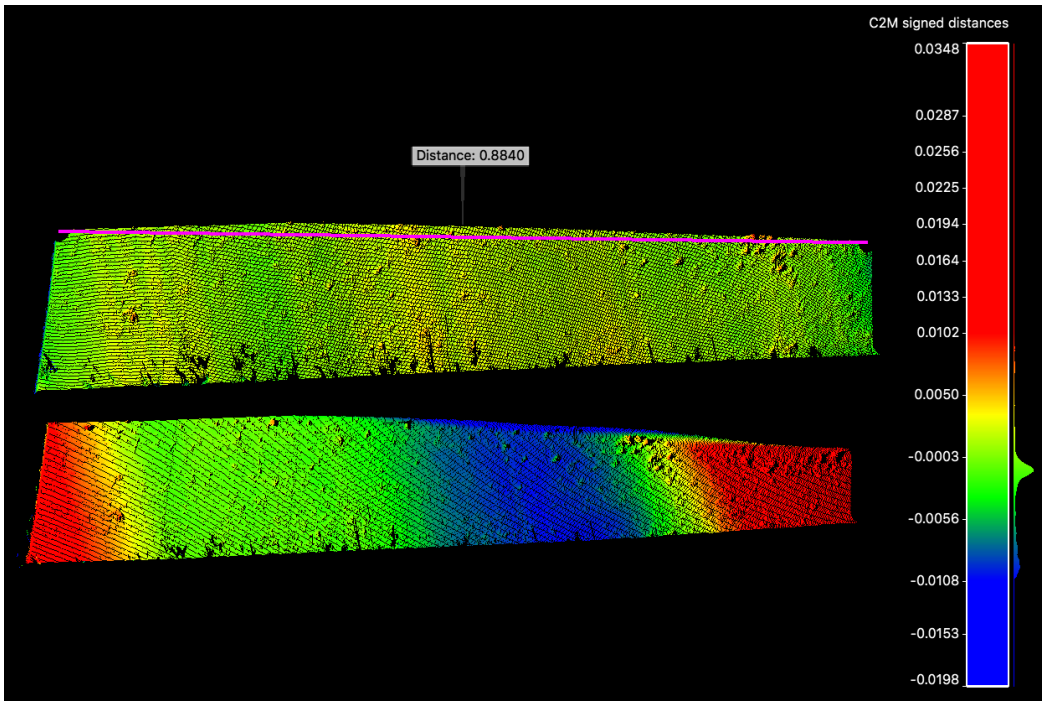


Figure 6.22: Signed distance error of a reconstructed hull section to a virtual curved reference plane assuming the real structure can be approximated by the plane. Here the upper part is reconstructed using the visual odometry for motion compensation resulting in a standard deviation of $1.4mm$. The lower part is the exact same hull patch but reconstructed with the help of the INS resulting in a standard deviation of $8.0mm$.

Conclusion

The first experiment, layed out in this chapter, shows that the proposed underwater flat-port model accurately tracks the displacement of a camera with respect to a flat physical interface. Following this, it is verified that the underlying geometric model of an underwater line-projector can be approximated by a simpler pinhole model in combination with tangential distortion. This approximation only holds if the angle between the principal ray of the laser projector and the glass interface normal is below 5°, which is usually even the case for manually aligned lasers. This combination allows implementing an accurate underwater line structured light system based on flat-port housings. Here, the unavoidable geometric distortion due to light refraction on the interfaces is effectively compensated in software for its full working range without compromising its accuracy.

In addition, several experiments demonstrate that sensor motion compensation is required for 3D scans performed from hovering platforms such as remotely operated vehicles. Here, an inertial navigation system is usually not accurate enough to compensate for the motion with an accuracy in the same order as the 3D range measurements are performed. Therefore, for improving the accuracy of such measurements, the sensor motion is tracked in the following with the help of raw sensor data utilizing background features of the scene. This so-called active-passive method is further successfully deployed to reconstruct several 3D underwater objects. Here, depending on the sensor movements, the system can reduce the overall reconstruction error by a factor of five in comparison to a motion compensation based on an inertial system. This improvement not only allows to perform accurate 3D reconstructions in the millimeter range from moving platforms. It also requires only a minimal sensor integration and removes the need for a co-calibration between the sensor and the inertial system of the vehicle. This circumstance makes it a very cost-effective method for accurately digitizing small underwater objects such as but not limited to corals, mooring chains, pipes, and ship hull sections.

Chapter 7

Conclusion and Outlook

7.1 Thesis Summary

We are only at the beginning of technological developments allowing the usage of optical sensors for more and more underwater applications. Many of these applications are currently impractical or limited by the available sensor resolutions, including inspection of submerged objects, such as mooring chains, risers, or ship hulls. In these cases, unavoidable relative motion between the sensor and the object introduces significant errors, often outside of the allowed measurement tolerances to accomplish the underlying task. Compensating these errors to a high degree usually requires motion tracking capabilities of the sensor system itself. In general, this capability relies on the correlation of multiple measurements over time by an underlying motion model. This trend is analog to what can be currently observed for in-air applications. Here, more and more off the shelf optical systems are nowadays available, supporting simultaneous location and mapping (SLAM) in 6D using embedded cameras. Even recent smartphones are now able to utilize their camera for 3D measurements, including gesture recognition and estimating the traveled sensor distance. Also, in the area of self-driving cars, cameras are becoming an essential sensor for 3D perception and situation awareness. Although the medium water profoundly impacts the performance of submerged cameras, there is no fundamental reason why they cannot be used similarly in the underwater domain.

For transferring machine vision technologies into the underwater domain, the most important aspects are the wavelength-dependent absorption of the medium water and the light refraction on air-glass and glass-water boundaries. Here, the absorption puts a hard limit on the achievable working range down to a few dozen meters. It also enforces that optical systems must use visible light to penetrate a reasonable water body. In addition, the refraction alters the light path invalidating standard camera

models used for motion estimation and 3D scene reconstruction. This circumstance leads to measurement errors, which eventually grow to the extent that it can no longer be ignored. Counteracting these errors requires either special camera ports or wet lenses, compensating for the additional refraction. This modification increases the hardware costs. It also comes with other drawbacks, such as a curved depth of focus in the case of dome ports. Whereas, flat-port cameras introduce in general significant chromatic and geometric distortions. However, the geometric distortion of flat-port cameras can be fully compensated in software with the help of flat refractive camera models. Therefore, they can still be a better choice for many underwater machine vision applications. Here, a new efficient approximation for the flat refractive forward projection is presented. It allows for the integration of flat-port cameras into bundle adjustments seamlessly and to calibrate the model parameters using standard calibration methods.

Furthermore, as stated, optical systems must use the visible spectrum of light to avoid substantial absorption by the water body. This circumstance is the main reason why many in-air systems cannot be easily transferred to the underwater domain. In general, all sensors are affected, relying on an infrared light emitter to measure the scene depth. For example, this includes all standard LIDAR systems and RGB-D cameras. Here, as the infrared spectrum is not available underwater, the depth sensor must be shifted to the visible spectrum also used by standard color and mono cameras. This overlap introduces considerable crosstalk between active and passive optical sensors. Therefore, a new passive-active approach is presented, which uses a single sensor element to perform active and passive measurements simultaneously. This approach is explicitly aware of active and passive features in the image domain, limiting the noise each modality has to cope with. Based on this, sensor self-localization can be realized while simultaneously using an active light pattern for dense 3D scene reconstructions with high accuracy.

As a demonstration, multiple difficult underwater objects are reconstructed with the new passive-active method under real-world conditions. The results are promising and are hardly achievable with a single modality alone. This also includes the reconstruction of objects in the water column subject to motion while scanning. In these cases, the motion tracking capability of the presented method can adequately compensate any relative motion between sensor and object while performing a dense 3D reconstruction in the millimeter range using structured light techniques.

7.2 Limitations and Future Work

Optical systems rely on the ability to sense electromagnetic signals with a frequency close to visible light. If an additional medium between an optical sensor and an object prevents transmission of these frequencies, the optical system is effectively disabled. Therefore, the transparency of the medium water is the leading factor for the ability to use optical systems for underwater sensing. Depending on the dissolved and none-dissolved particles in the water column, the working range of an optical system can vary between several dozen meters and only a few centimeters. This possible working range reduction by multiple orders of magnitude requires a robust fallback strategy, which is less affected by water turbidity, such as the usage of acoustical sensors. However, if the water body allows sufficient transmission for a given working range, optical sensors can outperform other sensing modalities with an order of magnitude in terms of resolution and accuracy. Here, an increase in the number of optical sensors also increases the number of redundant measurements robustifying 3D scene reconstruction and allowing to push the boundaries a system can operate in. Using only a single camera combined with a line laser can be seen as the lower hardware bound for performing metric correct 3D reconstructions. However, it is also the most sensitive as it cannot build up enough constraints to filter out biased noise. Therefore, it requires a certain amount of planning to change environmental conditions to its favor. For example, this includes operating during night time to reduce environmental light, using specialized illumination schemas for the operator, and carefully selecting sensor poses with a minimum object distance.

Future work will, therefore, concentrate on multi-sensor setups using the presented basic building blocks to increase the robustness of passive-active systems. A promising setup is, for example, a multi-camera setup allowing to directly label image features into passive and active features using the underlying trifocal tensor without requiring any motion cues. Furthermore, because the working range of underwater optical systems is in general smaller than of acoustical systems, surveys are usually carried out with acoustical sensors as primary sensors maximizing the available footprint. State of the art synthetic aperture sonar (SAS) can deliver footprints of around $800m$ orthogonal to the traveled path (across-track) at a resolution of around $3cm$. Therefore, future work will also include the fusion of optical passive-active systems with SAS imagery, which can serve as a global map reducing the drift of the optical motion compensation. Here, the underlying concept is that acoustical range data

will usually not improve a single range measurement of an optical system as the covariance of both sensor modalities is too different. However, it is often the case that SAS imagery is not delivering enough information to identify all objects of interest reliably. Therefore, one possible schema is to pre-survey a particular region with SAS and mark all unknown objects. In a second step, the marked objects are revisited, and a passive-active system is used to reconstruct them with sub-millimeter resolution. Following this, the embedded motion tracking localizes itself in the SAS imagery, which is already geo-referenced. This leads to a consist map where regions of high interest are enriched with an optical system. Any improvements in the area of visual odometry and or visual SLAM frameworks can be directly transferred to this schema as long as the refraction at interface boundaries is taken into account during pose refinement.

On a hardware level, the underwater community is usually too small to be the driving factor for significant sensor developments close to the visible spectrum. Unlike for the smartphone market, there is barely any camera manufacture that optimizes its sensor chips for the underwater case. The same is true for camera optics, LEDs, or new beam steering techniques used, for example, by solid-state LIDARs. Therefore, the underwater community heavily relies on new upcoming technologies that are transferrable to the underwater domain without requiring an extensive adaption of their core elements such as light emitters or receivers. This missing market traction limits somehow the possible developments in the near future. Nonetheless, exciting new sensor systems will emerge using the available building blocks to improve signal contrast further and to improve the utilization of the relatively small bandwidth of visible light available for underwater sensing.

Bibliography

P. Abeles. Tutorial Camera Calibration - https://boofcv.org, 2018.

S. Agarwal and K. Mierle. Ceres Solver - http://ceres-solver.org, 2018.

A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari. A theory of multi-layer flat refractive geometry. In *Computer Vision and Pattern Recognition*, pages 3346–3353, 2012. ISBN 9781467312264. doi: 10.1109/CVPR.2012.6248073.

J. Albiez, A. Duda, M. Fritsche, F. Rehrmann, and F. Kirchner. CSurvey—An autonomous optical inspection head for AUVs. *Robotics and Autonomous Systems*, 67:72–79, 2015. ISSN 09218890. doi: 10.1016/j.robot.2014.10.004.

A. Anwer, S. S. A. Ali, A. Khan, and F. Mériaudeau. Underwater 3D scanning using Kinect v2 time of flight camera. *Thirteenth International Conference on Quality Control by Artificial Vision 2017*, 10338(March):103380C, 2017. ISSN 1996756X. doi: 10.1117/12.2266834.

A. Arnaubec, J. Opderbecke, A. G. Allais, and L. Brignone. Optical mapping with the ARIANE HROV at IFREMER: The MATISSE processing tool. *MTS/IEEE OCEANS 2015 - Genova: Discovering Sustainable Ocean Energy for a New World*, 2015. doi: 10.1109/OCEANS-Genova.2015.7271713.

C. Balletti, C. Beltrame, E. Costa, F. Guerra, and P. Vernier. 3D reconstruction of marble shipwreck cargoes based on underwater multi-image photogrammetry. *Digital Applications in Archaeology and Cultural Heritage*, 3(1):1–8, 2015. ISSN 22120548. doi: 10.1016/j.daach.2015.11.003.

C. Beall, F. Dellaert, I. Mahon, and S. B. Williams. Bundle adjustment in large-scale 3D reconstructions based on underwater robotic surveys. In *OCEANS, 2011 IEEE - Spain*, pages 1–6. IEEE, 2011. doi: 10.1109/Oceans-Spain.2011.6003631.

G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa. a Comparison Between Active and Passive Techniques for Underwater 3D Applications. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-5/(March):357–363, 2012. ISSN 1682-1777. doi: 10.5194/isprsarchives-XXXVIII-5-W16-357-2011.

G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa. A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects. *Sensors (Switzerland)*, 13(8):11007–11031, 2013. ISSN 14248220. doi: 10.3390/s130811007.

M. Bleier and A. Nüchter. Low-Cost 3D Laser Scanning in Air Orwater Using Self-Calibrating Structured Light. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3:105–112, 2017. ISSN 2194-9034. doi: 10.5194/isprs-archives-XLII-2-W3-105-2017.

T. Bosch, R. Myllyla, and M. Rioux. Laser ranging: a critical review of usual techniques for distance measurement. *Optical Engineering*, 40(1):10, 2001. ISSN 0091-3286. doi: 10.1117/1.1330700.

S. S. d. C. Botelho, P. Drews, G. L. Oliveira, and M. da Silva Figueiredo. Visual odometry and mapping for Underwater Autonomous Vehicles. In *Robotics Symposium (LARS), 2009 6th Latin American*, pages 1–6. IEEE, 2009. doi: 10.1109/LARS.2009.5418320.

J. Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Technical Report 2, Intel Corporation, 2001.

G. Bradski. The OpenCV Library. *Dr Dobbs Journal of Software Tools*, 25:120–125, 2000. ISSN 1044-789X. doi: 10.1111/0023-8333.50.s1.10.

C. Bräuer-Burchardt, M. Heinze, I. Schmidt, P. Kühmstedt, and G. Notni. Underwater 3D surface measurement using fringe projection based scanning devices. *Sensors (Switzerland)*, 16(1), 2015. ISSN 14248220. doi: 10.3390/s16010013.

L. Brignone, M. Munaro, A. G. Allais, and J. Opderbecke. First sea trials of a laser aided three dimensional underwater image mosaicing technique. In *OCEANS 2011 IEEE - Spain*, pages 1–7. IEEE, jun 2011. doi: 10.1109/Oceans-Spain.2011.6003483.

D. C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8): 855–866, 1971. ISSN 03331024. doi: 10.1.1.14.6358.

F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and a.V. Razionale. Experimentation of structured light and stereo vision for underwater 3D reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):508–518, jul 2011. ISSN 09242716. doi: 10.1016/j.isprsjprs.2011.02.009.

Camera-wiki. Nikonos V, 2015.

P. L. N. Carrasco, F. Bonin-Font, M. M. Campos, and G. O. Codina. Stereo-vision graph-SLAM for robust navigation of the AUV SPARUS II. *IFAC-PapersOnLine*, 28(2):200–205, 2015. ISSN 24058963. doi: 10.1016/j.ifacol.2015.06.033.

P. L. N. Carrasco, F. Bonin-Font, and G. O. Codina. Stereo Graph-SLAM for Autonomous Underwater Vehicles. In *Advances in Intelligent Systems and Computing*, volume 302, pages 351–360. Springer, Cham, 2016. ISBN 9783319083377.

X. Chen. *Structured Light Methods: From Land to Undersea*. PhD thesis, University of Alberta, 2015.

S. Chi, Z. Xie, and W. Chen. A Laser Line auto-scanning system for underwater 3D reconstruction. *Sensors (Switzerland)*, 16(9), 2016. ISSN 14248220. doi: 10.3390/s16091534.

P. Church, W. Hou, G. Fournier, F. Dalgleish, D. Butler, S. Pari, M. Jamieson, and D. Pike. Overview of a hybrid underwater camera system. In W. W. Hou and R. A. Arnone, editors, *Proceedings of the SPIE*, volume 9111, pages 1–7, 2014. doi: 10.1117/12.2053365.

A. Concha, P. Drews-Jr, M. Campos, and J. Civera. Real-time localization and dense mapping in underwater environments from a monocular sequence. In *OCEANS 2015 - Genova*, pages 1–5, Genova, 2015. IEEE. ISBN 9781479987368. doi: 10.1109/OCEANS-Genova.2015.7271476.

J. L. Crowley and a. C. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE transactions on pattern analysis and machine intelligence*, 6(2):156–170, feb 1984. ISSN 0162-8828. doi: 10.1109/TPAMI.1984.4767500.

F. Dalgleish, B. Ouyang, A. Vuorenkoski, B. Ramos, G. Alsenas, B. Metzger, Z. Cao, and J. Principe. Undersea LiDAR imager for unobtrusive and eye safe marine wildlife detection and classification. *OCEANS 2017 - Aberdeen*, 2017-Octob:1–5, 2017. doi: 10.1109/OCEANSE.2017.8085029.

R. Detry, J. Koch, T. Pailevanian, M. Garrett, D. Levine, C. Yahnker, and M. Gildner. Turbid-water Subsea Infrastructure 3D Reconstruction with Assisted Stereo. In *2018 OCEANS - MTS/IEEE Kobe Techno-Ocean (OTO)*. IEEE, 2018. doi: 10.1109/OCEANSKOBE.2018.8559091.

V. Douskos, I. Kalisperakis, G. Karras, and E. Petsa. Fully automatic camera calibration using regular planar patterns. *Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci*, 37:21–26, 2008.

P. Drap, J. Seinturier, B. Hijazi, and D. Merad. The ROV 3D Project: Deep-Sea Underwater Survey Using Photogrammetry: Applications for Underwater Archaeology. *ACM Journal on Computing and Cultural Heritage*, 8(4):21:1–21:23, 2015a. ISSN 15564711. doi: 10.1145/2757283.

P. Drap, J. Seinturier, B. Hijazi, and D. Merad. The ROV 3D Project: Deep-Sea Underwater Survey Using Photogrammetry: Applications for Underwater Archaeology. *ACM Journal on Computing and Cultural Heritage*, 8(4):21:1–21:23, 2015b. ISSN 15564711. doi: 10.1145/2757283.

E. Dubrovinskaya, F. Dalgleish, B. Ouyang, and P. Casari. Underwater LiDAR signal processing for enhanced detection and localization of marine life. *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans, OCEANS - Kobe 2018*, 2018. doi: 10.1109/OCEANSKOBE.2018.8559113.

A. Duda and J. Albiez. Back Projection Algorithm for Line Structured Light Extraction. In *2013 OCEANS-San Diego*, pages 1–7. IEEE, 2013. ISBN 9780933957404.

A. Duda and U. Frese. Accurate Detection and Localization of Checkerboard Corners for Calibration. In *BMVC*, page 126, 2018.

A. Duda and C. Gaudig. Refractive forward projection for underwater flat port cameras. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2022–2027, oct 2016. doi: 10.1109/IROS.2016.7759318.

A. Duda, J. Schwendner, and C. Gaudig. SRSL : Monocular Self-Referenced Line Structured Light. *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2015. ISSN 21530866. doi: 10.1109/IROS.2015.7353451.

A. Duda, T. Kwasnitschka, J. Albiez, and F. Kirchner. Self-referenced laser system for optical 3D seafloor mapping. In *OCEANS 2016 - Monterey*, pages 1–6, 2016. ISBN 9781509015375. doi: 10.1109/OCEANS.2016.7761203.

S. Q. Duntley. Light in the Sea. *J. Opt. Soc. Am.*, 53(2):214–233, feb 1963. doi: 10.1364/JOSA.53.000214.

D. Eberly. Least Squares Fitting of Data by Linear or Quadratic Structures. In *Geometric Tools*, pages 1–62, 2018.

H. Fan, L. Qi, Y. Ju, J. Dong, and H. Yu. Refractive laser triangulation and photometric stereo in underwater environment. *Optical Engineering*, 56(11):1, 2017. ISSN 0091-3286. doi: 10.1117/1.OE.56.11.113101.

M. Ferrera, J. Moras, P. Trouvé-Peloux, and V. Creuze. Real-time Monocular Visual Odometry for Turbid and Dynamic Underwater Environments. *Sensors 2019*, pages 1–19, 2018. ISSN 14248220. doi: 10.3390/s19030687.

M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981. doi: 10.1145/358669.358692.

R. Fisher and D. Naidu. *A comparison of algorithms for subpixel peak detection.* Springer, Berlin, Heidelberg, 1996. ISBN 978-3-642-58288-2.

M. Forstner and E. Gulch. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. In *Proceedings of the ISPRS Intercommission Conference on Fast Processing of Phonogrammic Data*, pages 281–305, 1987.

X. Gao and X. Hou. Complete Solution Classification for the Perspective-Three-Point Problem. *IEEE transactions on pattern analysis and machine intelligence*, pages 930–943, 2003. doi: 10.1109/TPAMI.2003.1217599.

A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. doi: doi.org/10.1177/0278364913491297.

M. V. Ginkel, C. L. L. Hendriks, L. J. V. Vliet, C. L. Luengo Hendriks, and L. J. V. Vliet. A short introduction to the Radon and Hough transforms and how they relate to each other. Technical report, Faculty of Applied Science Delft University of Technology, 2004.

A. Handa, R. A. Newcombe, A. Angeli, and A. J. Davison. Real-time camera tracking: When is high frame-rate best? In *Computer Vision – ECCV 2012.*, volume 7578, pages 1–14. Springer, Berlin, Heidelberg, 2012. ISBN 9783642337857.

S. Haner and K. Astrom. Absolute Pose for Cameras Under Flat Refractive Interfaces. In *Computer Vision and Pattern Recognition*, pages 1428–1436, 2015. ISBN 9781467369640.

S. Hannon. Underwater Mapping. *LiDAR Magazine - Spatial Media*, 3(1):1–4, 2013.

C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Procedings of the Alvey Vision Conference 1988*, pages 23.1–23.6, 1988. doi: 10.5244/C.2.23.

R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, edition 2 edition, 2004. ISBN 9780521540513.

M. Heredia Conde. *Compressive sensing for the photonic mixer device: Fundamentals, methods and results*. Springer Vieweg, 2017. ISBN 9783658180577. doi: 10.1007/978-3-658-18057-7.

M. Hildebrandt. *Development, evaluation and validation of a stereo camera underwater SLAM Algorithm*. Dissertation, Universitat Bremen, 2014.

G. Huo, Z. Wu, J. Li, and S. Li. Underwater target detection and 3D reconstruction system based on binocular vision. *Sensors (Switzerland)*, 18(10), 2018. ISSN 14248220. doi: 10.3390/s18103570.

G. Inglis, C. Smart, I. Vaughn, and C. Roman. A pipeline for structured light bathymetric mapping. In *Intelligent Robots and Systems (IROS)*, pages 4425–4432, 2012. ISBN 9781467317368.

J. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, 1990. doi: 10.1109/48.50695.

J. Jaffe, K. Moore, J. McLean, and M. Strand. Underwater Optical Imaging: Status and Prospects. *Oceanography*, 14(3):64–75, 2001. ISSN 10428275. doi: 10.5670/oceanog.2001.24.

J. S. Jaffe. Underwater optical imaging: the design of optimal systems. *Oceanography*, 1(2):40–41, 1988. ISSN 10428275. doi: 10.5670/oceanog.1988.09.

J. S. Jaffe. Enhanced extended range underwater imaging via structured illumination. *Optics express*, 18(12):12328–12340, 2010. ISSN 1094-4087. doi: 10.1364/OE.18.012328.

J. S. Jaffe. Underwater Optical Imaging: The Past, the Present, and the Prospects. *IEEE Journal of Oceanic Engineering*, 40(3):683–700, 2015. ISSN 03649059. doi: 10.1109/JOE.2014.2350751.

S. Jain. A survey of Laser Range Finding. Technical report, 2003.

A. Jordt. *Underwater 3D Reconstruction Based on Physical Models for Refraction and Underwater Light Propagation*. Dissertation, Kiel University, 2013.

A. Jordt, K. Köser, and R. Koch. Refractive 3D reconstruction on underwater images. *Methods in Oceanography*, pages 1–24, 2015. ISSN 22111220. doi: 10.1016/j.mio.2016.03.001.

A. Jordt-Sedlazeck and R. Koch. Refractive Structure-from-Motion on Underwater Images. In *IEEE International Conference on Computer Vision (ICCV)*, pages 57–64. IEEE, 2013. doi: 10.1109/ICCV.2013.14.

A. Jordt-Sedlazeck, D. Jung, and R. Koch. Refractive plane sweep for underwater images. In J. Weickert, M. Hein, and B. Schiele, editors, *Pattern Recognition*, pages 333–342, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

H. Kan, C. Katagiri, Y. Nakanishi, S. Yoshizaki, M. Nagao, and R. Ono. Assessment and Significance of a World War II battle site: recording the USS Emmons using a High-Resolution DEM combining Multibeam Bathymetry and SfM Photogrammetry. *International Journal of Nautical Archaeology*, 47(2):267–280, 2018. ISSN 10959270. doi: 10.1111/1095-9270.12301.

L. Kang, L. Wu, and Y.-H. Yang. Two-View Underwater Structure and Motion for Cameras under Flat Refractive Interfaces. In *Computer Vision – ECCV 2012*, volume 7575, pages 303–316, 2012. ISBN 978-3-642-33764-2.

J. Kannala and S. S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, 2006. ISSN 01628828. doi: 10.1109/TPAMI.2006.153.

Kebes. Liquid water absorption spectrum across a wide wavelength range — Wikipedia, The Free Encyclopedia, 2008.

K. Kiger. Introduction of Particle Image Velocimetry. Technical report, University of Maryland, 2010.

J. Klepsvik, H. Torsen, and K. Thoresen. Laser imaging technology for subsea inspection: Principles and applications. In *IRM incorporating ROV 90*, page 16, 1990.

L. Kneip and P. Furgale. OpenGV: A unified and generalized approach to real-time calibrated geometric vision. *2014 IEEE International Conference on Robotics and Automation*, pages 1–8, 2014. doi: 10.1109/ICRA.2014.6906582.

T. Kovacovsky. Parameters of 3D sensing techniques in a nutshell, 2017.

M.-Y. Kuo and S. Nobuhara. One-Shot Underwater Active Stereo Through Refractive Parallel Flat Surfaces. In *PoTS Imaging Symposium*, volume 2, pages 1–8, 2017.

F. S. H. Leonard, R. M. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. J. Advanced perception , navigation and planning for autonomous in-water ship hull inspection. *The International Journal of Robotics Research*, 31(12):1445–1464, 2012. doi: 10.1177/0278364912461059.

L. Li. Time-of-Flight Camera–An Introduction. Technical report, Texas Instruments, 2014.

Y. Li, Y. Zhang, H. Li, W. Zhang, and Q. Zhang. Epipolar geometry and stereo matching algorithm for underwater fish-eye images. *International Journal of Advanced Robotic Systems*, 15(2):172988141876471, 2018. ISSN 1729-8814. doi: 10.1177/1729881418764715.

T. Lindeberg. *Discrete Scale-space Theory and Scale-space Primal Sketch*. PhD thesis, Royal Institute of Technology,Sweden, 1991.

J. Liu, A. Jakas, A. Al-Obaidi, and Y. Liu. Practical issues and development of underwater 3D laser scanners. In *Emerging Technologies and Factory Automation (ETFA)*, pages 1–8, Bibao, 2010. IEEE. doi: 10.1109/ETFA.2010.5641223.

J. V. D. Lucht, M. Bleier, F. Leutert, and K. Schilling. STRUCTURED-LIGHT BASED 3D LASER SCANNING OF SEMI-SUBMERGED. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV(June):4–7, 2018. ISSN 21949050. doi: 10.5194/isprs-annals-IV-2-287-2018.

T. Łuczyński, M. Pfingsthorn, and A. Birk. The Pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings. *Ocean Engineering*, 133:9–22, 2017. ISSN 00298018. doi: 10.1016/j.oceaneng.2017.01.029.

T. Luhmann. *Nahbereichsphotogrammetrie: Grundlagen, Methoden und Anwendungen*. Wichmann, 2010. ISBN 9783879074792.

T. Luhmann, S. Robson, S. Kyle, and J. Boehm. Close-Range Photogrammetry and 3D Imaging. *Close-Range Photogrammetry and 3D Imaging*, 2013. ISSN 00991112. doi: 10.1515/9783110302783.

M. R. Maire. *Contour detection and image segmentation*. PhD thesis, University of California, Berkeley, 2009.

P. Mariani, I. Quincoces, K. H. Haugholt, Y. Chardard, A. W. Visser, C. Yates, G. Piccinno, G. Reali, P. Risholm, and J. T. Thielemann. Range-gated imaging system for underwater monitoring in ocean environment. *Sustainability (Switzerland)*, 11(1), 2018. ISSN 20711050. doi: 10.3390/su11010162.

M. Massot-Campos, G. Oliver-Codina, H. Kemal, Y. Petillot, and F. Bonin-Font. Structured light and stereo vision for underwater 3D reconstruction. In *OCEANS 2015 - Genova*, pages 1–6, 2015. doi: 10.1109/OCEANS-Genova.2015.7271433.

B. L. McGlamery. A computer model for underwater camera systems. In *Proceedings of the SPIE*, volume 208, pages 221–231, 1980. doi: 10.1117/12.958279.

D. McLeod, J. Jacobson, M. Hardy, and C. Embry. Autonomous Inspection using an Underwater 3D LiDAR. In IEEE, editor, *OCEANS - San Diego*, pages 1–8, 2013. ISBN 9780933957404. doi: 10.23919/OCEANS.2013.6741175.

M. Meireles, R. Lourenço, A. Dias, J. M. Almeida, H. Silva, and A. Martins. Real time visual SLAM for underwater robotic inspection. *2014 Oceans - St. John's, OCEANS 2014*, pages 2–6, 2015. doi: 10.1109/OCEANS.2014.7003097.

F. Menna, E. Nocerino, and F. Remondino. Flat versus hemispherical dome ports in underwater photogrammetry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(2W3):481–487, 2017. ISSN 16821750. doi: 10.5194/isprs-archives-XLII-2-W3-481-2017.

C. Mertz and S. Koppal. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. In *Computer Vision and Pattern Recognition Workshops*, pages 15–22. Ieee, jun 2012. ISBN 978-1-4673-1612-5. doi: 10.1109/CVPRW.2012.6239194.

B. Micusik and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. *Computer Vision and Pattern Recognition*, 1:I–485, 2003. ISSN 1063-6919. doi: 10.1109/CVPR.2003.1211393.

C. D. Mobley. Optical Properties of Water. In *Light and waters: Radiative Transfer in Natural Waters*, pages 60–144. Academic Press, 1994. ISBN 978-0125027502.

K. Moore, J. Jaffe, and B. Ochoa. Development of a New Underwater Bathymetric Laser Imaging System: L-Bath. *Journal of Atmospheric and Technology*, 17(8): 1106–1117, 2000. doi: 10.1175/1520-0426(2000)017¡1106:DOANUB¿2.0.CO;2.

S. G. Narasimhan, S. Nayar, B. Sun, and S. J. Koppal. Structured light in scattering media. *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 1:420–427, 2005. doi: 10.1145/1508044.1508106.

M. M. Nawaf, D. Merad, J. P. Royer, J. M. Boï, M. Saccone, M. B. Ellefi, and P. Drap. Fast visual odometry for a low-cost underwater embedded stereo system†. *Sensors (Switzerland)*, 18(7), 2018. ISSN 14248220. doi: 10.3390/s18072313.

D. Nistér. An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence*, 26(6):756–77, jun 2004. ISSN 0162-8828. doi: 10.1109/TPAMI.2004.17.

D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *Computer Vision and Pattern Recognition*, pages 652–659. IEEE, 2004.

E. Nocerino, F. Menna, F. Fassi, and F. Remondino. Underwater calibration of dome port pressure housings. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40(3W4):127–134, 2016. ISSN 16821750. doi: 10.5194/isprsarchives-XL-3-W4-127-2016.

F. Oleari, F. Kallasi, D. L. Rizzini, J. Aleotti, and S. Caselli. An underwater stereo vision system: From design to deployment and dataset acquisition. In *OCEANS 2015 - Genova*. IEEE, 2015. ISBN 9781479987368. doi: 10.1109/OCEANS-Genova.2015.7271529.

B. Ouyang and F. Dalgleish. Experimental study of underwater stereo via pattern projection. In *Oceans, 2012*, pages 1–7. Ieee, oct 2012. ISBN 978-1-4673-0831-1. doi: 10.1109/OCEANS.2012.6404976.

M. Pfingsthorn, R. Rathnam, T. Luczynski, and A. Birk. Full 3D Navigation Correction using Low Frequency Visual Tracking with a Stereo Camera. In *OCEANS 2016 - Shanghai*, pages 1–6. IEEE, 2016. doi: 10.1109/OCEANSAP.2016.7485520.

F. Remondino and D. Stoppa. *TOF Range-Imaging Cameras*. Springer Berlin, 2013. ISBN 9783642275234. doi: 10.1007/978-3-642-27523-4.

C. Roman, G. Inglis, and J. Rutter. Application of structured light imaging for high resolution mapping of underwater archaeological sites. *Oceans'10 Ieee Sydney*, pages 1–9, 2010. doi: 10.1109/OCEANSSYD.2010.5603672.

M. Rossi, P. Trslić, S. Sivčev, J. Riordan, D. Toal, and G. Dooly. Real-Time Underwater StereoFusion. *Sensors (Basel, Switzerland)*, 18(11):1–17, 2018. ISSN 14248220. doi: 10.3390/s18113936.

M. Rufli, D. Scaramuzza, and R. Siegwart. Automatic detection of checkerboards on blurred and distorted images. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 3121–3126, 2008. ISBN 9781424420582. doi: 10.1109/IROS.2008.4650703.

H. Sarbolandi, D. Lefloch, and A. Kolb. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer Vision and Image Understanding*, 139:1–20, 2015. ISSN 1090235X. doi: 10.1016/j.cviu.2015.05.006.

R. Schattschneider. *Accurate high-resolution 3D surface reconstruction and localisation using a wide-angle flat port underwater stereo camera - towards autonomous ship hull inspection*. PhD thesis, University of Canterbury, Christchurch, New Zealand, 2014.

R. Schattschneider, G. Maurino, and W. Wang. Towards stereo vision SLAM based pose estimation for ship hull inspection. In *OCEANS'11 MTS*, pages 1–8. IEEE, 2011. doi: 10.23919/OCEANS.2011.6106988.

R. Schettini and S. Corchs. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP J. Adv. Signal Process*, 2010: 1–14, 2010. doi: 10.1155/2010/746052.

J. L. Schonberger and J.-M. Frahm. Structure-from-Motion Revisited. In *Computer Vision and Pattern Recognition (CVPR),*, pages 4104–4113. IEEE, 2016. doi: 10.1109/cvpr.2016.445.

D. Shea, P. Crocker, J. Dillon, and S. Chapman. AquaPix – A Low-Cost Interferometric Synthetic Aperture Sonar for AUVs : Sea Trials and Results. In *UUST*, 2013.

F. Shen, F. Cao, H. Qi, Y. Huang, W. Jin, G. Liu, and X. Wang. Range-gated underwater laser imaging system based on intensified gate imaging technology. *International Symposium on Photoelectronic Detection and Imaging 2007: Photoelectronic Imaging and Detection*, 6621(March 2008):66210L, 2008. doi: 10.1117/12.790665.

E. D. Sinzinger. A model-based approach to junction detection using radial energy. *Pattern Recognition*, 41(2):494–505, feb 2008. ISSN 00313203. doi: 10.1016/j.patcog.2007.06.032.

C. J. Smart, C. Roman, and S. N. Carey. Detection of diffuse seafloor venting using structured light imaging. *Geochemistry, Geophysics, Geosystems*, 14(11):4743–4757, nov 2013. ISSN 15252027. doi: 10.1002/ggge.20280.

R. C. Smith and K. S. Baker. Optical properties of the clearest natural waters (200–800 nm). *Applied Optics*, 20(2):177, 1981. ISSN 0003-6935. doi: 10.1364/AO.20.000177.

J. Solà, A. Monin, and M. Devy. BiCamSLAM: Two times mono is more than stereo. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 4795–4800, 2007. ISSN 10504729. doi: 10.1109/ROBOT.2007.364218.

K. Strobl and E. Mair. The self-referenced DLR 3D-modeler. In IEEE, editor, *International Conference on Intelligent Robots and Systems*, pages 21–28, 2009. doi: 10.1109/IROS.2009.5354708.

A. Suess. *High Performance CMOS Range Imaging*. CRC Press, 1 edition, 2016. ISBN 9778-1138029125.

Z. Tang, R. G. V. Gioi, P. Monasse, and J.-m. Morel. A Precision Analysis of Camera Distortion. In *IEEE Transactions on Image Processing*, pages 2694 – 2704. IEEE, 2017. doi: 10.1109/TIP.2017.2686001.

P. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15(8):591–605, 2002. ISSN 02628856. doi: 10.1016/s0262-8856(97)00010-3.

T. Treibitz, Y. Y. Schechner, C. Kunz, and H. Singh. Flat refractive geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):51–65, 2012. ISSN 01628828. doi: 10.1109/TPAMI.2011.105.

B. Ummenhofer and T. Brox. Dense 3D Reconstruction with a Hand-Held Camera. In *Pattern Recognition (Proc. DAGM)*, pages 1–10. Springer, 2012.

I. Vornicu, R. Carmona-Galán, and Á. Rodríguez-Vázquez. A CMOS imager for time-of-flight and photon counting based on single photon avalanche diodes and in-pixel time-to-digital converters. *Romanian Journal of Information Science and Technology*, 17(4):353–371, 2014. ISSN 14538245.

K. Yamafune, R. Torres, and F. Castro. Multi-Image Photogrammetry to Record and Reconstruct Underwater Shipwreck Sites. *Journal of Archaeological Method and Theory*, 24(3):703–725, 2017. ISSN 15737764. doi: 10.1007/s10816-016-9283-1.

Q. Yang, K.-h. Tan, B. Culbertson, J. Apostolopoulos, Q. Yang, and K.-h. Tan. Fusion of Active and Passive Sensors for Fast 3D Capture. In *IEEE International Workshop on Multimedia Signal Processing*, pages 69–74, Saint-Malo, France., 2010. doi: 10.1109/MMSP.2010.5661996.

T. Yau, M. Gong, and Y. H. Yang. Underwater camera calibration using wavelength triangulation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2499–2506, 2013. ISSN 10636919. doi: 10.1109/CVPR.2013.323.

W. Zeng and Z. Zhang. Microsoft Kinect Sensor and Its Effect. In *Multimedia - IEEEMM*, pages 4–10. IEEE, 2012. doi: 10.1109/MMUL.2012.24.

Z. Zhang. A flexible new technique for camera calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22, pages 1330–1334, 2000. doi: 10.1109/34.888718.