Perspective and Non-Perspective Camera Models in Underwater Imaging - Overview and Error Analysis

Anne Sedlazeck * and Reinhard Koch

Multimedia Information Processing, Institute of Computer Science,
Christian-Albrechts-University (CAU) of Kiel, Germany
{sedlazeck,rk}@mip.informatik.uni-kiel.de
http://www.mip.informatik.uni-kiel.de

Abstract. When capturing images underwater, image formation is affected in two major ways. First, the light rays traveling underwater are absorbed and scattered depending on their wavelength, creating effects on the image colors. Secondly, the glass interface between air and water refracts the ray entering the camera housing because of a different index of refraction of water, hence the ray is also affected in a geometrical way. This paper examines different camera models and their capabilities to deal with geometrical effects caused by refraction. Using imprecise camera models leads to systematic errors when computing 3D reconstructions or otherwise exploiting geometrical properties of images. In the literature, many authors have published work on underwater imaging by using the perspective pinhole camera model (single viewpoint model - SVP) with a different effective focal length and distortion to compensate for the error induced by refraction at the camera housing. On the other hand, methods were proposed, where refraction is modeled explicitly or where generic, non-single-view-point camera models are used. In addition to discussing all three model categories, an accuracy analysis of using the perspective model on underwater images is given and shows that the perspective model leads to systematic errors that compromise measurement accuracy.

1 Introduction

Underwater imaging is becoming more and more popular as technology becomes available to research the ocean floor at great water depths. Exemplary applications are the measurement of fish sizes or other organisms - in general observations of different ecosystems, (volumetric) measurements of deep sea structures like hydrothermal vents, offshore oil production, construction and maintenance of offshore wind parks, cable and pipe inspection, underwater archeology (e.g. ship wreck inspection, cave diving), and ship hull inspection as a measure against terrorism.

* This work was supported by the German Research Foundation (DFG), KO-2044/6-1 3D Modeling of Seafloor Structures from ROV-based Video Sequences

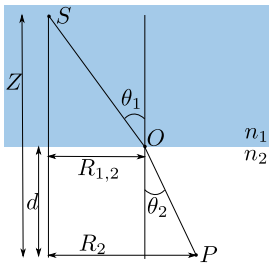

Fig. 1. Fermat's principle based on the ray from S to P being refracted at O.

In contrast to conventional computer vision, underwater image formation is effected in two ways. First, while traveling through the water, the light rays are partly absorbed and scattered, dependent on the wavelength. This leads to a green or blue hue on underwater images and has thus an effect on the colors. Secondly, refraction of light occurs at the boundary to the underwater housing, since the inside is usually occupied by air. Refraction effects the geometry of the image formation and is the subject of this work.

1.1 Refraction at Underwater Housings

The definition of refraction, as in [20], is the deviation of a light ray from its former path when entering a medium with a new optical density. While the frequency is constant, this causes the propagation velocity to change and all rays, not traveling perpendicularly to the interface, change their direction and enter the new medium under a different angle compared to the interface's normal. This effect is explained by Fermat's principle: the light traveling through two different media always travels the way that takes the least time to traverse. A derivation using the distances traveled and the speed of light in the different media yields Snell's law.

Following figure 1, the time the ray needs to travel from S to P is described by the following sum:

$$t = \frac{\sqrt{(Z-d)^2 + R_{1,2}^2}}{\nu_1} + \frac{\sqrt{d^2 + (R_2 - R_{1,2})^2}}{\nu_2}, \tag{1}$$

where ν_1 and ν_2 denote the speed of light in the corresponding medium. In order to minimize this equation, its derivative is computed:

$$\frac{\partial t}{\partial R_{1,2}} = \frac{R_{1,2}}{\nu_1 \sqrt{(Z-d)^2 + R_{1,2}^2}} + \frac{-(R_2 - R_{1,2})}{\nu_2 \sqrt{d^2 + (R_2 - R_{1,2})^2}} = 0, \tag{2}$$

which can also be expressed by:

$$\frac{\sin\theta_1}{\nu_1} = \frac{\sin\theta_2}{\nu_2}, \tag{3}$$

Medium	Index of Refraction
air ($\lambda = 589nm$)	1.0003
pure water ($\lambda = 700nm$, $30°C$, $p = 1.01e10^5 Pa$)	1.329
pure water ($\lambda = 700nm$, $30°C$, $p = 1.08e10^8 Pa$)	1.343
sea water ($\lambda = 700nm$, $30°C$, $p = 1.01e10^5 Pa$)	1.335
sea water ($\lambda = 400nm$, $30°C$, $p = 1.08e10^8 Pa$)	1.363
quartz glass ($\lambda = 589nm$)	1.4584
acrylic glass (Plexiglas, $\lambda = 589nm$)	1.51
crown glass ($\lambda = 589nm$)	1.52
light flint glass ($\lambda = 589nm$)	1.58
dense flint glass ($\lambda = 589nm$)	1.66
Lanthan flint glass ($\lambda = 589nm$)	1.80

Table 1. Indexes of refraction for air, different kinds of water, and glass as in [20] p. 163 and [36] p. 85.

and with c being the speed of light in vacuum and $n_1 = c/\nu_1$ and $n_2 = c/\nu_2$, Snell's law follows:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1}. \tag{4}$$

n_1 and n_2 are called indexes of refraction describing the phenomenon for both media. When setting the index of refraction to 1.0 for vacuum, all other indexes are determined relative to it. Important for this work are the indexes of refraction for water, glass, and air. The index of air is close to the index of vacuum, and is therefore usually set to 1.0. The index of water changes due to wavelength, salinity, pressure, and temperature, causing slight changes of the index of refraction when comparing different water bodies in the ocean (see table 1). According to [36], the dependency on all four parameters only induces a change of about 3% in the index of refraction in the whole relevant parameter range for ocean optics, thus the change can be ignored. In contrast to that, [20] lists the indexes of refraction for glass (see table 1) and shows a far stronger variation depending on the different materials, usually requiring them to be considered explicitly.

When using cameras to capture underwater images, those cameras need to be put into watertight underwater housings. These underwater housings have a piece of glass through which the image is taken, while the inside of the housing is filled with air. Hence, refraction, as described above, happens twice: first, at the water-glass interface and, secondly, at the glass-air interface (fig. 2, left), causing the ray to shift due to the double refraction depending on the glass thickness.

When working with camera housings, two different kinds of glass ports need to be considered. Planar glass, effecting most of the rays to be refracted just as depicted in figure 2 on the left, and dome ports (fig. 2, right), eliminating the refractive effect to some extend. In theory, the dome port completely removes refraction, due to zero angles between the interface normal and incoming rays.

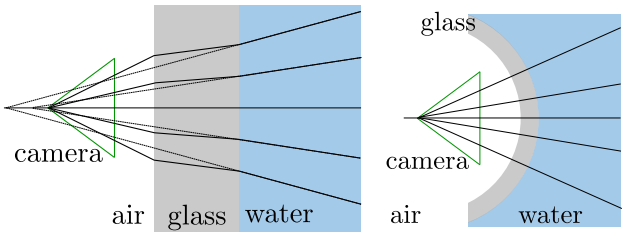

Fig. 2. Left: refraction at flat glass interface. Right: straight rays entering the underwater housing through a dome port.

However, the port and housing need to be manufactured and assembled such that the camera is centered perfectly with respect to the dome port's center for this to work. In case of a flat port or an imperfectly fit dome port, refraction of light rays invalidates the single-view-point camera model. This can be observed in figure 2 on the left for a flat port: the rays traveling in the water in the camera's direction are traced towards the optical center without refraction (dashed lines) and they do not intersect the optical axis in one common center of projection. Hence, the camera does not have a single view point (non-SVP camera model) and the commonly used pinhole camera model is invalid for underwater images.

In the literature, a large group of authors uses the perspective model, although their camera housings have flat ports, while others seek a complete physical model of the refraction effects to achieve greater accuracy. A third approach consists of using a more generic camera model, not requiring a single view point, only being based on rays. The goal of this work is to examine the wealth of approaches to underwater imaging and to discuss their benefits and shortcomings. We will show that the SVP assumption is not sufficient and will discuss a camera model that eliminates these shortcomings.

Sections 2 - 4 will analyze in depth the state of the art in underwater camera models and will give an overview of the publications on the above mentioned categories. A concise summary of all papers and their application area is given in tables 5-7 in an overview covering perspective models (28 papers), ray-based models (6 papers), and physical, refractive models (19 papers). In section 5, an error analysis of the usage of the perspective and the physical imaging model on underwater images is presented, followed by a conclusion.

2 The Perspective Camera Model

Throughout the article, geometric entities are described in a common notation, summarized in table 2. In addition, the conversion of Euclidean coordinates into cylinder coordinates is required:

$$\begin{pmatrix} R \\ \varphi \\ Z \end{pmatrix} = \begin{pmatrix} \sqrt{X^2+Y^2} \\ \arccos(\frac{X}{R}) \\ Z \end{pmatrix}. \tag{5}$$

Homogeneous Point in 3D	$\mathbf{X} = (X, Y, Z, 1)^T$
Homogeneous Point in 2D	$\mathbf{x} = (x, y, 1)^T$
Euclidean Vector in 3D	$\mathsf{X} = (X, Y, Z)^T$
Euclidean Vector in 2D	$\mathsf{x} = (x, y)^T$
Ray in 3D	$\tilde{\mathsf{X}} = (\tilde{X}, \tilde{Y}, \tilde{Z})^T$
3D vector in cylinder coordinates	$\mathsf{X}^c = (R, \varphi, Z)^T$
Ray in cylinder coordinates	$\tilde{\mathsf{X}}^c = (\tilde{R}, \tilde{\varphi}, \tilde{Z})^T$
distance camera center - interface in mm	d
glass thickness in mm	d_g
indexes of refraction (air, glass, water)	n_a, n_g, n_w

Table 2. Notations for rays and points in Euclidean, homogeneous, and cylinder coordinates. Note that in some cases, it is sufficient to use the radial coordinates $(R, Z)^T$, thus φ is omitted for the sake of readability.

The pinhole camera model with distortion is one of the established models for perspective cameras. It uses rays to describe how 3D points are projected to individual pixels and is parametrized by intrinsic parameters describing the camera's internal properties:

$$\mathbf{K} = \begin{pmatrix} f & s & c_x \\ 0 & af & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

with f being the focal length, a being the aspect ratio, s being the skew, and (c_x, c_y) being the principal point. Extrinsic parameters describe the camera pose, thus, the projection matrix follows with $\mathbf{R}$ being an orthonormal rotation matrix and C being a translation vector: $P = \mathbf{K}\mathbf{R}^T[\mathbf{I}| - \mathsf{C}]$. A homogeneous point $\mathbf{X}$ in 3D space is projected by the camera, resulting in a homogeneous 2D point $\mathbf{x} = \mathbf{P}\mathbf{X}$. In addition, it is possible to use this parametrization to back project 2D image points, i.e. to compute the ray in space on which the 3D point lies [18]. Imperfect lenses require an additional compensation for lens distortion [35], which is usually divided into a radial component and a de-centering or tangential component, approximated by a polynomial. Let (x, y) be a 2D image point without distortion. With $r = \sqrt{x^2 + y^2}$, the distorted point (x_d, y_d) is then retrieved by:

$$\begin{aligned} \begin{pmatrix} x_d \\ y_d \end{pmatrix} &= \begin{pmatrix} x + (x - c_x)[r_1 r^2 + r_2 r^4 + ...] + x_{tan} \\ y + (y - c_y)[r_1 r^2 + r_2 r^4 + ...] + y_{tan} \end{pmatrix} \\ x_{tan} &= [t_1(r^2 + 2(x - c_x)^2) + 2t_2(x - c_x)(y - c_y)](1 + t_3 r^2 + ...) \\ y_{tan} &= [2t_1(x - c_x)(y - c_y) + t_2(r^2 + 2(y - c_y)^2)](1 + t_3 r^2 + ...) \end{aligned} \tag{7}$$

where $r_1, r_2, ...$ and $t_1, t_2, ...$ are the radial and tangential distortion coefficients respectively. In the literature, there is no consensus about the number of coefficients that are necessary for perspective cameras with distortion. For example

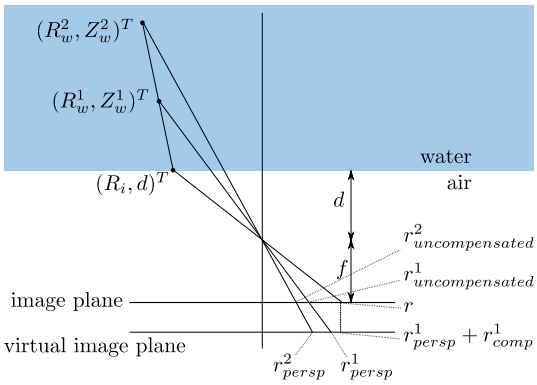

Fig. 3. Approximation of the underwater camera by the perspective model. A virtual image plane is used in combination with larger radial distortion to image the point onto the same radial coordinate. Even though, the two 3D points lying on the same ray in water are projected to the same pixel using the underwater model, but onto different pixels using the perspective model.

[21] uses two parameters each, while [61] uses only one coefficient for radial distortion and none for tangential distortion. Two coefficients for radial distortion and none for tangential distortion are used by Zhang in [66]. A description of a widely used toolbox for perspective camera calibration is found in [2]. For our own experiments we use [52] with two coefficients for both components.

When using the perspective model on underwater images captured through a glass port, a calibration based on above-water images is invalid underwater. Furthermore, the perspective model itself is invalid for underwater images due to the non-single view point. Despite that, focal length and distortion coefficients can be used to approximate the difference introduced by not modeling refraction explicitly. Figure 3 depicts this approximation using cylinder coordinates: (R_w^1, Z_w^1) is a 3D point in water, which would be imaged to $r^1_{uncompensated}$ without any compensation causing a large error compared to the true image r. By using a virtual image plane, which is moved further away from the center of projection, a part of this error can be compensated (r^1_{persp}). Stronger radial distortion r^1_{comp} can be used to eliminate the error ($r = r^1_{persp} + r^1_{comp}$). The second 3D point (R_w^2, Z_w^2) is imaged with a greater point-camera distance on the same ray in water and it immediately becomes obvious, that the required compensation by radial distortion r^2_{comp} differs from r^1_{comp} and is therefore depending on the camera-point distance, a feature not supported by the common pinhole camera model. Hence, the approximation can only be satisfying for the calibration distance.

In spite of these problems, such an approximation offers the possibility of calibrating a camera above water and compute its approximate calibration for the underwater scenario, which is examined in the following two presented methods. Freyer et al. [12] use the pinhole camera model (with 3 parameters for radial

distortion and 2 for tangential distortion) and compensate for refraction by multiplying the focal length in water by 1.34. More important in their opinion, is the change in the distortion parameters. When submerging a camera in water, the change in radial distortion is specified to be $\delta r = \left(\frac{\cos\theta_w}{\cos\theta_a} - 1\right) r$, with r being the radial distortion in air, θ_w being the angle between optical axis and water ray, and θ_a being the angle between optical axis and air ray. In common applications for perspective cameras, those angles are usually unknown.

Lavest et al. published a similar work in [30]. The paper explicitly models a thick lens, directly emerged in a medium other than air, which is then transferred into the pinhole model with distortion. Concerning the focal lengths in air and water, the major result matches the one introduced above:

$$1.333 f_{water} = f_{air}. \tag{8}$$

The computation of underwater distortion differs to the one in [12]: if $r_{d_{air}}$ and $r_{d_{water}}$ are the distorted coordinates in air and water respectively and $r_{rad_{air}}$ and $r_{rad_{water}}$ the corresponding radial distortion corrections, then

$$1.333(r_{d_{air}} - r_{rad_{water}}) = r_{d_{water}} - r_{rad_{water}}. \tag{9}$$

The authors experimented with two different cameras and their calibrations in air and water and found the theoretical equations (8) and (9) to be a good approximation. When considering the above discussion of figure 3, it becomes clear that unless $r_{rad_{water}}$ is depending on the imaging distance, the model error is still not eliminated completely.

In case of using a dome port with a perfect fit, meaning the sphere's center coincides with the camera's center of projection, a calibration done in air is valid below water. According to the entry pupil model used for lens systems [1], the locus of the center of projection is determined by the lens system of the camera and can even lie in front of the physical camera and its lens. Consequently, it is a difficult task to perfectly align the camera center and the dome port's center. Alignment errors lead to even more complicated aberrations from the pinhole model than in the flat port case.

Despite of this usually inevitable geometric error, the literature contains a lot of methods (refer to tables 5-7), where the perspective model is used in underwater scenarios. Examples for calibrating a camera underwater are found in [4] or by Pessel et al. in [45, 43, 44]. Application areas utilize the implicitly contained geometric properties of the images to measure distances in stereo images [19, 8], to compute dense stereo [51, 39], to aid navigation by computing mosaics [15, 16, 13, 5, 63, 9, 60, 42], or to reconstruct 3D structure (called Structure from Motion or SFM) [23, 22, 24, 3, 53, 46, 47, 25, 40, 41]. The nature of these applications requires accurate geometric estimation. Especially the SfM approaches utilize navigation data, often available on a ROV (Remotely Operated Vehicle for underwater operations) in order to gain more accurately estimated camera poses and/or rely on extensive global optimization (bundle adjustment, refer to [59]). Otherwise, drift, i.e. an accumulating error in the recovery of the camera path,

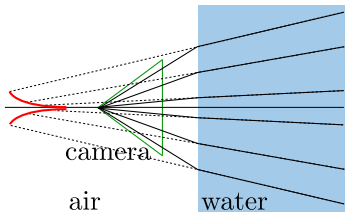

Fig. 4. Radial image of a caustic (marked in red) caused by refraction at a water-air interface. When tracing the rays in water (dashed lines), they are tangents to the caustic.

is a major problem sometimes causing the results to be useless. Some of the authors even mention the erroneous camera model as one of the error sources, but even though, up until now, most Structure from Motion approaches neglect the error caused by refraction. Only recently, research has begun to explicitly incorporate refraction in a specialized SfM-system [6]. Section 5.3 will try to answer the question as to how severe the introduced error is for the applications.

3 Ray-Based Generic Camera Models

A possibility to account for refraction in underwater imaging more explicitly is to use a more generic camera model in underwater scenarios. Such ray-based cameras do not need to have a single viewpoint and are capable of dealing with dome ports and flat ports alike.

Grossberg et al. [17] introduce a generic camera model, where incoming rays are 'somehow' captured by corresponding pixels on the sensor. It is assumed that each pixel records exactly one main ray, no matter where on the ray the sensor array is. Therefore, the central definition of the paper, the raxel, describes one ray per pixel. When parameterizing all rays of an imaging system, there is usually a singularity in the bundle of rays (not true for e.g. orthographic cameras). The locus of this singularity is the caustic (fig. 4), uniquely describing the imaging system. In case of a single view point system, the caustic encompasses only a single point - the center of projection.

In order to compute the caustic, the mapping from image coordinates to rays is differentiated.

$$\mathsf{X}(x,y,\alpha) = \begin{pmatrix} X(x,y,\alpha) \\ Y(x,y,\alpha) \\ Z(x,y,\alpha) \end{pmatrix} = \mathsf{X}_s(x,y) + \alpha\tilde{\mathsf{X}}(x,y), \tag{10}$$

with X_s being the starting point and $\tilde{\mathsf{X}}$ being the direction of the ray starting at image point (x, y) and $\alpha \in \mathbb{R}$ describing the position on the ray. The determinant of the Jacobi matrix of this parametrization is set to zero and solved for the parameter α.

$$\det(J(\mathsf{X}(x,y,\alpha))) = 0 \tag{11}$$

Using α in (10) allows to compute the corresponding point on the caustic for each pixel position (x, y). Grossberg et al. develop a method to compute caustics for arbitrary cameras numerically by projecting differing calibration patterns using an active display. Unfortunately such active displays are not feasible in underwater environments, but as was noted by [58] (see also below), caustics provide a natural connection between the underwater non-single-view-point camera, generic ray-based cameras, and the common pinhole model.

A different work by Narasimhan et al. [38] researches light sheet reconstruction as an application of the described raxel model for small scale underwater images in laboratory settings. A camera is put in front of a water tank, and calibrated by placing two planes into the tank vertically with respect to the optical axis and therefore gaining two points in space for each ray.

In addition to the raxel model, Sturm et al. [55, 54] work with another ray-based model, where each pixel is simply represented by a ray defined by its starting point and its direction of travel. By only assuming that neighboring rays are close to each other, this model is independent of the physical location of the sensor array and does not require an existing caustic, thus making the camera model even more generic than the raxel model. A camera is calibrated by taking several images of a calibration plane, however, the authors mention problems with the calibration robustness. In [54], algorithms for pose estimation, triangulation, multi-view geometry, in short for SfM, are derived and the theory is applicable to the underwater case. [7] concentrates on the case of a refractive plane in an underwater scenario. The derivation only works for one refractive interface (thin glass) and has not yet been implemented.

Another possibility to deal with refraction by approximating ray-based cameras is described in [62]. Here, the camera is viewed as a non-SVP camera having a caustic instead of the single view point. Instead of modeling the refraction effect physically or using a generic ray-based camera, the camera is approximated by several perspective cameras for the different areas of the image. The number of virtual perspective cameras determines the accuracy of this system.

In summary, it can be said, that using a more generic camera model than the pinhole model with distortion allows to deal with refractive effects. However, using independent 3D origins and directions for each ray leads to a high degree of freedom, making the robust calibration of generic camera models difficult, especially in open water. The following section shows that far less parameters need to be determined if refraction is modeled explicitly.

4 Physical Models for Refraction

The third possibility to deal with refraction is to use a physical model that explicitly computes the refraction of rays at the underwater housing. Several methods for achieving this will be compared in this section. They differ in the assumptions made about the glass thickness, normal between interface and image sensor plane, or indexes of refraction and in their derivation.

The two papers presented next describe the theory and calibration method for calibrating underwater cameras with the assumption of a thin piece of flat glass as an interface of the underwater housing.

In [58] by Treibitz et al., the derivation of a refractive model and its calibration for a perspective camera behind a flat port are presented. The authors' underwater housing has a glass thickness of about $5mm$. The ray's shift due to traveling through the glass interface is approximated to be about $0.28mm$ and therefore neglected. In addition, it is assumed that the image sensor and the interface are parallel. This allows examining the projection through a refractive interface by using radial image coordinates, thus making it possible to derive all required equations analytically.

The derivation is based on Fermat's principle (see 1.1):

$$\frac{dt}{dR_i} = n_w \frac{-(R_w - R_i)}{\sqrt{(Z_w - d)^2 + (R_w - R_i)^2}} + n_a \frac{R_i}{\sqrt{d^2 + R_i^2}} = 0, \tag{12}$$

where, (R_i, d) is the radial coordinate on the interface and (R_w, Z_w) is the radial coordinate of the 3D point in the water. For common perspective systems with only small amounts of radial distortion, the following equation holds for all radial coordinates:

$$f \approx \frac{Z_w r}{R_w}, \tag{13}$$

with f being the focal length. This equation can be used to project (radial) coordinates on the glass interface (R_i, d) into the perspective camera:

$$R_i = rd/f \tag{14}$$

Using this in equation (12) yields:

$$\left(R_w - \frac{d}{f} r\right)^2 \left[\left(\frac{f n_w}{r}\right)^2 + (n_w^2 - 1)\right] = Z_w^2 \tag{15}$$

relating r and (R_w, Z_w) in the underwater case. In order to calibrate the camera model, the common parameters for perspective cameras (f, (c_x, c_y), r_1, r_2) as well as the interface distance d are calibrated. Based on (15), the following equation needs to be satisfied:

$$R_w = \frac{Z_w}{\sqrt{\left(\frac{f n_w}{r}\right)^2 + (n_w^2 - 1)}} + \frac{d}{f} r_i \tag{16}$$

which is extended to account for lens distortion. When using two points X_{w_1} and X_{w_2}, they are parametrized by $(R_{w_i}, \varphi_{w_i}, Z_{w_i})$ and their distance in space is estimated using the law of cosines:

$$\hat{s} = \sqrt{\hat{R}_{w_1}^2 + \hat{R}_{w_2}^2 - 2\hat{R}_{w_1}\hat{R}_{w_2} \cos|\varphi_{w_1} - \varphi_{w_2}|}. \tag{17}$$

Using the true distances s, a non-linear optimization is used to solve for the camera parameters. Usually, the intrinsic parameters apart from focal length are estimated beforehand, leaving only the interface distance and the focal length to be calibrated.

As introduced by [17], caustics can be used as a measure of the deviation from the single view point model. [58] derives the caustic analytically using equation (16) (see section 3):

$$R_{caustic} = \left(1 - \frac{1}{n_w^2}\right)\left(\frac{r}{f}\right)^3 d \tag{18}$$

$$Z_{caustic} = -\,n\left[1 + \left(1 - \frac{1}{n_w^2}\right)\left(\frac{r}{f}\right)^2\right]^{1.5} d \tag{19}$$

Obviously, the caustic's extent is directly depending on the interface distance d, therefore, the extent of the caustic can be diminished by moving the entry pupil as close to the glass interface as possible.

Telem et al. describe in [56, 57] a different approach to model refraction. As in the approach described above, the authors use a model with thin glass and parallelism between image sensor and interface in the first paper, but in their photogrammetric model, the authors relate the measured 2D image coordinates to image coordinates eligible for perspective projection. Note that the camera center is not valid for these points, so the intersection with the ray coming from the water and the optical axis is computed for each set of image coordinates as well. The non-refracted rays do not meet in one common center of projection. For each point (in radial coordinates) a value

$$\Delta d = d\left(\frac{n_w}{n_a f}\sqrt{f^2 + r^2\left(1 - \frac{n_a^2}{n_w^2}\right)} - 1\right) \tag{20}$$

specifies the distance between the center of projection and the actual crossing of the non-refracted ray with the optical axis. The measured image points (x, y, f) are modified in the underwater case to fit the perspective projection depending on Δd, the water ray's crossing with the optical axis:

$$\begin{pmatrix} x' \\ y' \\ f' \end{pmatrix} = \begin{pmatrix} x\frac{d}{f} \\ y\frac{d}{f} \\ d + \Delta d \end{pmatrix}. \tag{21}$$

This allows to write the back projection to an underwater ray as:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = (\mathsf{C} + \Delta\mathsf{C}) + \lambda \mathbf{R}^T \begin{pmatrix} x' \\ y' \\ f' \end{pmatrix} \tag{22}$$

with $\Delta\mathsf{C} = -\Delta d\mathbf{r}_3$ being the deviation from the principle point and $\mathbf{r}_3$ being the third row of the rotation matrix $\mathbf{R}$. In a second paper [57], the authors extend

their approach to incorporate glass interfaces that are not parallel to the image sensor, causing Δd to become more complicated. However, in our tests, we found that the ray coming from the water not necessarily intersects the optical axis if the interface is not parallel to the image sensor. Errors introduced by a non-zero glass thickness are absorbed by the interface distance. In the calibration tests, the intrinsics are estimated first, then four additional parameters for the underwater case are calibrated: d, $n = n_w/n_a$, and two parameters for the interface rotation. The results show that it is possible to estimate the required underwater parameters and the missing camera poses without getting large correlations between the parameters.

An often cited method [28, 29] establishes a way to combine refraction with the pose estimation using the Direct Linear Transform (DLT [18]). However, parallelism between interface and image sensor was be achieved by manually rotating the hardware, and the distance between interface and camera center is measured. The authors have so far not included an estimation of those parameters into their algorithm.

Up until now, all described methods considering a physical refraction model assumed thin glass and, except for one, parallelism between interface and image sensor. Li et al. [32, 31] (see also [35]) describe an approach based on back projecting image points, with a stereo rig where the complete physical model is calibrated: the double refraction of rays at the air-glass and the glass-water interfaces is modeled explicitly. Here, the light is assumed to travel through $p+1$ different refractive media and thus is refracted p times. This is derived using Snell's law instead of Fermat's principle: the points (X_i, Y_i, Z_i) and $(X_{i-1}, Y_{i-1}, Z_{i-1})$ denote the points on the i-th and i-1-th interface surfaces. The path length of the ray between those interfaces is:

$$\rho = \sqrt{(X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2 + (Z_i - Z_{i-1})^2}. \tag{23}$$

In addition, it is assumed that the start and end points (X_0, Y_0, Z_0) and $(X_{p+1}, Y_{p+1}, Z_{p+1})$ are known as well as the functions of the refractive surfaces $F_i(X_i, Y_i, Z_i) = 0$ with their existing derivation: $\left[\frac{\partial F_i}{\partial X_i}, \frac{\partial F_i}{\partial Y_i}, \frac{\partial F_i}{\partial Z_i}\right]^T$. Using those notations, Snell's law is applied at each refractive point: $n_i \sin\theta_i = n_{i+1} \sin\theta_i'$ and the ray between the interfaces is determined by:

$$\tilde{\mathsf{X}}_i = \frac{1}{\rho_i} \begin{pmatrix} X_i - X_{i-1} \\ Y_i - Y_{i-1} \\ Z_i - Z_{i-1} \end{pmatrix}. \tag{24}$$

With $\mathbf{n}_i$ being the normal at the interface point (computed using the derivatives of function F_i), θ_i and θ_i' are computed using the scalar product:

$$\cos\theta_i = \mathbf{n}^T \tilde{\mathsf{X}}_i \qquad\qquad \cos\theta_i' = \mathbf{n}^T \tilde{\mathsf{X}}_{i+1}, \tag{25}$$

allowing the computation of the following function using Snell's law and the fact that both rays and the normal form the same plane:

$$\tilde{\mathsf{X}}_{i+1} = \frac{n_i}{n_{i+1}} \tilde{\mathsf{X}}_i - \left(\frac{n_i}{n_{i+1}} \cos\theta_i - \cos\theta_i' \right) \mathbf{n}_i. \tag{26}$$

Using (26), inner interface points are back projected, then refracted twice resulting in outer interface points and rays in water eligible for triangulation using the stereo rig. The calibration routine assumes known indexes of refraction and estimates the intrinsics and rig extrinsics from images taken in air. Then the underwater parameters are calibrated by taking images of a three-dimensional calibration object underwater using linearized versions of the equations derived above to find an initial solution. The accuracy evaluation in [32] showed that the errors of reconstructed 3D points are between $6mm$ and $6cm$ for the optical axis and $6mm$ and $1cm$ for the x- and y-axes. In [31], an additional reduced central projection allows to project points from 3D through a refractive interface onto the image plane with an iterative method that solves for the required unknowns on the interfaces.

In [27], the usage of a perspective camera in an underwater scenario is examined as well as a flat port and dome port model. The back projection is derived by computing rays in air, glass, and water using Snell's law and quaternion rotations (refer to section 5.1). Projections are computed numerically. In addition, a calibration routine is proposed assuming intrinsics, indexes of refraction, and glass thickness to be known. Nested loops of a Levenberg Marquardt routine [48] are used to solve for the remaining interface parameters and the camera's poses with respect to a calibration pattern. Unfortunately, the authors did not implement and examine their calibration routine, but conclude that consideration of refraction is necessary when exploring the implicitly contained geometric information from images due to the model error (see section 5). Chang and Chen [6] made a promising start in developing an actual 3D-reconstruction algorithm with explicit consideration of refraction. The cameras are assumed to view the object of interest through the planar water surface. The vertical direction of the camera is assumed to be known, so only the heading of the camera needs to be computed.

Another approach to using physical models is found in the works of Maas, [33, 34] and a follow-up work by Putze [50, 49]. The goal of both methods is optical fluid flow analysis in fairly small laboratory tanks, where the fluid has been marked with a set of particles. In the model, the actual 3D points in space are substituted by their corresponding virtual 3D points, fitting the perspective back projection. The computation of these points is based on an iteration with known interface parameters and indexes of refraction. In order to calibrate the system, a calibration pattern below water at known distances is used and optimized by a bundle adjustment routine. The method has been found to perform well if the indexes of refraction, especially for the glass are known. A correlation analysis shows high correlation between focal length and distance between camera center and glass interface for all three calibrated cameras. The works of Maas also contain an introduction to epipolar geometry [18] in case of refractive imaging, where the epipolar lines are bent into curves. If the ray in water from one camera is known, several points on this ray are projected into the second image defining a linear approximation of the epipolar curve. This is for example used in [11] examining surface reconstruction.

In addition, there exist some more exotic applications also considering refraction explicitly. In contrast to the approaches described above, where the indexes of refraction are assumed to be known, here, they can be calibrated only in very confined laboratory scenarios. See [37, 64, 65, 26, 10] for more detailed information.

The methods for using a physical model of refraction on underwater images show that calibrating such systems is possible only if extra assumptions about interface-sensor parallelism, indexes of refraction, or glass thickness are made or a stereo rig is used. Until now, methods utilizing geometric information contained in images usually rely on the perspective camera model, but [27] already showed that a considerable error is caused by using the wrong camera model, however, we found that an inclination angle between housing interface and image sensor is even worse than different interface - camera distances. Therefore, the analysis in [27] will be extended in the following section.

5 Accuracy Analysis of the Perspective Model

In this section, the exact derivation of the physical underwater ray cast will be explained. This ray cast is then used to compute synthetic data compliant with the underwater model allowing to compute for example the caustic for the case of non-parallel interface and sensor plane. As shown in section 2, most authors still work using a perspective camera on underwater images and this section aims at examining the resulting error and its compensation in detail by using the synthetic data computed by physically modeling refraction.

5.1 Physical Underwater Projection

The derivation of the ray cast in the physical underwater model presented here follows [27], but is more detailed and considers projection routines. Note that other papers using Snell's law for the derivation come to similar conclusions.

Flat Port Back Projection in case of a flat port in front of an underwater housing, the distance to the port, the glass thickness, and the normal of the glass within the camera coordinate system are important parameters. Here, the inner interface plane is parametrized by $\Pi_i = (\underbrace{n_1, n_2, n_3}_{\mathbf{n}_\Pi}, -d)$ containing the normal and the port's distance to the camera origin. In addition, the outer interface plane is parametrized by the same normal and the glass thickness d_g: $\Pi_o = (n_1, n_2, n_3, -(d+d_g))$ (fig. 5). When back projecting an image point in the underwater case, the goal is the computation of the point on the outer interface plane and the direction of the ray within the water. First, the image point $\mathbf{x}$ needs to be turned into a ray within the camera's underwater housing:

$$\tilde{\mathsf{X}}_a = \mathbf{K}^{-1}\mathbf{x}, \tag{27}$$

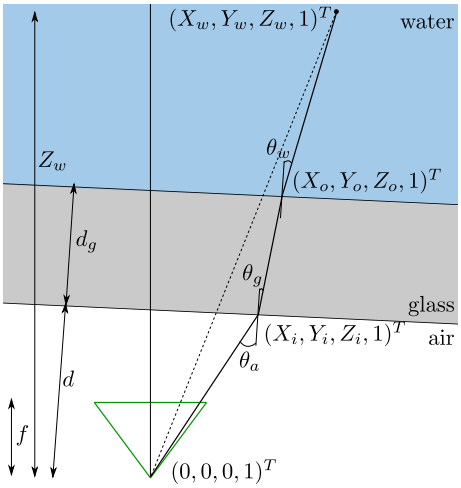

Fig. 5. When back projecting a point (solid line), the ray travels from the camera through air until it intersects the inner interface plane $(X_i, Y_i, Z_i, 1)^T$. After being refracted, the ray travels through glass until intersecting the outer interface plane $(X_o, Y_o, Z_o, 1)^T$, is then refracted, and finally travels through water reaching the 3D point in water $(X_w, Y_w, Z_w, 1)^T$. Projecting $(X_w, Y_w, Z_w, 1)^T$ without refraction (dashed line) yields a different pixel in the image.

with the subscript a denoting the coordinates within the underwater housing, in air, and $\mathbf{K}$ being the camera matrix containing the intrinsic parameters. The ray is in the camera coordinate system, i.e. the center of projection is in the origin.

In order to find the intersection X_i between ray and interface the following equation is used:

$$\Pi_i^T \begin{pmatrix} \lambda_g \tilde{X}_a \\ \lambda_g \tilde{Y}_a \\ \lambda_g \tilde{Z}_a \\ 1 \end{pmatrix} = 0 \qquad \Rightarrow \lambda_g = \frac{d}{\mathbf{n}_\Pi^T \tilde{\mathsf{X}}_a} \qquad \Rightarrow \mathsf{X}_i = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_g \tilde{\mathsf{X}}_a. \tag{28}$$

The intersection of the port's inner plane and the ray, parametrized by λ_g, is used to determine the point on the inner plane of the interface X_i. After that, the ray's direction within the glass is computed: the scalar product between the plane's normal $\mathbf{n}_\Pi$ and the ray in air yields the angle between normal and incident ray before refraction, and is then refracted:

$$\theta_a = \arccos\left(\frac{\mathbf{n}_\Pi^T \tilde{\mathsf{X}}_a}{\|\mathbf{n}_\Pi\| \|\tilde{\mathsf{X}}_a\|}\right) \qquad \theta_g = \arcsin(\sin\theta_a \frac{n_a}{n_g}). \tag{29}$$

The ray being incident upon the inner interface plane needs to be rotated/refracted. This is described by a rotation around the normal resulting from the cross prod-

uct of the plane normal and the incoming ray:

$$\mathbf{n}_{rot} = \frac{\mathbf{n}_{\Pi} \times \tilde{\mathsf{X}}_a}{\|\mathbf{n}_{\Pi}\| \|\tilde{\mathsf{X}}_a\| \sin\theta_a}. \tag{30}$$

The rotation angle is $\theta_{rot} = \theta_g - \theta_a$ and the rotation around the axis $\mathbf{n}_{rot}$ is best described by a unit quaternion:

$$\mathbf{q} = \begin{pmatrix} \frac{\sin\left(\frac{\theta_{rot}}{2}\right)}{\|\mathbf{n}_{rot}\|} \mathbf{n}_{rot} \\ \cos\left(\frac{\theta_{rot}}{2}\right) \end{pmatrix}. \tag{31}$$

This quaternion is applied to the ray $\tilde{\mathsf{X}}_a$, yielding the refracted ray $\tilde{\mathsf{X}}_g$, which describes the light's traveling direction within the glass. Now, the point on the outer interface needs to be computed:

$$\lambda_w = \frac{(d_g + d - \mathbf{n}_{\Pi}^T \mathsf{X}_i)}{\mathbf{n}_{\Pi}^T \tilde{\mathsf{X}}_g} \qquad \Rightarrow \mathsf{X}_o = \mathsf{X}_i + \lambda_w \tilde{\mathsf{X}}_g. \tag{32}$$

The ray within the glass is refracted again, using the indexes of refraction for glass and water, the cross product, and the unit quaternion rotation. The result is the ray in water $\tilde{\mathsf{X}}_w$. The 3D point can be computed, if the distance *dist* between the camera center and the 3D point is known:

$$||\mathsf{X}_o + \alpha_w \tilde{\mathsf{X}}_w|| = dist \tag{33}$$

This equation can be solved for α_w yielding the distance the ray needs to travel from the interface point:

$$\mathsf{X}_w = \mathsf{X}_o + \alpha_w \tilde{\mathsf{X}}_w. \tag{34}$$

X_w is still in the camera coordinate system, but using the transform of the camera pose, the point can easily be transformed into the world coordinate system.

Dome Port Back Projection the dome is parametrized by its center with respect to the camera's center of projection and its inner and outer radius. In case of perfect alignment of the dome center and the camera's center of projection, the project and back project functions are equal to the common pinhole camera model. Otherwise, the refraction at the dome needs to be modeled explicitly, but the only difference to the method described above is found in the intersection of the rays in air or glass and the inner and outer interface respectively. To compute the intersection point, the inner and outer dome spheres are parametrized by using the quadric:

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{35}$$

A transformation containing the sphere's inner r_i or outer r_o radius and the translation of the dome's center $(X_d, Y_d, Z_d)^T$ are applied to the quadric to get the matrix describing the dome:

$$\begin{aligned}\mathbf{H}_i =& \begin{pmatrix} r_i & 0 & 0 & X_d \\ 0 & r_i & 0 & Y_d \\ 0 & 0 & r_i & Z_d \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ \mathbf{D}_i =& (\mathbf{H}^{-1})^T \mathbf{Q} \mathbf{H}^{-1}.\end{aligned} \tag{36}$$

A homogeneous point $\mathbf{X}$ lies on the quadric $\mathbf{D}$ if $\mathbf{X}^T\mathbf{D}\mathbf{X} = 0$. Using the parametrization for the ray in air or in glass, the intersections of the rays with the inner or outer dome surface can be determined. The normals at those intersection points can be found by using the line from the center of the dome to the intersection points. Once the normals, the intersection points, and the ray directions in air and glass are known, the remaining derivation of the refraction is exactly the same as in the flat port case.

Projection in contrast to [27], we analyze the projection of 3D points into the camera in more detail, using an approach building upon [58]. The problem in this case is caused by the unknown points on the inner and outer interface. In order to derive a formula for the projection, Fermat's principle is applied. The total traveling time of the ray is the sum of three components: the time spent in the underwater housing (in air), the time spent in the glass of the interface, and the time spent in the water. The derived equation contains four unknowns, the x- and y-coordinates on the inner and outer interface planes (X_i and Y_i and X_o and Y_o respectively):

$$\begin{aligned} t(X_i, Y_i, X_o, Y_o) = & \\ & n_{air}\sqrt{X_i^2 + Y_i^2 + Z_i^2} + \\ & n_{glass}\sqrt{(X_o - X_i)^2 + (Y_o - Y_i)^2 + (Z_o - Z_i)^2} + \\ & n_{water}\sqrt{(X_w - X_o)^2 + (Y_w - Y_o)^2 + (Z_w - Z_o)^2}. \end{aligned} \tag{37}$$

Since the light always travels the distance in the least time, this equation's partial derivatives are used to minimize the traveling time with respect to the unknowns:

$$\frac{\partial t}{\partial X_i} = 0 \qquad \frac{\partial t}{\partial Y_i} = 0 \qquad \frac{\partial t}{\partial X_o} = 0 \qquad \frac{\partial t}{\partial Y_o} = 0. \tag{38}$$

The plane equations are utilized to eliminate the Z-components:

$$\begin{aligned} Z_i =& \frac{d - n_1 X_i - n_2 Y_i}{n_3} \\ Z_o =& \frac{d + d_g - n_1 X_o - n_2 Y_o}{n_3}. \end{aligned} \tag{39}$$

focal length	1100 px
image size	1001×801 px
principal point	middle of image
distortion	$r_1 = 0$, $r_2 = 0$, $t_1 = 0$, and $t_2 = 0$
aspect ratio	1
skew	0
index of refraction water	1.333
index of refraction glass	1.5
index of refraction air	1
interface distance	$20mm$
glass thickness	$30mm$
interface tilt	1.5°

Table 3. Parameters used for caustic computation.

The resulting system of equations with four equations and four unknowns is solved numerically using e.g. Powell's hybrid method [1] [48]. After that, the points on the inner and outer interface planes are determined, however, only the point on the inner interface plane is relevant for projecting it onto the image plane with the usual perspective projection.

In our tests, we found that it is difficult to find the correct solution using this method, especially in case of a negative camera-interface distance. This occasionally happens, when the entry-pupil of the camera lies in front of the physical camera housing (refer to [1, 58]). In case of thin or no glass, parallelism between interface and image sensor and positive interface distance d, (38) is only depending on the radial coordinate on the refractive plane. The derivative in this direction becomes a polynomial of fourth degree [14, 58]. For this special case, [14] proved that the correct/practical root is found in the interval $[0, R_w]$. In experiments in our more general case, with possibly negative d and non-parallel interface, this is no longer true. In order to deal with all possible cases, the projection can also be solved numerically (as in [27]). This is accomplished by an optimization, which is initialized using the common perspective projection. After that, the Nelder-Mead-Simplex routine[2] [48] is used to compute the correct 2D point.

5.2 Caustics as a Measure of Deviation from the SVP

Caustics present the bridge between physically modeled underwater cameras and more generic camera models. The extent of a caustic is also a measure of the deviation from the perspective single view point camera. [17, 62, 58] describe methods for deriving caustics analytically. In more generic models with thick glass and no parallelism between the sensor and the interface, the analytic derivation becomes infeasible.

[1] e.g. in GSL library from www.gnu.org/software/gsl/
[2] NLOPT toolbox from ab-initio.mit.edu/nlopt/

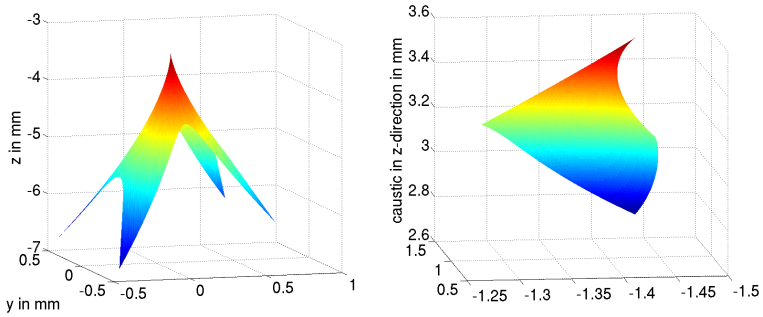

Fig. 6. Left: the caustic for a flat port camera housing with imperfect sensor-interface alignment. Right: caustic in the dome port case with imperfect alignment.

Alternatively, the outer interface points and directions of the ray in water are computed using the back project function described above. The derivatives for the Jacobi matrix are computed numerically. α (parameterizing the point on each ray, which lies on the caustic) is expressed in terms of the entries of the Jacobian. Once α is known, the ray parametrization is used to compute the caustic point for each image point (x, y). Figure 6 shows an exemplary caustic for the parameters in table 3,and figure 7 is an example for the extent in x- y- and z-direction, which changes with focal length and distance between camera center and interface, and can be in the order of centimeters.

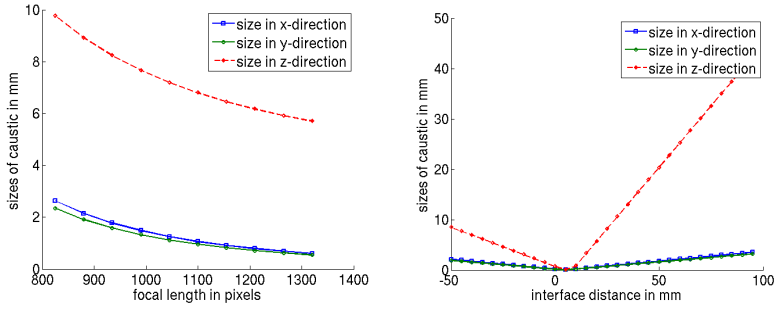

Fig. 7. Left: caustic size depending on focal length for $30mm$ glass thickness and interface tilt $= 1.5°$. Right: caustic size depending on interface distance for $30mm$ glass thickness and interface tilt $= 1.5°$.

focal length	800,1000 px
image size	600×800 px
principal point	middle of image
distortion	no distortion
aspect ratio	1
skew	0
index of refraction water	1.333
index of refraction glass	1.45
index of refraction air	1
interface distance	10 - 80 mm
glass thickness	5, 30, 60 mm
interface rotation	0° or $2 - 3^\circ$
rig baseline (no rotation)	200 mm
distance range checkerboard images	1000-10000 mm

Table 4. Parameters used for synthetic tests.

5.3 Accuracy of the Perspective Model in Calibration, Triangulation, and SfM

In this section, results of the accuracy analysis of using the perspective model on underwater images from [58, 27] are extended, especially considering slight rotations of the interface plane.

Error Compensation in Perspective Calibrations Using an implementation of the model described above, a thorough examination based on synthetic data is possible. The synthetic images were rendered according to the underwater projection model for a stereo camera rig. In order to examine the influence of different underwater housings, different sets of calibration images (50 for each set) showing a checkerboard pattern were rendered with different parameters. Using the exact checkerboard corners to eliminate effects from erroneous corner detection, the camera rigs were calibrated perspectively using [52]. Table 4 summarizes the parameters for different test cases.

When using the error-free 2D3D correspondences from perspective projections for calibration in [52], the final re-projection error is in the order of 10^{-8} (model and data fit perfectly). When using 2D3D correspondences compliant with the underwater model, the final re-projection error is in the order of ($\varnothing <$ 0.05 pixel), which still suggests a good fit to the perspective model. As stated by [12, 30], the focal length changes according to the refractive index of water when calibrating perspectively, see figure 8 on the left. The underwater images were rendered without any distortion, so the four resulting parameters (fig. 8, right) give an idea about how much the images are altered by refraction. Obviously, tangential distortion does not compensate the error induced by tilting the interface. Figure 9 on the left shows the resulting errors in principal points of the calibration. In the case of a slightly rotated interface plane, part of the error

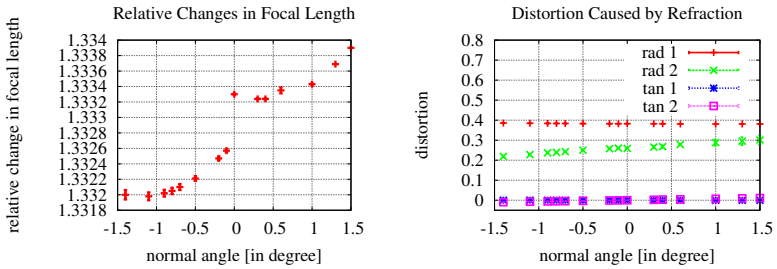

Fig. 8. Left: relative change in focal length when calibrating perspectively: $\frac{fl_{persp}}{fl_{water}}$. The true index of refraction for water was 1.333, thus, refraction is partly compensated for by using a virtual sensor plane. Right: distortion introduced by using the perspective calibration (camera within the underwater housing had zero distortion). Plotted are the four used coefficients for distortion: two each for radial and tangential distortion. Radial distortion mainly compensates for refraction in general, and tangential distortion does not compensate a rotation of the interface plane.

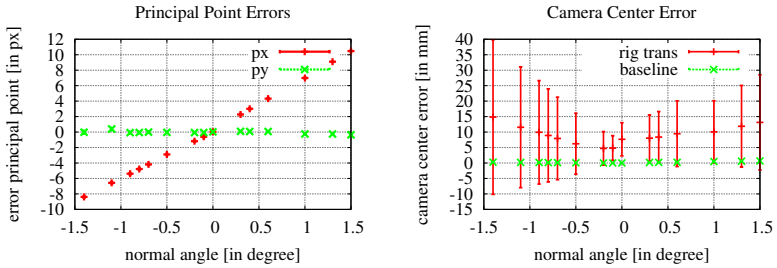

Fig. 9. Left: errors of estimated principal points in pixels depending on the interface angle. Right: errors of estimated camera centers in mm depending on the interface angle.

is absorbed by moving the principal point. Furthermore, the computed camera centers have an increasing error and increasing covariances (see figure 9 on the right), not only suggesting problems with robustness, but an error in the extrinsic parameters during the calibration causes errors during applications later on.

In case of dome ports, [27] came to the conclusion that perspective models are accurate enough if the camera center does not move more than $1cm$ from the dome center.

Triangulation Errors When using a perspectively calibrated underwater camera for tasks such as measuring using stereo rigs or computing 3D reconstructions, accuracy and drift reduction play important roles. The error induced by using the perspective model for triangulating points is shown in figure 10. The left image shows a rendering of 3 cameras and 20 triangulated points. In dark

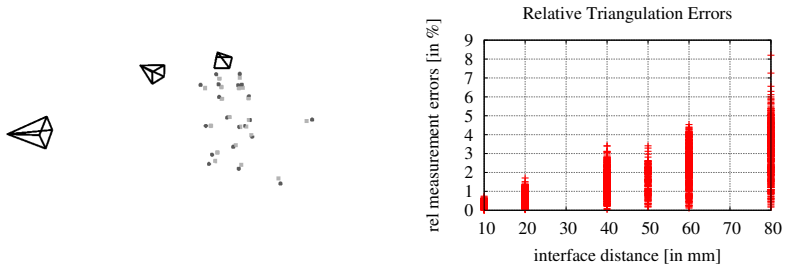

Fig. 10. Left: triangulation results after triangulating with three cameras. Dark gray: correct points triangulated with physical model for underwater camera, light gray: points triangulated with perspective calibration of the underwater images. Right: relative triangulation errors using the perspective calibrations depending on the interface distance (relative with respect to camera-point distance in %).

gray are the true points triangulated using the underwater model, while the light gray points were triangulated using the perspective model. It can be seen clearly that the perspective calibration has an area where it fits well, allowing fairly accurate reconstruction, while in other areas of the 3D space high triangulation errors occur. The right figure 10 shows triangulation errors depending on the interface distance for the stereo rig calibrated in different parameter configurations. In addition to the dependence on the interface distance, the error also depends on the distance of the points used for triangulation to the center of projection and the distance range of the calibration pattern with respect to the camera. Figure 11 extends the analysis of plane triangulation in [27] by comparing parallel and non-parallel interfaces: the black (red in color version) planes are triangulated using the underwater model (interface distance $20mm$, glass thickness $30mm$), while the gray (green in color version) planes are triangulated using the perspective calibration. In case of parallelism between interface and image sensor, the error is radially symmetric (11, left), while in the right image, a slight rotation of the interface plane causes far higher errors.

Errors in Pose Estimation Figure 12 shows the reconstruction of a cylinder captured from cameras moved on an orbit with slight interface rotation. The error induced by the wrong camera model clearly accumulates. Note that the correspondences used are synthetic and therefore not biased by feature detection and matching methods, so all of the drift in this scenario is caused by the model error alone, increasing especially in case of even slight rotations between interface and image sensor.

Aside from other sources of error not present in the synthetic data presented here (e.g. errors in checkerboard detection), the measurement errors induced by using an incorrect imaging model do not bode well for exact measurements of underwater structures. This matches the conclusions drawn in [27, 6]: underwater

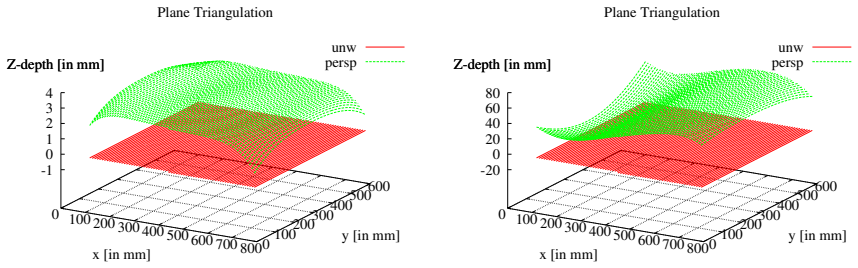

Fig. 11. Two of the perspective calibration scenarios, both with interface distance $20mm$ were used to triangulate points on the xy-plane, with the camera being $2m$ away, viewing the xy-plane at an $45°$ angle. The left scenario has a parallel interface with respect to the imaging sensor, while in the right scenario the interface was tilted by $(-1°, 1°)$ with the resulting errors in the perspective calibration. 'persp' stands for perspective triangulation, while 'unw' stands for refractive triangulation. Note the different scales of the z-axis.

SfM so far works even in case of several thousand images, however, navigational data or time consuming bundle adjustment methods are required to stabilize the motion computation and reduce drift.

6 Conclusion and Future Work

We have discussed three different types of camera models, which are used to deal with refraction effects on underwater images.

First, it was shown that the often used pinhole camera model is invalid due to refraction at the camera housing, although it is common in the literature. The accuracy analysis for the perspective model shows that the model error is not negligible and grows with increasing interface distance and with stronger tilt of the interface with respect to the image sensor. Applications like stereo measurements, mosaicking for navigation, and Structure from Motion all rely on accurate geometrical measurements. Especially Structure from Motion is prone to errors due to drift in pose estimation and we believe that the systematic error caused by using a wrong model for refractive effects adds an unnecessary source of drift.

Second, the ray-based camera models have a completely derived theory for SfM, but no implementation has been tried on real underwater images yet. In addition, the high degree of freedom caused by individually parametrized rays for each pixel makes robust calibration difficult or even infeasible in underwater environments.

Third, physically modeled interfaces allow to compute refraction explicitly without needing a high degree of freedom. Only the parameters describing the underwater housing with respect to the camera are required in addition to the

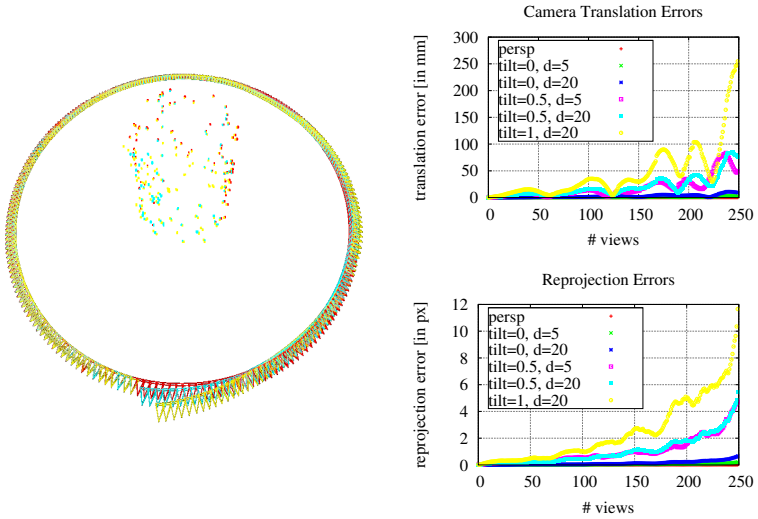

Fig. 12. The reconstructed points in the middle all lie on a cylinder (scene size = $2500mm$). In dark gray (red in the color version) are the cameras and points in the perspective scenario. In medium gray (cyan in the color version) ($tilt = 0.5°$, $d = 20$) and light gray (yellow in the color version) ($tilt = 1°$, $d = 20$) are the points and cameras, when the points originate from the underwater model, while a perspectively calibrated camera is used to compute the reconstruction using a classical SfM approach. On the right, error curves for the different scenarios are shown.

classic perspective camera model. Applications like SfM, mosaicking, and stereo based measurements could therefore profit from using such a model because the systematic error induced by using an approximate, perspective camera model can be eliminated by modeling refraction explicitly. Future Work will include robust calibration of the interface parameters and application of the physical model to underwater images.

Authors	Application	Method
Perspective Model		
Freyer et al. [12]	calibration	calibrate in air and find perspective water calibration by adapting focal length ($f_{water}1.333 = f_{air}$) and distortion ($\delta r = \left(\frac{\cos\theta_w}{\cos\theta_a} - 1\right) r$, with r being the radial distortion in air, θ_w being the angle between optical axis and water ray, and θ_a being the angle between optical axis and air ray)
Lavest et al. [30]	calibration	calibrate in air and find perspective water calibration by adapting focal length ($f_{water}1.333 = f_{air}$) and distortion ($1.333u_{air} + du_{air}$) $= u_{water} + du_{water}$, with u_{air} and u_{water} being the distorted coordinates and du_{air} and du_{water} being the distortion corrections)
Bryant et al. [4]	calibration	finding checkerboard corners robustly in turbid environments; calibrate based on underwater images; using one coefficient for radial lens distortion
Pessel et al. [45, 43, 44]	calibration	checkerboard-free self calibration approach using predefined trajectory; no distortion modeled because lens system eliminated distortion effects by 99%; calibrate on-site to adapt calibration to changing index of refraction of water
Harvey et al. [19]	stereo measurement	usage of stereo rig to measure underwater structures; examination of calibration robustness in different water bodies; 3D calibration frame
Costa et al. [8]	stereo measurement	automatically measure fish size using a stereo rig; automatic contour detection and interest point triangulation; initial calibration without distortion, removal of inconsistencies by training neural network; 5% measuring accuracy
Gracias et al. [15, 16]	mosaicing	a mosaic computation; used for navigation after wards; self calibration with rotating camera on pan-tilt unit, sequential mosaic building followed by global optimization; geo-referenced
Garcia, Carreras et al. [13, 5]	mosaicing	a mosaic computation; used for navigation after wards; one parameter for radial distortion; second paper using robot in pool with coded pattern on the ground for estimating the accuracy of other on-board navigation devices
Xu, Negahdaripour et al. [42, 63]	mosaicing	first: simultaneous mosaicing, navigation, and station keeping; second: statistical combination of image-based registration data and other navigation data sources applied to mosaic computation
Eustice et al. [9]	mosaicing	compares different methods for mosaicing under consideration of movement with growing complexity (from translation to full projective transformations) in underwater environments
Trucco et al. [60]	mosaicing	mosaicing approach via tracked features and homography estimation; registered images are warped into common image

Table 5. Paper overview.

Authors	Application	Method
Hogue et al. [23, 22]	3D reconstruction	a bumblebee stereo camera and IMU are combined in one underwater housing and used to reconstruct and register 3D structure; reconstruction shows a lot of drift if IMU is not used and authors presume erroneous camera model to cause part of it
Jasiobedzki et al. [24]	3D reconstruction	real-time reconstruction using stereo images, registered using ICP; resulting model is bent, authors plan to incorporate refraction to eliminate the error
Sedlazeck et al. [53]	3D reconstruction	classical, sequential SfM with adaptations to underwater environment; calibration below water, 2 coefficients for radial distortion, dense depth maps are used for model computation; additional color correction; absolute scale from navigation data
Pizarro et al. [46, 47]	3D reconstruction	calibration below water including distortion; reconstructions based on 2 or 3 images are registered against each other by a graph based algorithm; Delaunay triangulation; usage of navigation data
Johnson et al. [25]	3D reconstruction	sparse sets of 3D points are meshed using a Delaunay triangulation and registered via SLAM utilizing navigation data; additional loop closing and color correction; can process thousands of images
Brandou et al. [3]	3D reconstruction	stereo rig is moved on predefined trajectory by a ROV arm; model is computed using dense depth maps; camera is calibrated on-site by deploying checkerboard on the sea floor
Negahdaripour et al. [40, 41]	3D reconstruction	combination of optical and acoustic systems in one rig; calibration and reconstruction theory in presence of both: euclidean and spherical coordinates
Queiroz-Neto, Nascimento et al. [51, 39]	underwater stereo	color correction routine is combined with stereo to match more robustly; no consideration of refraction
Ray-based Models		
Grossberg, Narasimhan et al. [17, 38]	calibration, reconstruction	cameras are defined via their caustics; calibration routine using an active display; second paper specializes on underwater case with light sheet based reconstruction in small tank environments
Sturm, Chari et al. [55, 54, 7]	calibration, reconstruction	development of theory for generic cameras described only by their rays (assumption is that neighboring rays are close to each other); calibration by taking several checkerboard images; theory, but no implementation of ray-based cameras for refractive case (thin glass)
Wolff [62]	sea floor reconstruction	reconstruction of sea floor in simulator (small tank); ray-based, generic camera is approximated by several perspective cameras suitable for different image regions

Table 6. Paper overview.

Authors	Application	Method
Physical Underwater Models		
Treibitz et al. [58]	calibration	physical model for underwater imaging is developed assuming thin glass and interface-sensor parallelism; analytical derivation of projection using cylinder coordinates; includes calibration routine; bridge to caustics
Telem et al. [56, 57]	calibration	each point is mapped to a point eligible for perspective projection by moving the point in the image and computing the correct intersection with the optical axis
Kwon et al. [28, 29]	calibration, measurement	refraction is modeled in combination with the DLT for pose estimation; assumed thin glass; no rotation between glass and camera sensor
Kunz et al. [27]	calibration	hemispherical and flat ports are modeled and synthetic data is used to experiment with inaccuracies using the perspective model; calibration routine is described, but not implemented; general case for non-parallel interfaces and thick glass
Li et al. [31, 32]	calibration	development of complete physical model and its calibration; stereo rig is used to triangulate the derived rays in water; indexes of refraction are assumed to be known
Maas [33, 34]	fluid flow measurement	a complete physical model for a rig is derived and implemented, then used for calibration; 3D points are moved to positions eligible for perspective projection by an iteration; rig is used to reconstruct fluid flow marked by suspended particles; parallelism is assumed; indexes of refraction cannot be calibrated
Putze [50, 49]	calibration, fluid flow measurement	follow-up work to the above gaining more robustness
Förstner et al. [11]	reconstruction	[33] and its specialized epipolar geometry are used for surface reconstruction
Morris et al. [37]	wave surface reconstruction	a calibrated stereo rig views the bottom of a water tank on which a checkerboard pattern is placed; refraction is used to determine the wave's normals on the liquid's surface
Yamashita, Kawai et al. [64, 65, 26]	measurements	in small water tanks within a lab, a stereo system, a laser beam, or active patterns are used to gain reconstructions of objects completely or half emerged in the water
Ferreira et al. [10]	underwater stereo	the underwater model is linearized to compensate of the majority of the errors induced by using the perspective model for stereo
Chang et al [6]	underwater SfM	underwater SfM with known vertical rotation (IMU) and explicit consideration of refraction

Table 7. Paper overview.

References

1. Aggarwal, M., Ahuja, N.: A pupil-centric model of image formation. International Journal of Computer Vision 48(3), 195–214 (2002)
2. Bouguet, J.Y.: Visual methods for three-dimensional modelling. Ph.D. thesis, California Institute of Technology Pasadena, CA, USA (1999), http://etd.caltech.edu/etd/available/etd-02072008-115723/
3. Brandou, V., Allais, A., Perrier, M., Malis, E., Rives, P., Sarrazin, J., Sarradin, P.: 3d reconstruction of natural underwater scenes using the stereovision system iris. In: Proc. OCEANS 2007 - Europe. pp. 1–6 (2007)
4. Bryant, M., Wettergreen, D., Abdallah, S., Zelinsky, A.: Robust camera calibration for an autonomous underwater vehicle. In: Australian Conference on Robotics and Automation (ACRA 2000) (August 2000)
5. Carreras, M., Ridao, P., Garcia, R., Nicosevici, T.: Vision-based localization of an underwater robot in a structured environment. In: Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on. vol. 1, pp. 971 – 976 vol.1 (sep 2003)
6. Chang, Y.J., Chen, T.: Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction. In: IEEE International Conference on Computer Vision (ICCV) (2011)
7. Chari, V., Sturm, P.: Multiple-view geometry of the refractive plane. In: Proceedings of the 20th British Machine Vision Conference, London, UK (sep 2009), http://perception.inrialpes.fr/Publications/2009/CS09
8. Costa, C., Loy, A., Cataudella, S., Davis, D., Scardi, M.: Extracting fish size using dual underwater cameras. Aquacultural Engineering 35(3), 218 – 227 (2006), http://www.sciencedirect.com/science/article/B6T4C-4K4WMTT-1/2/d8103dd1d6946795645bf447642c7813
9. Eustice, R., Singh, H., Howland, J.: Image registration underwater for fluid flow measurements and mosaicking. In: OCEANS 2000 MTS/IEEE Conference and Exhibition. vol. 3, pp. 1529 –1534 vol.3 (2000)
10. Ferreira, R., Costeira, J.P., Santos, J.A.: Stereo reconstruction of a submerged scene. In: Marques, J.S., Prez de la Blanca, N., Pina, P. (eds.) Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, vol. 3522, pp. 102–109. Springer Berlin / Heidelberg (2005)
11. Foerstner, W., Wolff, K.: Exploiting the multi view geometry for automatic surfaces reconstruction using feature based matching in multi media photogrammetry. In: Proceedings of the 19th ISPRS Congress. pp. 5B 900–907 (2000)
12. Fryer, J.G., Fraser, C.S.: On the calibration of underwater cameras. The Photogrammetric Record 12 (1986)
13. Garcia, R., Batlle, J., Cufi, X., Amat, J.: Positioning an underwater vehicle through image mosaicking. In: Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on. vol. 3, pp. 2779 – 2784 vol.3 (2001)
14. Glaeser, G., Schröcker, H.P.: Reflections on refractions. Journal for Geometry and Graphics (JGG) 4, 1–18 (2000)
15. Gracias, N., Santos Victor, J.: Underwater video mosaics as visual navigation maps. Computer Vision and Image Understanding, Journal of (CVIU) 79(1), 66–91 (July 2000)
16. Gracias, N., van der Zwaan, S., Bernardino, A., Santos-Victor, J.: Mosaic-based navigation for autonomous underwater vehicles. Oceanic Engineering, IEEE Journal of 28(4), 609 – 624 (oct 2003)

17. Grossberg, M.D., Nayar, S.K.: The raxel imaging model and ray-based calibration. International Journal of Computer Vision 61(2), 119–137 (2005)
18. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision (Second Edition). Cambridge University Press, second edn. (2004), http://www.amazon.com/Multiple-View-Geometry-Computer-Vision/dp/0521540518
19. Harvey, E.S., Shortis, M.R.: Calibration stability of an underwater stereo-video system : Implications for measurement accuracy and precision. Marine Technology Society journal 32, 3–17 (1998)
20. Hecht, E.: Optik. Oldenburg Verlag Muenchen Wien (2005)
21. Heikkila, J., Silven, O.: A four-step camera calibration procedure with implicit image correction. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 0, 1106 (1997)
22. Hogue, A., German, A., Jenkin, M.: Underwater environment reconstruction using stereo and inertial data. In: Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on. pp. 2372 –2377 (7-10 2007)
23. Hogue, A., German, A., Zacher, J., Jenkin, M.: Underwater 3d mapping: Experiences and lessons learned. In: Computer and Robot Vision, 2006. The 3rd Canadian Conference on. pp. 24 – 24 (07-09 2006)
24. Jasiobedzki, P., Se, S., Bondy, M., Jakola, R.: Underwater 3d mapping and pose estimation for rov operations. In: OCEANS 2008. pp. 1 –6 (15-18 2008)
25. Johnson-Roberson, M., Pizarro, O., Williams, S., Mahon, I.: Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. Journal of Field Robotics 27 (2010)
26. Kawai, R., Yamashita, A., Kaneko, T.: Three-dimensional measurement of objects in water by using space encoding method. In: Robotics and Automation, 2009. ICRA '09. IEEE International Conference on. pp. 2830 –2835 (12-17 2009)
27. Kunz, C., Singh, H.: Hemispherical refraction and camera calibration in underwater vision. In: OCEANS 2008. pp. 1 –7 (15-18 2008)
28. Kwon, Y.: A camera calibration algorithm for the underwater motion analysis. In: ISBS - Conference Proceedings Archive, 17 International Symposium on Biomechanics in Sports (1999) (1999)
29. Kwon, Y., Casebolt, J.: Effects of light refraction on the accuracy of camera calibration and reconstruction in underwater motion analysis. Sports Biomech 5(1), 95–120 (2006)
30. Lavest, J.M., Rives, G., Lapresté, J.T.: Underwater camera calibration. In: ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II. pp. 654–668 (2000)
31. Li, R., Tao, C., Zou, W.: An underwater digital photogrammetric system for fishery geomatics. In: Intl. Archives of PRS. vol. Vol.XXXI, pp. pp.319–323 (1996)
32. Li, R., Li, H., Zou, W., Smith, R., Curran, T.: Quantitative photogrammetric analysis of digital underwater video imagery. Oceanic Engineering, IEEE Journal of 22(2), 364 –375 (apr 1997)
33. Maas, H.G.: Digitale Photogrammetrie in der dreidimensionalen Stroemungsmesstechnik. Ph.D. thesis, Eidgenoessische Technische Hochschule Zuerich (1992)
34. Maas, H.G.: New developments in multimedia photogrammetry. In: Optical 3-D Measurement Techniques III. Wichmann Verlag, Karlsruhe (1995)
35. McGlone, J.C. (ed.): Manual of Photogrammetry. ASPRS, 5th edn. (2004), http://www.amazon.de/Manual-Photogrammetry-American-Society/dp/0937294012
36. Mobley, C.D.: Light and Water: Radiative Transfer in Natural Waters. Academic Press (1994)

37. Morris, N., Kutulakos, K.N.: Dynamic refraction stereo. In: Proc. 10th Int. Conf. Computer Vision. pp. 1573–1580 (2005)
38. Narasimhan, S.G., Nayar, S.: Structured light methods for underwater imaging: light stripe scanning and photometric stereo. In: Proceedings of 2005 MTS/IEEE OCEANS. vol. 3, pp. 2610 – 2617 (September 2005)
39. Nascimento, E.R.d., Campos, M.F.M., Barros, W.F.d.: Stereo based structure recovery of underwater scenes from automatically restored images. In: Nonato, L.G., Scharcanski, J. (eds.) Proceedings SIBGRAPI 09 (Brazilian Symposium on Computer Graphics and Image Processing). IEEE Computer Society, Los Alamitos (Oct 11–14, 2009 2009), http://urlib.net/sid.inpe.br/sibgrapi@80/2009/08.18.16.07
40. Negahdaripour, S., Sekkati, H., Pirsiavash, H.: Opti-acoustic stereo imaging, system calibration and 3-d reconstruction. In: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on. pp. 1 –8 (17-22 2007)
41. Negahdaripour, S., Sekkati, H., Pirsiavash, H.: Opti-acoustic stereo imaging: On system calibration and 3-d target reconstruction. Image Processing, IEEE Transactions on 18(6), 1203 –1214 (jun 2009)
42. Negahdaripour, S., Xu, X., Khamene, A., Awan, Z.: 3-d motion and depth estimation from sea-floor images for mosaic-based station-keeping and navigation of rovs/auvs and high-resolution sea-floor mapping. In: Autonomous Underwater Vehicles, 1998. AUV'98. Proceedings Of The1998 Workshop on. pp. 191 –200 (aug 1998)
43. Pessel, N., Opderbecke, J., Aldon, M.J.: Camera self-calibration in underwater environment. In: WSCG (2003)
44. Pessel, N.: Auto-Calibrage d'une Camra en Milieu Sous-Marin. Ph.D. thesis, Universit Montpellier II (2003)
45. PESSEL, N., J., O., M.-J., A.: An experimental study of a robust self-calibration method for a single camera. In: 3rd International Symposium on Image and Signal Processing and Analysis, ISPA'2003, sponsored by IEEE and EURASIP. Rome, Italie (september 2003)
46. Pizarro, O., Eustice, R., Singh, H.: Relative pose estimation for instrumented, calibrated imaging platforms. In: DICTA. pp. 601–612 (2003)
47. Pizarro, O., Eustice, R., Singh, H.: Large area 3d reconstructions from underwater surveys. In: Proc. OCEANS '04. MTTS/IEEE TECHNO-OCEAN '04. vol. 2, pp. 678–687 Vol.2 (2004)
48. Press, W.H., Vetterling, W.T., Teukolsky, S.A., Flannery, B.P.: Numerical Recipes in C++: the art of scientific computing. Cambridge University Press, New York, NY, USA, 2nd edn. (2002)
49. Putze, T.: Erweiterte verfahren zur mehrmedienphotogrammetrie komplexer krper. In: Beitrge der Oldenburger 3D-Tage 2008. Herbert Wichmann Verlag, Heidelberg (2008)
50. Putze, T.: Geometrische und stochastische Modelle zur Optimierung der Leistungsfaehigkeit des Stroemungsmessverfahrens 3D-PTV. Ph.D. thesis, Technische Universitaet Dresden (2008)
51. Queiroz-Neto, J.P., Carceroni, R., Barros, W., Campos, M.: Underwater stereo. In: Proc. 17th Brazilian Symposium on Computer Graphics and Image Processing. pp. 170–177 (Oct 17–20, 2004)
52. Schiller, I., Beder, C., Koch, R.: Calibration of a pmd camera using a planar calibration object together with a multi-camera setup. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. Vol.

XXXVII. Part B3a, pp. 297–302. Beijing, China (2008), www.mip.informatik.uni-kiel.de/tiki-index.php?page=Ingo+Schiller-23k-, xXI. ISPRS Congress
53. Sedlazeck, A., Koser, K., Koch, R.: 3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions. In: Proc. OCEANS '09. OCEANS 2009-EUROPE. pp. 1–10 (May 11–14, 2009)
54. Sturm, P., Ramalingam, S., Lodha, S.: On calibration, structure from motion and multi-view geometry for generic camera models. In: Daniilidis, K., Klette, R. (eds.) Imaging Beyond the Pinhole Camera, Computational Imaging and Vision, vol. 33. Springer (aug 2006), http://perception.inrialpes.fr/Publications/2006/SRL06
55. Sturm, P.F., Ramalingam, S.: A generic concept for camera calibration. In: ECCV (2). pp. 1–13 (2004)
56. Telem, G., Filin, S.: Calibration of consumer cameras in a multimedia environment. In: ASPERS 2006 Annual Conference (2006)
57. Telem, G., Filin, S.: Photogrammetric modeling of underwater environments. ISPRS Journal of Photogrammetry and Remote Sensing 65(5), 433 – 444 (2010), http://www.sciencedirect.com/science/article/B6VF4-50F9H66-1/2/d8dba566f79b0a207e13a6aa2bf3f69d
58. Treibitz, T., Schechner, Y., Singh, H.: Flat refractive geometry. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008. pp. 1–8 (2008)
59. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle adjustment – a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) Vision Algorithms: Theory and Practice. Lecture Notes in Computer Science, vol. 1883, pp. 298–372. Springer-Verlag (2000)
60. Trucco, E., Doull, A., Odone, F., Fusiello, A., Lane, D.: Dynamic video mosaicing and augmented reality for subsea inspection and monitoring. In: In Oceanology International, United Kingdom (2000)
61. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses, an efficient and accurate camera calibration technique. IEEE Journal of Robotics and Automation RA-3(4), 323–344 (1987)
62. Wolff, K.: Zur Approximation allgemeiner optischer Abbildungsmodelle und deren Anwendung auf eine geometrisch basierte Mehrbildzuordnung am Beispiel einer Mehrmedienabbildung. Ph.D. thesis, Rheinische Friedrich-Wilhelms-Universitaet Bonn (2007)
63. Xu, X., Negahdaripour, S.: Application of extended covariance intersection principle for mosaic-based optical positioning and navigation of underwatervehicle. In: ICRA'01. pp. 2759–2766 (2001)
64. Yamashita, A., Hayashimoto, E., Kaneko, T., Kawata, Y.: 3-d measurement of objects in a cylindrical glass water tank with a laser range finder. In: Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on. vol. 2, pp. 1578 – 1583 vol.2 (27-31 2003)
65. Yamashita, A., Fujii, A., Kaneko, T.: Three dimensional measurement of objects in liquid and estimation of refractive index of liquid by using images of water surface with a stereo vision system. In: ICRA. pp. 974–979 (2008)
66. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the International Conference on Computer Vision. pp. 666–673. Corfu, Greece (1999), http://www.citeulike.org/user/snsinha/article/238276