An Efficient Solution to Ray Tracing Problems in Multimedia Photogrammetry for Flat Refractive Interfaces

Robin Rofallski¹ · Thomas Luhmann¹

Received: 14 September 2021 / Accepted: 7 January 2022 / Published online: 2 March 2022
© The Author(s) 2022

Abstract

Understanding and considering refraction effects are important parts of the demanding task of multimedia photogrammetry, especially with planar interfaces, so-called “flat ports”. Yet, it remains challenging to determine reliable calibration results that are both quickly acquired and physically interpretable. In this contribution, a novel object-based optimization algorithm, relying on ray tracing methods, is introduced. It enables calibrating physical parameters of all involved refractive properties with reduced computational effort, compared to other standard algorithms in ray tracing. We show that this solution produces equally accurate results as other ray tracing approaches while improving processing speed by a factor of approximately ten and providing a statistical metric in object space. Furthermore, we show in a laboratory investigation that explicit calibration of refractive properties is crucial even with orthogonally aligned bundle-invariant interfaces for highest accuracy, as accuracy in object space is decreased by about 10% with implicit calibration. With deviation from orthogonality by about ten degrees this decreases even further to almost no useful results and accuracy loss of more than 50% compared to explicit calibration results.

Keywords Multimedia photogrammetry · Underwater photogrammetry · Camera calibration · Ray tracing · Geometric modeling · Bundle adjustment

Zusammenfassung

Eine effiziente Lösung für Raytracing Probleme in der Mehrmedienphotogrammetrie für flache refraktive Trennflächen. Das Verständnis und die Berücksichtigung von Brechungseffekten sind wichtige Bestandteile der anspruchsvollen Aufgabe der Mehrmedienphotogrammetrie, insbesondere bei ebenen Trennflächen, sogenannten “Flat Ports”. Dennoch bleibt es eine Herausforderung, zuverlässige Kalibrierungsergebnisse zu bestimmen, die sowohl schnell erfasst als auch physikalisch interpretierbar sind. In diesem Beitrag wird ein neuartiger objektbasierter Optimierungsalgorithmus vorgestellt, der auf Raytracing-Methoden beruht. Er ermöglicht die Kalibrierung der physikalischen Parameter aller beteiligten Brechungseigenschaften mit geringerem Rechenaufwand als andere Standardalgorithmen mit Raytracing. Wir zeigen, dass diese Lösung ebenso genaue Ergebnisse liefert wie andere Raytracing-Ansätze, während sie die Verarbeitungsgeschwindigkeit um das zehnfache erhöht und ein statistisches Genauigkeitsmaß im Objektraum liefert. Darüber hinaus zeigen wir in einer Laboruntersuchung, dass die explizite Kalibrierung der Brechungseigenschaften selbst bei orthogonal ausgerichteten, bündelinvarianten Trennflächen für höchste Genauigkeit entscheidend ist, da die Genauigkeit im Objektraum bei impliziter Kalibrierung um etwa zehn Prozent abnimmt. Bei einer Abweichung von der Orthogonalität um etwa zehn Grad verringert sich die Genauigkeit weiter, bis hin zu kaum nutzbaren Ergebnissen und einem Genauigkeitsverlust von mehr als fünfzig Prozent im Vergleich zu den Ergebnissen der expliziten Kalibrierung.

✉ Robin Rofallski
Robin.Rofallski@jade-hs.de

Thomas Luhmann
Thomas.Luhmann@jade-hs.de

¹ Jade University of Applied Sciences, Institute for Applied Photogrammetry and Geoinformatics, Ofener Str. 16/19, 26121 Oldenburg, Germany

1 Introduction

Image acquisition in or through multiple media, especially water, suffers from many quality-degrading and geometry-altering influences, compared to single-media image acquisition. First, the light from the object travels through

multiple media (air, glass, water) and is refracted at the interfaces separating these. This introduces astigmatism, chromatic aberration and color shift towards green or blue, hence decreasing image quality and sharpness. Moreover, the ray path is altered at refracting interfaces according to Snell's law, thus rendering the standard pinhole model with additional distortion parameters invalid (Treibitz et al. 2012). This is often compensated by either calibrating under water with standard parameters of photogrammetric models (e.g. Brown (1971)), designed for air-applications to implicitly calibrate the imaging geometry or by explicit calibration of the refractive imaging properties. In turn, refraction can also be compensated by employing a hemispherical interface (dome port). If the entrance pupil of the lens is perfectly aligned with the hemisphere's center, all rays pass the interface orthogonally and hence no refraction occurs. However, mechanical alignment to the accuracy that no significant errors are introduced is a rather complicated task and requires very fine positioning with special methods (She et al. 2019). Even when aligned, the position of the entry pupil changes with different focal distances which enforces realignment for different settings (Menna et al. 2016). As investigated by Kunz and Singh (2008), unresolved misalignments result in nonlinear distortions that cannot be fully compensated by standard distortion models. Therefore, flat ports are widely used and it remains of high interest to fully understand and calibrate the imaging geometry for photogrammetric analyses, especially when requiring high accuracy. However, many calibration models developed over the years have some backlashes that either introduce a significant amount of additional computation time or complicate full understanding of parameters. Especially for robotic navigation tasks underwater, e.g. by remotely operated vehicles (ROV), time constraints, preferably real-time, are of major importance. Algorithms for optical real-time orientation, such as simultaneous localization and mapping (SLAM) require high performance. This also applies to large structure from motion (SfM) networks which are impracticable with many models due to long computation times. Based on well-known ray tracing procedures, we hence introduce a novel optimization algorithm in object space. This improves these disadvantages by introducing a different error function to reduce computational complexity significantly.

These aspects are important in a variety of underwater applications, ranging from navigation with optical sensors to measuring objects and sea species underwater, such as offshore pillars, welding seams (Ekkel et al. 2015), nuclear power plants (Przybilla et al. 1990), reefs (Rossi et al. 2021; Rofallski et al. 2020a), wrecks (Drap et al. 2015; Prado et al. 2019; Rofallski et al. 2020b) or archaeological sites (Costa 2019). Furthermore, multimedia photogrammetry plays

an important role in tasks on land, e.g. in tracking of fluid dynamics (Particle Image Velocimetry), requiring highest accuracy for particle tracking (Maas et al. 2009).

This paper outlines as follows: After giving an overview on related work with emphasis on multimedia photogrammetry with single cameras and flat ports, we present our methodology on solving ray tracing problems and describe our implementation in a powerful bundle adjustment. The approach is then investigated in an underwater evaluation in a water tank at Jade University. Special focus of the evaluation is on the influence of different angles between the optical axis and the interface normals. After presenting and discussing results, the paper closes with a conclusion and an outlook.

2 Related Work

Several models and calibration approaches for multimedia photogrammetry have been developed over the years. These can be separated into three categories regarding the inclusion of refraction:

1. Implicit compensation by underwater calibration and use of standard pinhole camera with distortion models, such as Brown (1971).
2. Strict modeling of refractive indices and interfaces in the ray path from object to sensor.
3. Alternative or virtual camera models, discarding the pinhole model.

A complete review on all approaches is not within the scope of this article. We limit ourselves to referring to significant contributions, describing monocular approaches with flat interfaces. Extensive reviews on related work can be obtained from Sedlazeck and Koch (2012), Shortis (2015, 2019) or Dolereit (2018).

Implicit compensation is strictly spoken incorrect, not physically interpretable and creates distance-dependent distortions, as shown by Treibitz et al. (2012). However, in several practical applications that do not rely on highest accuracy or fulfill favorable positioning of the camera towards the interfaces, this can be a negligible influence. Since no commercial photogrammetric or structure-from-motion (SfM) software offer explicit modeling of refraction, this is a widely used practical approach to multi-media photogrammetry. Kang et al. (2012), Sedlazeck and Koch (2012) and Kahmen et al. (2019) systematically investigated the feasibility of this approach. All studies show that cameras facing orthogonally to the interface with a high percentage of water in the optical ray path, can well be calibrated applying the standard pinhole model with additional distortion parameters. This is congruent with other studies

observing similar behavior (Kotowski 1988; Treibitz et al. 2012; Shortis 2015; Kahmen et al. 2020b). However, while Kang et al. (2012) and Kahmen et al. (2019) solely focus on this case, Sedlazeck and Koch (2012) and Hemmati and Rashtchi (2019) show that deviating from this optimum setup (i.e. non-orthogonality of optical axis and interface) results in unresolvable errors when employing the pinhole camera model. Unfortunately, no specific threshold values or equations can be generated to evaluate resulting errors and have to be investigated for each specific setup individually. Lavest et al. (2000, 2003) investigated the change in interior orientation parameters and proposed a formula to predict resulting underwater distortions from in-air calibrations in two-media cases. Taking up on their research, Agrafiotis and Georgopoulos (2015) extended the approach, considering the percentage of water and air in the optical path.

Strict modeling of the ray path was developed by several authors, relying on an explicit description of all physical properties (camera model, refractive indices, positioning and shape of interfaces) by applying Snell's law. Kotowski (1987, 1988) introduced a prominent approach explicitly calculating the ray path through an arbitrary amount of interfaces and media (ray tracing). Refracting points for bundle adjustment are calculated by a separate module that iteratively optimizes three conditions by means of Gauß–Newton-optimization. The approach was quite flexible but suffered in terms of robustness and ambiguity, as stated later by Mulsow (2010). Maas (1992, 1995) developed a simplification of the former for the specific three-media case (air–glass–water) and introduced a lookup-table for avoiding computationally expensive calculations in bundle adjustment and other photogrammetric analyses. Furthermore, Mulsow (2010) and later Mulsow and Maas (2014) and Mulsow (2016), implemented an approach for a flexible module, based on ray tracing. The approach, called "minimum distance alternating forward ray tracing", is an iterative procedure to achieve a strict ray tracing projection to an imaging sensor. To overcome the limiting issue of an unknown initial ray direction, the model coordinate in image space is obtained by minimizing the distance between ray and object point in an inner loop. These procedures demand a high number of additional iterations to calculate the approximate image observations. Thus, analytic derivatives are disabled and computation times increased, possibly up to an impractical duration for bundles with many observations. To overcome this issue, Agrawal et al. (2012) proposed solving a 4th- (two-media case: air–water) or 12th-order (three-media case: air–glass–water) polynomial. Other authors use approximations, such as Duda and Gaudig (2016), proposing a Taylor approximation or Yau et al. (2013) employing a bisection search. Henrion et al. (2015) developed a ray tracing approach for measurement of marine fauna, based on known object points and refractive indices. The distance between a known object point coordinate

and a ray is minimized. However, no explicit calculation methods on the exact procedure are given, which makes the functional approach hard to reproduce. Furthermore, it relies on a priori known object points, hence not including a bundle approach. Elnashef and Filin (2019) developed a model description that separates the orientation estimation from the refractive parameters. They derive a direct estimation of the parameters by first solving for five degrees of freedom that are independent of refraction and then including an adjusted focal length and ratio of refractive indices of water and air, absorbing the refraction effect of glass by the focal length. Most recently, Nocerino et al. (2021) proposed to include a polynomial, explicitly addressing the distance dependence without introducing interface and refraction parameters. Simulation results show a good compensation of refraction effects for both flat and dome ports.

Alternative camera models based on altering the projection either by assuming multiple projection centers that invalidates the pinhole model or transforming the projection to computationally favorable positions. An example of such models is presented in Treibitz et al. (2012) who state that the system is rather to be described as a set of projection centers on a caustic surface. These depend on the distance and intersection angle of the object point with the sensor and are thus not unique. They proposed calibrating a ray map to assign a parametric ray to each pixel that includes refraction but does not have a single projection center. Agrawal et al. (2012) showed that a multimedia system can be described as an axis of projection centers, assuming all interfaces are parallel. Furthermore, it is stated that, employing their approach, multiple interfaces may well be approximated by a single interface assumption, hence simplifying calculations to a fourth order polynomial. However, as residual errors remain in simulations, this holds true only if highest accuracy is not demanded. Chadebecq et al. (2017) took up the idea of an axial camera for two-media photogrammetry and formulated a four-view-constraint to enforce consistency in exterior orientation. The algorithmic description of Agrawal et al. (2012) is claimed to suffer from computational instabilities and hence the introduced four-view-constraint overcomes this disadvantage in case of simulations and laboratory data with endoscopes. Based on the axial-camera assumption, Łuczyński et al. (2017) developed their "pinax-model" for a runtime-optimized procedure for dense matching. Imagery is undistorted by a pre-generated lookup-table for efficient computation of 3D depth maps.

Jordt-Sedlazeck and Koch (2012) proposed creating a virtual camera at the intersection point of an unrefracted ray and the interface normal, assuming interfaces are parallel. All object points are projected onto this virtual sensor, rendering a synthetic image and differences to the real images are minimized for each pixel (analysis by synthesis). Similarly, Telem and Filin (2010) apply a perspective

shift of the projection center, specifically designed for the refractive geometry of flat ports.

The contribution of this paper is twofold:

1. A novel optimization procedure to significantly speed up strict ray tracing calculations is introduced for flat ports. The idea bases on the approaches by Mulsow and Maas (2014) but shifts the optimization procedure from image space to object space, hence eliminating computationally expensive iterations over each image observation while explicitly accounting for refraction. Furthermore, no additional transformations have to be performed while still allowing for analytical derivatives.
2. The significance of strict modeling is investigated with respect to different intersection angles of the optical axis with a flat interface

The main goal is to employ the given mathematical description of ray tracing procedures, such as in Kotowski (1987) and Mulsow and Maas (2014), and simplify it for speed considerations. Furthermore, to include additional sensor observations (e.g. from environmental sensors), interpretability and decoupling refractive effects from the interior orientation is of higher importance, hence disabling implicit and alternative models. Eventually, the actual use for explicit models is to be discussed in typical ROV applications, considering the refraction geometry.

3 Methodology

Similar to approaches by Kotowski (1988) and Mulsow and Maas (2014), our contribution is based on ray tracing and equally assumes homogeneity as well as isotropy of all involved media. Ray tracing is a common practice in computer graphics to render synthetic images from 3D data. Explicitly, the light path between an object and the sensor is calculated to simulate interaction of objects with light. Hence, the image ray is considered as a 3D vector, thus transferring the analytical collinearity equations into vector calculus. The procedure can either be described from the sensor to the object or the other way around. The intuitive photogrammetric bundle description from an object to a projected image coordinate (backwards ray tracing) is significantly more complicated. Therefore, we first consider the opposite direction from the sensor to the object (forward ray tracing). Note that the denotation of forward/backward ray tracing is consistent with Mulsow (2010) as it follows the photogrammetric terms. This is contrary to the denotation in computer graphics and computer vision and should be kept in mind when comparing approaches.

3.1 Forward Ray Tracing

The procedure is depicted in Fig. 1. Starting from a measured image coordinate and given (approximate) parameters of interior orientation, distortion correction is applied iteratively by Gauß–Newton minimization. Hence, coordinates are corrected for radial-symmetric and tangential-asymmetric (decentering) distortion, affinity and shear and principal point before continuing with computations (Luhmann et al. 2020). The undistorted coordinate is then transformed to a vector starting at the respective position on the sensor and passing through the projection center by employing parameters of interior and exterior orientation. This creates an arbitrarily scaled directional 3D vector $\vec{A}_1$.

$$\begin{pmatrix} A_{1x} \\ A_{1y} \\ A_{1z} \end{pmatrix} = \mathbf{R} \cdot \begin{pmatrix} x' \\ y' \\ c \end{pmatrix} \quad (1)$$

$\mathbf{R}$ Rotation matrix
 A_{1x}, A_{1y}, A_{1z} Directional vector in first medium
 x', y' Undistorted image coordinates
 c Principal distance

Combined with the position of the projection center $P_0 = (X_0, Y_0, Z_0)^T$, the image ray can be represented as a line in vector form in the global coordinate system, as follows.

$$g : x = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + s \cdot \begin{pmatrix} A_{1x} \\ A_{1y} \\ A_{1z} \end{pmatrix} \quad (2)$$

Fig. 1 Forward ray tracing through n media. From the measured point p'_m , distortion correction is applied to gain p'_u . The outgoing ray $\vec{A}_i$ (red) is then refracted at $n - 1$ interfaces $t_i(N_i, d_i)$ according to the refractive indices μ_i and the normal vectors of the interfaces $\vec{N}_i$. An unrefracted ray (green) with the projected point p'^* is depicted for reference [extended after Mulsow and Maas (2014)]

Following Snell's law, we then intersect the ray with the first interface from the projection center and apply refraction in 3D space. The interface is defined in the camera coordinate system for bundle-invariant interfaces and hence has to be transformed to the global coordinate system by a Helmert-Transformation (6DOF). However, object-invariant, i.e. non-moving, interfaces can easily be integrated by simply omitting the aforementioned transformation, hence fixing it to the global coordinate system. It is worth noting that interfaces can be of any shape that allow for calculating an intersection point and a surface normal, such as planes, spheres or other implicitly described shapes. As we deal with flat ports in this study, the description as a plane and hence line-plane intersection will be assumed in the following. The plane is described for the given i th projection center P_{0_i} lying on the plane in the Hesse normal form, fulfilling:

$$\vec{N}_i \cdot \vec{P}_{0_i} - d_i = 0. \quad (3)$$

We constrain $\|\vec{N}_i\| = 1$ for which the scalar d_i equals the orthogonal distance from the interface to the projection center. Hence, this description is practically interpretable. The intersection point results from calculating the scaling s and inserting into Eq. (2) as follows:

$$s = \frac{d_i - \vec{N}_i \cdot \vec{P}_{0_i}}{\vec{N}_i \cdot \vec{A}_i}. \quad (4)$$

This results in the intersection point on the next interface and a refracted directional vector that travels through the next medium (Glassner 1989).

$$\vec{A}_{i+1} = \mu_{rel} \cdot \vec{A}_i + \left(\mu_{rel} \cdot C_i - \sqrt{1 + \mu_{rel}^2 \cdot (C_i^2 - 1)} \right) \cdot \vec{N}_i \quad (5)$$

$\vec{A}_i$ Directional vector in i th medium
 $\vec{N}_i$ Normal vector on i th interface
 C_i Reversed incidence angle $-\cos(\theta_i) = -\vec{N}_i \cdot \vec{A}_i$
 μ_{rel} Relative refractive index μ_i/μ_{i+1}

This procedure is performed recursively until all involved media are passed, e.g. twice for the bundle-invariant case of underwater photogrammetry, air–glass–water. The last vector with a point on the outer interface $P_{o_{i,j}}$ then represents the outgoing ray A_n .

3.2 Error Function in Object Space

This process, however, is only straight forward in the uni-directional case from the sensor to object space. Since the initial direction of the ray starting at the object point is unknown, the opposite direction cannot be calculated

directly. Other authors who base their strict algorithm on ray tracing perform separate iterative procedures to find the projected image coordinate by either solving for conditions (Kotowski 1987, 1988; Maas 1992, 1995) or using iterative forward ray tracing (Mulsow 2010; Mulsow and Maas 2014; Mulsow 2016). However, the approaches include additional iterations and hence pose significant computational increase to the entire procedure and furthermore prevent analytical derivatives to decrease speed even further.

Our approach overcomes these limitations by introducing a novel error function in object space while maintaining a similar mathematical representation from "minimum distance alternating forward ray tracing" (MDFRT), introduced by Mulsow and Maas (2014) for photogrammetric applications, especially bundle adjustment. Instead of minimizing the difference between projected and measured 2D image coordinate Δ_{xy} (reprojection error), the function is defined in object space as a 3D vector between the outgoing ray of forward ray tracing and the approximate value of the object coordinate Δ_{xyz} . This corresponds to the minimum distance algorithm for forward intersection but is extended to an arbitrary amount of observations per point and is hence capable of solving for bundle adjustment (Fig. 2). Hence, in contrast to Schinstock et al. (2009), we do not explicitly find stereo observations for a given point and perform forward intersection but rather calculate the distance of each image ray to the approximate 3D coordinate of the point.

The orthogonal distance $\Delta_{xyz_{i,j}}$ between the outgoing image ray and the approximate point coordinate is calculated as follows, resulting in a 3D residual vector (Mulsow and Maas 2014):

Fig. 2 Visualization of the error function in object space. The orthogonal line-point distance $\Delta_{xyz_{i,j}}$ is calculated as a 3D vector that is subject to minimization by means of least squares

$$\Delta_{xyz_{i,j}} = \vec{P}_j - \vec{P}_{o_{i,j}} - \vec{A}_n \cdot \left(\frac{\vec{A}_n \cdot (\vec{P}_j - \vec{P}_{o_{i,j}})}{|\vec{A}_n|^2} \right) \quad (6)$$

$\vec{P}_j$ j th point in object space

$\vec{P}_{o_{i,j}}$ Starting point of the outgoing ray for point j
and exterior orientation i

$\vec{A}_n$ Direction vector of outgoing ray.

The procedure from Eqs. (1–6) can be calculated directly. Provided, no distortion parameters are adjusted, the entire process can be solved analytically and hence, analytical derivatives are well computable for linearization. In this case, distortion parameters are assumed to remain constant and undistortion is performed once for all iterations. In turn, distortion parameters can also be simultaneously estimated which then would disable analytical derivatives, as nonlinear undistortion has to be performed for each iteration.

3.3 Optimization in Object Space by Nonlinear Least Squares

For optimization, we assume given approximate values that can be obtained from a single-media approach bundle adjustment and rough estimates of the interface parameters (e.g. orthogonal orientation, distances d_i measured with a ruler) and refractive indices (e.g. from tables). From there, we apply nonlinear least-squares adjustment to find the minimum of the error function over all image observations projected to object space n_{obs} .

$$\sum_{i=1}^{n_{\text{obs}}} \|A_{xyz_i}\| \Rightarrow \min \quad (7)$$

Since the functional approach includes nonlinear parameters, linearization has to be performed at approximate values for the unknown parameters to obtain the Jacobian and the reduced observation vector. Equation (8) is solved for $\vec{dx}$ which is then added to the approximate values of the unknowns before the next iteration. The procedure is repeated until convergence below a given threshold ϵ .

$$(\mathbf{J}^T \cdot \mathbf{P} \cdot \mathbf{J}) \cdot \vec{dx} = \mathbf{J}^T \cdot \mathbf{P} \cdot \vec{dl}$$

$\mathbf{J}$ Jacobian containing partial derivatives

$\mathbf{P}$ Weight matrix

$\vec{dx}$ Reduced vector of unknowns

$\vec{dl}$ Reduced vector of observations

Similar to the collinearity equations, this model can be employed for all photogrammetric 3D applications. This includes spatial resection, forward intersection or

bundle adjustment with arbitrary numbers of points, exterior orientations, interior orientations, interfaces and image observations.

3.4 Implementation into a Powerful Bundle Adjustment

For practical applications, a bundle adjustment software, based on the C++ optimization library *Ceres Solver* (Agarwal et al. 2021) was developed and the model integrated. Ceres can be set up either to automatically calculate derivatives (analytic and numeric) or to use provided pre-determined derivatives for speed optimization. Additionally, the software provides a modular flexibility for a wide range of applications and enables changing camera models quite easily within the bundle. The entire optimization process can be performed by Ceres automatically. However, due to limitations regarding the introduction of linear constraints, e.g. for datum and scale definition, we only rely on self-provided derivatives (collinearity equations), numeric derivatives (for our own model) and function evaluation. The latter will be transferred to analytic derivatives in the future for a significant decrease in computation time. Optimization and solving equation systems are performed separately in our own implementation.

For datum definition, we use Helmert constraints for six degrees of freedom (translation and rotation) and distance constraints for scale definition and hence a free network adjustment. The constraints are formalized in the matrix of constraints $\mathbf{B}$ and the vector of discrepancies $\vec{w}$. These are bordered on the normal equation matrix and absolute term vector. Further datum definitions, e.g. by ground control points or constraints between the unknown parameters are fundamentally possible and shall be introduced in the near future.

$$\mathbf{C} = \begin{pmatrix} \mathbf{J}^T \mathbf{P} \mathbf{J} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \quad \vec{c} = \begin{pmatrix} \mathbf{J}^T \mathbf{P} \vec{dl} \\ \vec{w} \end{pmatrix} \quad (9)$$

For computational stability, we employ an implementation of the well-known Levenberg–Marquardt algorithm with multiplicative extension, according to Marquardt (1963). This accounts for possible instabilities in the bundle, avoiding divergence during iteration. This is performed by introducing a damping factor λ that is added to the main diagonal elements of the normal equation matrix. If the iteration results in a higher standard deviation, the parameters are discarded, λ is increased and the system of equations is solved again to follow the steepest gradient. Otherwise, the new parameters are accepted and λ is decreased towards a (faster) Gauß–Newton optimization. This constrains the next iteration to result in a lower standard deviation than its preceding

iteration. Extending Eq. (8), the full system of equations that is solved for each iteration is:

$$(\mathbf{C} + \lambda \cdot \text{diag}(\mathbf{C})) \cdot \vec{dx} = \vec{c}. \quad (10)$$

For computational efficiency, the system of equations is solved by Cholesky decomposition. Elements of the cofactor matrix $\mathbf{Q}_x = \mathbf{C}^{-1}$ are determined explicitly after convergence, employing the standard algorithm for sparse block matrices (Wolf et al. 2014).

The bundle has a number of advantages, building on both an efficient optimization for ray tracing and a performant computing library. Thus, we are capable of solving large systems of equations, containing several ten or hundred thousands of observations with reasonable time requirements (minutes on Intel i7, 7th Gen. 2.90 GHz, 16Gb RAM). Additionally, flexible integration of different camera models and refraction correction enables use for a wide variety of applications, other than multimedia photogrammetry, as it has been used to implement a local distortion model and alternative radial-symmetric distortion parameters for Hastedt et al. (2021). However, practical limitations exist, concerning both the state of implementations and determinability of parameters.

3.5 Advantages and Limitations of the Approach

In the following, advantages and limitations are mainly presented for the case of bundle-invariant interfaces, i.e. the camera is housed inside a pressure housing and moved rigidly together. This mostly regards the parallel estimation of refractive and interior orientation parameters.

Theoretically, only one refractive index has to remain constant in all ray tracing models, as only the relation between refractive indices is taken into account. However, empirical tests showed that in the usual air–glass–water case, determining the refractive index of more than one medium is challenging and often leads to insignificant results. Hence, we assume refractive indices of both air and glass to be known, as they can be well determined beforehand, using a refractometer or environmental sensors. Furthermore, the limited range of temperature, humidity and pressure in an underwater housing have a rather small influence on the refractive index of air. The same applies to the orientation and distance of the second glass interface and determination of interior orientation. These parameters are highly correlated if estimated with the full model. For practical applications, the thickness (i.e. $d_2 - d_1$) can be measured with a caliper and the approximation of parallel interfaces (i.e. $N_1 = N_2$) can be assumed to be sufficiently accurate. Furthermore, the interior orientation parameters can be calibrated in air beforehand and assumed to remain constant under water. However, all parameters (interior

orientation, exterior orientation, object point coordinates, interface parameters and refractive indices) can be treated as unknowns if the bundle configuration allows for significant results. All limitations are comparable to other models regarding bundle-invariant interfaces and hence no further practical restrictions arise from our approach. However, some authors (Kotowski 1987; Mulsow 2010; Mulsow and Maas 2014) show parallel estimation of the interior orientation with refractive interfaces in simulations and practical experiments. These are focused on object-invariant cases which allow for observing the interfaces with more depth and hence lower correlations between both parameter sets. Thus, these experiments are not fully comparable to our current state of implementation and practical experiment.

Additionally, we create three observations for the Jacobian (vs. two in standard models). Hence, the Jacobian and the required memory are increased by 50% without a significant increase in computation time. Since the derived normal matrix $\mathbf{N}$ only contains the unknowns, the size remains equal and hence no additional computation time for the solution of the system of equations is needed. Additional matrix multiplications, to obtain the normal matrix, hereby only pose a minor influence.

For a stochastic model, i.e. the weight matrix $\mathbf{P}$, a 2D metric in image space is used for the reprojection error, e.g. the accuracy of semi-major and -minor axes from an ellipse operator. This usually results in a diagonal matrix. In our case, a 3D metric is needed which could be obtained by variance propagation from any given 2D metric. In contrast, the resulting weight matrix would be a 3×3 block diagonal matrix due to non-negligible correlations between the single vector components in Δ_{xyz} . Higher memory requirements and more calculations for matrix multiplications are thus necessary for a strict stochastic model.

Besides the significantly reduced computation time, the approach's main advantage is the description of residuals in object space. This results in a number of simplifications and convenient results. First and foremost, the additional iterations to project a point to the sensor or an interface, necessary for other ray tracing approaches, become obsolete. Admittedly, additional iterations for undistorting an image coordinate have to be performed. However, if no distortion parameters are optimized during adjustment (which is usually the case, as results are highly correlated and calibration is performed beforehand), this has to be performed only once before the first iteration. Nonetheless, image undistortion is computationally far less expensive than additional refraction iterations. Other than that, each projection is a direct function for which analytical derivatives can be calculated. Additionally, the stochastic information, given in σ_0 , is provided in object space. The parameter, therefore, provides a metric for an average coordinate component of an image observation of unit weight, projected into object

space. Stated σ_0 hence reflects the accuracy of the distance from the ray's plumb foot point to the adjusted object point in a single coordinate direction. To obtain an approximation of accuracy for a single point in 3D space, σ_0 values should be multiplied by $\sqrt{3}$. First empirical simulations for different observation accuracies and error-free object points showed that this metric slightly underestimates the actual accuracy by about 10–15% (i.e. $\sigma_0 \cdot \sqrt{3}$ is 10–15% higher than RMS_{xyz} of resulting object coordinates). Still, this gives an easy estimate of the anticipated accuracy, assuming no systematic or gross errors. We will discuss this aspect further in future work. However, for easier comparison, we also project the final results back into image space once after convergence. Hence, we also obtain σ_0 in image space which enables comparison with the standard ray tracing or Brown model. The error function can fundamentally also be transferred to single-media photogrammetry. Advantages of explicit modeling using ray tracing, as in Mulsow and Maas (2014) and Kotowski (1987), are:

- Arbitrary number of interfaces can be modeled (depending on significance of results)
- Arbitrary orientation of camera towards interfaces (limited by resulting image quality)
- Explicit calculation of interface orientation
- Explicit calculation of refractive indices

These advantages over implicit modeling and other camera models are equal to our representation, as the underlying physics are equal. The explicit calculation of the entire imaging geometry decouples the media (glass, water) from the interior orientation and can hence be extracted from the bundle. This is particularly useful in cases where only an a priori calibration in a laboratory can be performed but in-situ calibration (e.g. on the open sea) is hardly possible. In such case, the interface orientation and other calibration parameters can be calibrated beforehand and only an estimate of the refractive index of water has to be obtained through calibration or measurement in situ.

4 Evaluation on Underwater Dataset

Investigation of the model was performed in a water tank of $1.0\text{ m} \times 1.5\text{ m} \times 1.0\text{ m}$ size that was filled with clear fresh water. A single camera observed a calibration artifact that was marked with ring-coded targets. Based on the well-known best practice to employ a 3D test field for air-applications and as investigated by Boutros et al. (2015) for underwater cases, we relied on a cuboidal calibration artifact for optimum calibration results. For an optimum geometric bundle configuration, the artifact was moved and rotated

Fig. 3 Evaluation setup in water tank. A $300\text{ mm} \times 350\text{ mm} \times 480\text{ mm}$ cuboidal calibration artifact with coded ring targets (gray) was moved and rotated around all axes in a water tank. The single camera (red) observed the calibration scene through a flat port and remained static

Fig. 4 Visualization of an exemplary bundle configuration in object space for non-orthogonal left case. Cameras are shown in red, object points are colored in blue

around all axes, resulting in an apparent spherical imaging setup (see Fig. 3).

Figure 4 shows an exemplary bundle which provides very good geometry. All points with less than four intersecting rays and exterior orientations with less than six corresponding points in object space were excluded. Hence, well-determined camera stations were computed and intersection angles on object points were close to the optimum 90° .

The camera (see Table 1) was enclosed in a waterproof housing and observed the scene through a flat port, both

Table 1 Relevant setup parameters

Camera parameters	
Camera	Basler Ace acA2040-25 gm
Focal length	10 mm
Resolution	2048 × 2048 px
Pixel pitch	5.5 × 5.5 μm
Interface parameters	
Interface material	Flat acrylic glass
Interface thickness	8.42 mm
μ_{glass}	1.49
μ_{air}	1.00028

made of acrylic glass. To investigate the influence of the interface orientation w.r.t. the camera's optical axis, the camera is set up orthogonally and rotated about 10° to the left and right, respectively, resulting in three positions. For each position, the calibration procedure is repeated three times.

Table 1 shows the relevant evaluation parameters. The interface parameters were measured, using a caliper and a refractometer. The refractive index of the air inside the camera housing was calculated based on average environmental parameters (temperature, humidity, pressure and wavelength) and the well-known formula by Edlén (1966).

A variable acquisition distance of 300–1200 mm was chosen to enhance distance-dependent distortion effects, introduced by refraction. Hence, in this setup, calibration is performed with unfavorable conditions for implicit Brown modeling. The percentage of water in the ray path from the projection center to the minimum and maximum acquisition distance ranged from 92–98%, a common air/water ratio in applications with ROV.

For analysis, all observations were weighted equally, omitting stochastic modeling and outlier detection to show pure effects of functional modeling. However, initial values were provided by standard photogrammetry software "AICON 3D Studio" and hence, gross errors, e.g. by mixed-up point observations or erroneous ellipse measurements have most likely been eliminated from the data beforehand. All datasets were adjusted, using free network analysis (six Helmert constraints) with three distance constraints for scaling by three a priori-determined distances. The distances were spread over the measurement volume, covering all dimensions, as recommended by Luhmann et al. (2020) to be the most reliable configuration with a moving test field.

For analysis on precision, we focused on the reprojection error and residual patterns in the image to conclude possible systematic errors, visible in the model. For accuracy, we compared 3D coordinates of the bundle under water with pre-determined coordinates by a photogrammetric bundle in air (reference). The underwater dataset was transformed to the coordinates of the reference by a best-fit 6DOF Helmert

transformation with fixed scale. As a result, we compared RMS_{xyz} values and calculated relative accuracy as the ratio of the calibration artifact's diagonal and the RMS_{xyz} value.

In the following, we will discuss the results of the orthogonal setup, followed by the non-orthogonal datasets. Each dataset was evaluated using three different ways of refraction compensation which were integrated into our own bundle implementation:

1. Brown model with simultaneous estimation of all distortion parameters for radial-symmetric distortion, tangential-asymmetric (decentering) distortion and affinity and shear (BRN)
2. Ray tracing (RT), as proposed by Kotowski (1987); implementation according to Mulsow and Maas (2014) with MDFRT. Parameters of interior orientation result from a pre-calibration and are held constant during optimization.
3. Our own ray tracing implementation, referred to as *Ray Tracing with Minimization in Object Space* (RTO). Parameters of interior orientation result from a pre-calibration and are held constant during optimization.

For further analysis, parameters of interior orientation (for the Brown model only) are not considered as this is not the focus of this work. Furthermore, changes mostly follow predictable behavior, such as an increase in camera constant underwater by a factor of approximately the refractive index, as shown, e.g. by Agrafiotis and Georgopoulos (2015). Non-orthogonal setups showed a shift of the principal point and all parameters were determined within the expected precision of image measurement σ_0 .

We assume the results of the standard ray tracing and our method to be very similar, as both employ the same imaging geometry with differences in the optimization procedure. However, especially for speed comparison and sensitivity to noise, it remains of interest to compare both approaches. All configurations were performed three times with equal setups. However, for analysis, we focused on one representative dataset and refer to salient differences, if present.

4.1 Calibration in Air

Right before each of the investigated setups, a calibration in air without any interfaces was performed, to obtain the best accuracy possible with the same cuboidal calibration artifact. We assumed this to be the benchmark for the best possible result we could achieve with our setup due to a favorable bundle configuration and the medium air. Hence, we could evaluate the loss of accuracy under water.

Despite having performed several calibrations, we only show one exemplary result for reference. All other datasets have very similar results and do not show any considerable

differences towards the results shown here. Calibration was performed, using the Brown model with all aforementioned distortion parameters. Figure 5a, e shows an exemplary image and the resulting distortion plot for the calibration. The image shows no prominent distortion and the residual plot shows a typical barrel distortion with vectors pointing outward from the middle and vice versa in the corners.

Figure 7a shows the image residuals for this calibration. It can be observed, that no systematic residuals occur and the entire sensor was captured, as recommended in literature (Luhmann et al. 2020).

Figure 6 shows the residuals of the calibration. Systematic effects cannot be observed and overall accuracy is homogeneous. The overall RMS_{xyz} (1σ) was $30\ \mu\text{m}$ with a maximum residual of $102\ \mu\text{m}$ and the precision, represented by the standard deviation of unit weight, was $\sigma_0 = 0.397\ \mu\text{m}$, equaling $1/14\ \text{px}$. All parameters of interior orientation were computed significantly and no gross errors were identified.

4.2 Orthogonal Orientation

First, we show results from the orthogonally arranged setup. This is the common case for many applications in underwater photogrammetry (e.g. for ROV cameras). The orientation was set by eye and was hence expected to be

Fig. 6 Residual errors in 3D space after best fit transformation onto reference data for in-air calibration

accurate up to 1° . For the results of the Brown model, all distortion and camera parameters were adjusted, whereas the interior orientation for the ray tracing models based on an in-air calibration and parameters were fixed. Hence, interface parameters, refractive index of water, exterior

Fig. 5 Exemplary images (top row) and corresponding distortion plots (bottom row) of different imaging setups. The case in air shows no prominent distortion on the calibration artifact and barrel distortion in the plot. Orthogonally arranged interface **b** shows pincushion

distortion on both sides, slanted perspective from the left shows pincushion distortion mostly on the left side and vice versa on the right hand side. Note: Scale factor for distortion plots of left and right (g, h) is reduced by factor 5 compared to the other plots

Fig. 7 Image residuals for all setups and implicit (top row) and explicit (bottom row) modeling

orientation and object coordinates were part of the adjustment for these models.

An exemplary image and the corresponding distortion plot for the orthogonal underwater setup is depicted in Fig. 5b, f, showing prominent distortion in the image, resulting in bent straight lines towards the image center. The distortion plot verifies this impression by showing pin-cushion distortion and hence basically inverting the vectors from the in-air calibration. Figure 7b, e shows exemplary residual plots for the orthogonal setup with the Brown and the explicit RT/RTO model, reprojected to image space. Distinction between the two explicit models was omitted in this case because of no salient differences, whatsoever. The implicit model shows systematic effects in the image corners with residual vectors pointing inwards. In contrast, this effect is hardly visible in the explicit model. Only few vectors point inwards from the bottom left corner but the overall picture does not show a systematic pattern.

Figure 8a–c shows the residual errors in 3D space of coordinates resulting from calibration and the reference data. It can be observed that a minimal vertical systematic scaling effect occurs in all three models. However, since this effect is present in all datasets and the margin is small enough to result from temperature-induced effects, this will not be discussed any further. Other systematic effects are not present in any model and results cannot be distinguished from the figure. Table 2 on the other hand shows a slight accuracy

improvement of $8 \mu\text{m}$ in the explicit modelings of both RT and RTO. The precision in turn decreases significantly which can be observed from the value of σ_0 at $1/7 \text{ px}$ and is about 50% larger than with the explicit modelings ($1/11 \text{ px}$). The two explicit models show similar values for σ_0 which are slightly higher with our RTO model but result in the same accuracy. The values for the interface and refraction parameters are slightly different. The parameters' standard deviations are nearly identical for both ray tracing approaches and are lower than the deviations between the two models. The biggest difference between the two models is in computation time per iteration. While needing roughly the same amount of iterations, it can be observed that our model decreases computation time by a factor of 10 compared to RT, though still being about 30 times slower than the Brown model per iteration. However, it has to be considered that derivatives for the Brown model were analytic, whereas the explicit models both ran on numeric derivatives.

4.3 Non-orthogonal Orientation

Second, two non-orthogonal setups were investigated. This can occur when multiple cameras are included in the same housing or the desired convergence cannot be achieved by rotating the housings for space reasons. Analysis was equal to the procedure with the orthogonal setup. First, the left setup is shown with the camera rotated about 10° to the left

Fig. 8 Residual errors in 3D space after best-fit transformation onto reference data for all camera setups

Table 2 Results of orthogonal setup with $n_{\text{obs}} = 9112$

	σ_0 [μm]	RMS_{xyz} [mm]	RA [1 : N]	α_z [$^\circ$]	d_2 [mm]	μ_{water}	t_{it} [s]
BRN	0.758	0.063	10582				0.3
RT	0.500	0.055	12188	0.790 (3E–5)	– 42.55 (0.050)	1.3341 (2E–5)	100.8
RTO	0.531 (19.74)	0.055	12109	0.786 (3E–5)	– 41.38 (0.044)	1.3342 (2E–5)	9.5

Bracketed σ_0 value for RTO depicts standard deviation in object space. RA is relative accuracy, referring the ratio of the diagonal distance in the calibration artifact and the RMS_{xyz} value. α_z denotes the angle between the interface normal and the optical axis, d_2 is the distance from the projection center to the outer interface. Bracketed values below depict 1σ standard deviation of estimated parameter. t_{it} represents the time for a single iteration

around the vertical camera axis w.r.t. the interface normal. Afterwards, we show the opposite direction with approximately 8° to the right hand side.

4.3.1 Left

Figure 5c, g shows an exemplary image and the corresponding distortion plot for this setup. The vertical structure on the left hand side is bent toward the inside, whereas on the right hand side, this behavior is not prominently visible. The distortion plot supports this and shows a different pattern than the orthogonal and in-air setup with high amounts of asymmetric distortion. Since the distortion was very high in the image corners, targets could not be measured and hence the total amount of observations was reduced, compared to the orthogonal setup. Residual plots for this setup are depicted in Fig. 7c, f. The left setup shows major systematic residuals and the reduced image coverage in the top and bottom left corner due to high distortions in these areas, disabling ellipse measurement. The residual vectors point radially inwards with only a radial area around the second zero-crossing of the radial distortion function r_0 that shows lower amounts. This radial pattern is only interrupted on the right hand side where vectors show a rather random pattern. Figure 8d–f shows the 3D residuals of analysis for all three investigated models. The most prominent difference can be observed in the systematic residuals of the Brown model. All residuals in the middle of the verticals tend towards the center of mass. This behavior is not observable in both explicit models, whose residuals appear almost identical. Considering Table 3, this impression is supported by the RMS values for accuracy, with the RTO having a slightly higher accuracy than the RT model. However, the precision in σ_0 shows slightly lower values for the RT model at $1/8$ px.

σ_0 for the implicit model is reduced to $1/3$ px. In total, the accuracy is about 50% higher for each of the explicit models than for the implicit Brown model. However, the RMS values for the explicit models are about 13–20% lower in the orthogonal case. The angle α_z differs about 0.02° which is larger than for the orthogonal setup. The difference of distance d_2 in turn is closer than with the orthogonal setup, only differing 0.04 mm. Standard deviations of the parameters show an increase in the RT approach which is still smaller than the deviations between the parameter values shown. Compared to the orthogonal setup, the values are higher, especially for the determination of α_z . Again, the time per iteration for our model is about 10 times faster than the standard ray tracing but also about 30 times slower than the Brown model. The total number of iterations was comparable for all models.

4.3.2 Right

Figure 5d, h shows an exemplary image and the corresponding distortion plot for this setup. Both images show the mirrored behavior from Fig. 5c, g, such as a bent vertical line on the right hand side and the corresponding distortion plot, mostly affecting the right hand side and no standard rotational symmetric distortion pattern. The image residuals in Fig. 7 show a mirrored systematic pattern, compared to the left setup. The magnitude is comparable and coverage, again is significantly lower than for the orthogonal setup in the image corners on the right hand side. Figure 8g–i shows the 3D residuals in object space. The first obvious observation can be made on the Brown model, which again shows higher systematic residuals with a similar pattern to the case on the left side. The explicit modelings again show significantly lower residuals, although a small systematic rotational

Table 3 Results of non-orthogonal left setup with $n_{\text{obs}} = 5364$

	σ_0 [μm]	RMS_{xyz} [mm]	RA [1 : N]	α_z [$^\circ$]	d_2 [mm]	μ_{water}	t_{it} [s]
BRN	1.881	0.134	4972				0.2
RT	0.677	0.068	9743	10.246 (7E–4)	– 36.57 (0.075)	1.3336 (4E–5)	57.1
RTO	0.718 (25.94)	0.063	10628	10.224 (5E–4)	– 36.53 (0.060)	1.3337 (3E–5)	5.4

Depicted parameters are equal to Table 2

Table 4 Results of non-orthogonal right setup with $n_{\text{obs}} = 5020$

	σ_0 [μm]	RMS_{xyz} [mm]	RA [1 : N]	α_z [$^\circ$]	d_2 [mm]	μ_{water}	t_{it} [s]
BRN	1.682	0.117	5692				0.2
RT	0.612	0.069	9576	8.545 (6E–4)	– 36.52 (0.077)	1.3338 (4E–5)	53.9
RTO	0.651 (22.57)	0.069	9697	8.524 (4E–4)	– 36.11 (0.061)	1.3338 (3E–5)	5.2

pattern appears on the upper part of the artifact. In general, again no explicit model appears to present different results, compared to the other. Table 4 also shows no difference in accuracy between these. However, precision is again higher with the RT model, though only by a relatively small margin (about 1/9 px). The Brown model shows lower accuracy (about 70% higher RMS value) and precision (about 175% higher σ_0 value of about 1/3 px). The angle α_z , like in the case on the left side, differs about 0.02° . The distance d_2 differs about 0.4mm, which is higher than in the left case but lower than in the orthogonal case. Standard deviations show nearly the pattern as with the left setup. The highest increase in standard deviation can again be observed for α_z . Again, the time per iteration for our model is about 10 times faster than the standard ray tracing but also about 30 times slower than the Brown model.

4.4 Discussion

Results show that our model is capable of producing comparable results to the standard ray tracing procedure. Accuracy in object space is almost equal in all setups and 3D residual patterns show no salient difference to the RT approach. However, differences arise in precision, interface and refraction parameters. The precision value σ_0 tends to be higher for our RTO model than for the RT model, when comparing reprojected residuals. This was not expected since the imaging geometry is equal in both approaches. No complete explanation can be given for this behavior but it seems likely that the error function in object space is less sensible to small changes in image observations than the ones in image space. This is hence a numeric issue, which has not shown a significant influence on the results in object space in our investigation. In fact, in the left dataset, our RTO approach achieved better results than the RT did. However, despite showing a similar tendency in some of our other (not presented) datasets, it is likely that this behavior is rather a coincidence and not generally applicable. High correlations between the unknown parameters, especially with exterior orientation and among underwater parameters could have been another reason. However, during evaluation, no crucially high correlations ($|r| \geq 0.85$) were detected between these parameters for any setup. Considering the refractive parameters' standard deviations, nearly all differences between both ray tracing approaches are significant. However, several sets for interface and refraction parameters (RT

and RTO) lead to similar results in object space, possibly due to minor correlations. This is particularly of interest when calibration parameters are extracted from the bundle (e.g. interface orientation and position) and set up in a different environment with different refractive properties.

The standard deviations of the parameters are increased in the non-orthogonal setups compared to the orthogonal setup. This complies with the overall increased precision and accuracy values which are about 30% higher with non-orthogonal setups. However, the standard deviation for α_z is increased by more than factor ten whereas precision of the refractive index of water is only increased by about factor two compared to the orthogonal setup. Taking the overall lower precision of the non-orthogonal setup into account, the refractive index can hence be determined better when the interface is observed with depth. This complicates simultaneous calibration of interior orientation for the achievable depth with our bundle-invariant setup.

Timing behavior is approximately linear with the number of observations for all models. Also, the proportions between the timing of the three models remains almost equal with the Brown model being about 30 times faster than the RTO model, which in turn is about ten times faster than the standard RT model. In particular, large SfM bundles benefit from this efficient model and should be solvable with reasonable time consumption on a capable machine.

The results show a clear tendency towards an accuracy loss without explicit modeling, even for the orthogonal interface right in front of the camera. This finding in the orthogonal case contradicts other authors, such as Kotowski (1988), Shortis (2015) or Kahmen et al. (2020b). This may be caused by the variable acquisition distance, which enforces distance-dependent distortions even more. However, whereas the orthogonal setup obtained almost comparable results in object space with an accuracy loss of about 10%, computation time is significantly lower than with the explicit models and is hence to be considered when choosing the camera model, depending on the specific demands. However, our model still runs on numeric derivatives, which is yet to be improved and should reduce computation time at least by factor 5.

Overall accuracy is approximately reduced by a factor of two compared to calibration in air for the orthogonal setup. With non-orthogonal setups, accuracy decreases further, although explicit models have a reduced accuracy of a maximum of 20%, as well. This may partly be caused by

the lower image coverage in the corners and hence no optimum calibration setup. However, we were unable to improve parameters with the software for ellipse measurement to obtain targets in these areas.

The σ_0 of RTO states the distance of a single coordinate component from the image ray projected to object space. Hence, $\sigma_0 \cdot \sqrt{3}$ reflects accuracy of the distance from the ray's plumb foot point to the adjusted object point. However, with this factor applied, values are still smaller than the respective RMS_{xyz} values. This is likely caused by two factors: For one, the accuracy metric assumes an error-free plumb foot point, which in practice is equally affected by errors and hence introduces additional noise to the system. Second, finite accuracy at reference measurement and temperature variance due to submersion in water create further uncertainty.

The general precision and accuracy is in accordance with other literature, summarized by Shortis (2015, 2019). Here, a general precision of 1/10—1/3 px is stated for implicit models. Unfortunately, only two publications with explicit modeling are stated that rely on natural points. Hence, stated results of 1 and 1/3 px are not quite comparable to our findings. However, Shortis (2015) and Maas (2015) state a relative accuracy loss, compared to in-air-calibration, by a factor of 3–10, which reflects our findings, particularly for the orthogonal case, quite well. Relative accuracy of about 1:10,000 with explicit modeling or orthogonal setups is again on the same order of magnitude as the results from Shortis (2015). Hemmati and Rashtchi (2019) showed an accuracy loss from a orthogonal setup towards a non-orthogonal setup under water of about 58% with implicit modeling. Despite a smaller margin, the tendency shown in our findings reflects the same behavior, mostly affected by reduced imaging quality. Furthermore, ellipse eccentricity may have a significantly higher impact due to large distortions occurring, especially in the corner of the images. Hence, compensation for eccentricity, as investigated in Luhmann (2014) for spatial intersection in air, may be necessary. This should already be integrated at the processing stage of ellipse measurement, e.g. by image correction, to enable more targets to be measured and hence higher image coverage and improved accuracy for both implicit and explicit calibrations. Alternatively, stochastic modeling with a radial weight function towards the image corners can be performed, similar to Menna et al. (2020).

Implicit Brown modeling shows a major accuracy loss and should be avoided when employing a non-orthogonal setup. It can be concluded that the non-orthogonal interface produces distortions which cannot be fully modeled by tangential and asymmetric distortion and hence need explicit modeling or other specially designed non-standard functions. In the orthogonal case, this can mostly be compensated by radial distortion, as the distance to the interface

is almost constant for all areas of the image. With non-orthogonal orientation, this is no longer the case and the left side of the lens faces more air than the right part, creating complex distortion patterns. This can be observed in the residual plots, as well. Systematic residuals occur in all implicit cases with smallest margins in the orthogonal setup. This is in accordance with results from other publications, e.g. Menna et al. (2017).

5 Conclusion

In this contribution, an alternative, novel—yet physically strict—optimization procedure for ray tracing in underwater and multimedia photogrammetry was presented. We compared outcomes with state-of-the-art-procedures in both theory and a practical experiment. Furthermore, several camera setups for standard applications were investigated and evaluated towards the need for explicit models in these settings.

Results clearly show that our model is capable of providing equal results with only one tenth of time consumption, compared to the standard ray tracing approach by Mulsow and Maas (2014). Contrary to other algorithms, our approach is fundamentally capable of analytic derivatives, reducing a significant amount of computational cost. Additionally, we were able to show a significant accuracy loss for an incidence angle of less than 1° on the interface, with implicit calibration. Further deviation from orthogonality results in a major accuracy loss and disabled ellipse measurement in corners due to high distortion. With the error function, all kinds of shapes and camera models are fundamentally possible to be implemented and is hence not limited to the flat port underwater case. Moreover, all shapes and even single-media or completely different imaging geometries can be implemented with little effort.

The investigated setup was very favorable for a photogrammetric bundle, which in practice is not always the case. It remains of interest to observe differences for planar test fields, worse image quality or an unfavorable configuration of camera stations in the future that should increase correlations and other unwanted effects. So far, we could not observe any significantly different behavior of the RTO model, compared to the RT with other bundles but in-depth analysis for other datasets is yet to be performed.

In this study and our model, we focused on the modeling itself, omitting the correspondence problem that is especially present with natural features rather than coded targets. We assume our model to still be as performant as our study shows. However, as we see certain numeric discrepancies between both models, it would be of interest to investigate the model's behavior with natural feature points. The depicted speed improvement should enable calculating

such, generally very large, bundles that usually arise from SfM data.

Further work will focus on the influence, the wide acquisition distance has on the outcome for implicit models. Simulations in Kahmen et al. (2020a) show a correlation of accuracy with deviation from calibration distance and should be further analyzed for standard applications with real imagery (e.g. ROV, ultra-close range, etc.). Also, the intersection angle of the optical axis with the interface normal should be further investigated to obtain limits of deviation from orthogonality for both Brown and explicit models. The model is furthermore to be expanded towards a hemispherical dome port to account for residual errors, resulting from imperfect positioning.

Our results showed discrepancies between single parameters with similar results in object space. This may be an issue, when calibration is performed in a lab before actual measurements take place since correlations of extracted parameters, such as the interface orientation, lose the relation to the exterior orientation. It is to be investigated how this behaves in practical measurements and the actual influence is to be determined.

Funding This work was funded by Volkswagen Foundation (ZN3253). Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal S, Mierle K, Others (2021) Ceres solver. <http://ceres-solver.org>
- Agrafiotis P, Georgopoulos A (2015) Camera constant in the case of two media photogrammetry. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XL-5/W5:1–6*. <https://doi.org/10.5194/isprs-archives-XL-5-W5-1-2015>
- Agrawal A, Ramalingam S, Taguchi Y, Chari V (2012) A theory of multi-layer flat refractive geometry. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, pp 3346–3353. <https://doi.org/10.1109/CVPR.2012.6248073>
- Boutros N, Shortis MR, Harvey ES (2015) A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. *Limnol Oceanogr Methods* 13(5):224–236. <https://doi.org/10.1002/lom3.10020>
- Brown DC (1971) Close-range camera calibration. *Photogramm Eng* 37(8):855–866
- Chadebecq F, Vasconcelos F, Dwyer G, Lacher R, Ourselin S, Vercauteren T, Stoyanov D (2017) Refractive structure-from-motion through a flat refractive interface. In: *2017 IEEE international conference on computer vision (ICCV)*, IEEE, pp 5325–5333. <https://doi.org/10.1109/ICCV.2017.568>
- Costa E (2019) The progress of survey techniques in underwater sites: the case study of cape stoba shipwreck. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLII-2/W10:69–75*. <https://doi.org/10.5194/isprs-archives-XLII-2-W10-69-2019>
- Dolereit T (2018) A virtual object point model for the calibration of underwater stereo cameras to recover accurate 3D information. *Universität Rostock* https://doi.org/10.18453/ROSDOK_ID00002323
- Drap P, Merad D, Hijazi B, Gaoua L, Nawaf M, Saccone M, Chemisky B, Seinturier J, Sourisseau JC, Gambin T, Castro F (2015) Underwater photogrammetry and object modeling: a case study of Xlendi wreck in Malta. *Sensors* 15(12):30351–30384. <https://doi.org/10.3390/s151229802>
- Duda A, Gaudig C (2016) Refractive forward projection for underwater flat port cameras. In: *IROS 2016, IEEE, Piscataway, NJ*, pp 2022–2027. <https://doi.org/10.1109/IROS.2016.7759318>
- Edlén B (1966) The refractive index of air. *Metrologia* 2(2):71–80. <https://doi.org/10.1088/0026-1394/2/2/002>
- Ekkel T, Schmik J, Luhmann T, Hastedt H (2015) Precise laser-based optical 3D measurement of welding seams under water. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XL-5/W5:117–122*. <https://doi.org/10.5194/isprs-archives-XL-5-W5-117-2015>
- Elnashef B, Filin S (2019) Direct linear and refraction-invariant pose estimation and calibration model for underwater imaging. *ISPRS J Photogramm Remote Sens* 154:259–271. <https://doi.org/10.1016/j.isprsjprs.2019.06.004>
- Glassner AS (1989) Surface physics for ray tracing. In: *Glassner AS (ed) An introduction to ray tracing*. Acad. Press, London, pp 121–160
- Hastedt H, Luhmann T, Przybilla HJ, Rofallski R (2021) Evaluation of interior orientation modelling for cameras with aspheric lenses and image pre-processing with special emphasis to SFM reconstruction. *Int Arch Photogramm Remote Sens Spatial Inf Sci XLIII-B2–2021:17–24*. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-17-2021>
- Hemmati M, Rashtchi V (2019) A study on the refractive effect of glass in vision systems. In: *2019 4th international conference on pattern recognition and image analysis (IPRIA)*, IEEE, pp 19–25. <https://doi.org/10.1109/IPRIA.2019.8785995>
- Henrion S, Spoor CW, Pieters RPM, Müller UK, van Leeuwen JL (2015) Refraction corrected calibration for aquatic locomotion research: application of Snell's law improves spatial accuracy. *Bioinspir Biomim* 10(4):046009. <https://doi.org/10.1088/1748-3190/10/4/046009>
- Jordt-Sedlazeck A, Koch R (2012) Refractive calibration of underwater cameras. In: *Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C (eds) Computer vision—ECCV 2012, vol 7576. lecture notes in computer science*. Springer, Berlin, pp 846–859. https://doi.org/10.1007/978-3-642-33715-4_61
- Kahmen O, Rofallski R, Conen N, Luhmann T (2019) On scale definition within calibration of multi-camera systems in multimedia photogrammetry. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLII-2/W10:93–100*. <https://doi.org/10.5194/isprs-archives-XLII-2-W10-93-2019>

- Kahmen O, Haase N, Luhmann T (2020a) Orientation of point clouds for complex surfaces in medical surgery using trinocular visual odometry and stereo ORB-SLAM2. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLIII-B2-2020*:35–42. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-35-2020>
- Kahmen O, Rofallski R, Luhmann T (2020b) Impact of stereo camera calibration to object accuracy in multimedia photogrammetry. *Remote Sens* 12(12):2057–2087. <https://doi.org/10.3390/rs12122057>
- Kang L, Wu L, Yang YH (2012) Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the SVP camera model. *Appl Opt* 51(31):7591–7603. <https://doi.org/10.1364/AO.51.007591>
- Kotowski R (1987) Zur Berücksichtigung lichtbrechender Flächen im Strahlenbündel: Zugl.: Bonn, Univ., Diss., (1986) Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften Reihe C, vol 330. Beck, München
- Kotowski R (1988) Phototriangulation in multi-media photogrammetry. *Int Arch Photogramm Remote Sens XXVII*:324–334
- Kunz C, Singh H (2008) Hemispherical refraction and camera calibration in underwater vision. *Oceans* 2008:1–7. <https://doi.org/10.1109/OCEANS.2008.5151967>
- Lavest JM, Rives G, Lapresté JT (2000) Underwater camera calibration. In: Vernon D (ed) *Computer vision—ECCV 2000*. Springer, Berlin, pp 654–668. https://doi.org/10.1007/3-540-45053-x_42
- Lavest JM, Rives G, Lapresté JT (2003) Dry camera calibration for underwater applications. *Mach Vis Appl* 13(5–6):245–253. <https://doi.org/10.1007/s00138-002-0112-z>
- Łuczynski T, Pflingsthorst M, Birk A (2017) The Pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings. *Ocean Eng* 133:9–22. <https://doi.org/10.1016/j.oceaneng.2017.01.029>
- Luhmann T (2014) Eccentricity in images of circular and spherical targets and its impact on spatial intersection. *Photogram Rec* 29(148):417–433. <https://doi.org/10.1111/phor.12084>
- Luhmann T, Robson S, Kyle S, Boehm J (2020) *Close-range photogrammetry and 3D imaging*, 3rd edn. De Gruyter, Berlin–Boston. <https://doi.org/10.1515/9783110607253>
- Maas HG (1992) *Digitale Photogrammetrie in der dreidimensionalen Strömungsmesstechnik*. ETH Zürich—Dissertation Nr. 9665
- Maas HG (1995) New developments in multimedia photogrammetry. In: Grün A, Kahmen H (eds) *Optical 3D measurement techniques III*. Wichmann, Karlsruhe, pp 362–372
- Maas HG (2015) A modular geometric model for underwater photogrammetry. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XL-5/W5*:139–141. <https://doi.org/10.5194/isprsarchives-XL-5-W5-139-2015>
- Maas HG, Putze T, Westfeld P (2009) Recent developments in 3D-PTV and tomo-PIV. In: Hirschel EH, Dobriloff C, Fujii K, Haase W, Leer B, Leschziner MA, Nitsche W, Pandolfi M, Periaux J, Rizzi A, Roux B (eds) *Imaging measurement methods for flow analysis, notes on numerical fluid mechanics and multi-disciplinary design*. Springer, Berlin, pp 53–62. https://doi.org/10.1007/978-3-642-01106-1_6
- Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11(2):431–441. <https://doi.org/10.1137/0111030>
- Menna F, Nocerino E, Fassi F, Remondino F (2016) Geometric and optic characterization of a hemispherical dome port for underwater photogrammetry. *Sensors (Basel, Switzerland)* 16(1):48–68. <https://doi.org/10.3390/s16010048>
- Menna F, Nocerino E, Remondino F (2017) Flat versus hemispherical dome ports in underwater photogrammetry. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLII-2/W3*:481–487. <https://doi.org/10.5194/isprs-archives-XLII-2-W3-481-2017>
- Menna F, Nocerino E, Ural S, Gruen A (2020) Mitigating image residuals systematic patterns in underwater photogrammetry. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLIII-B2-2020*:977–984. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-977-2020>
- Mulrow C (2010) A flexible multi-media bundle approach. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XXXVIII/5*:472–477
- Mulrow C (2016) Ein universeller ansatz zur mehrmedien-bündel-triangulation. In: Luhmann T, Schumacher C (eds) *Photogrammetrie—Laserscanning—optische 3D-Messtechnik*. Wichmann, Berlin and Offenbach, pp 266–273
- Mulrow C, Maas HG (2014) A universal approach for geometric modelling in underwater stereo image processing. In: 2014 ICPR workshop on computer vision for analysis of underwater imagery (CVAUI), pp 49–56. <https://doi.org/10.1109/CVAUI.2014.14>
- Nocerino E, Menna F, Gruen F (2021) Bundle adjustment with polynomial point-to-camera distance dependent corrections for underwater photogrammetry. *Int Arch Photogramm Remote Sens Spatial Inf Sci XLIII-B2-2021*:673–679. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-673-2021>
- Prado E, Gómez-Ballesteros M, Cobo A, Sánchez F, Rodríguez-Basalo A, Arrese B, Rodríguez-Cobo L (2019) 3D modeling of Rio Miera wreck ship merging optical and multibeam high resolution points cloud. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLII-2/W10*:159–165. <https://doi.org/10.5194/isprs-archives-XLII-2-W10-159-2019>
- Przybilla HJ, Kotowski R, Meid A, Weber B (1990) Geometrical quality control in nuclear power stations: an application of high precision underwater photogrammetry. *Photogramm Rec* 13(76):577–589. <https://doi.org/10.1111/j.1477-9730.1990.tb00718.x>
- Rofallski R, Tholen C, Helmholz P, Parnum I, Luhmann T (2020a) Measuring artificial reefs using a multi-camera-system for unmanned underwater vehicles. *ISPRS Int Arch Photogramm Remote Sens Spatial Inf Sci XLIII-B2-2020*:999–1008. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-999-2020>
- Rofallski R, Westfeld P, Nistad JG, Büttner A, Luhmann T (2020b) Fusing ROV-based photogrammetric underwater imagery with multibeam soundings for reconstructing wrecks in turbid waters. *J Appl Hydrogr* 116:23–31. <https://doi.org/10.23784/HN116-03>
- Rossi P, Ponti M, Righi S, Castagnetti C, Simonini R, Mancini F, Agraftiotis P, Bassani L, Bruno F, Cerrano C, Cignoni P, Corsini M, Drap P, Dubbini M, Garrabou J, Gori A, Gracías N, Ledoux JB, Linares C, Mantas TP, Menna F, Nocerino E, Palma M, Pavoni G, Ridolfi A, Rossi S, Skarlatos D, Treibitz T, Turicchia E, Yuval M, Capra A (2021) Needs and gaps in optical underwater technologies and methods for the investigation of marine animal forest 3D-structural complexity. *Front Mar Sci* 8:171–179. <https://doi.org/10.3389/fmars.2021.591292>
- Schinstock DE, Lewis C, Buckley C (2009) An alternative cost function to bundle adjustment used for aerial photography from UAVs. In: *American Society for Photogrammetry and Remote Sensing (ed) ASPRS annual conference 2009*, Baltimore, MD, Curran Associates, Red Hook, NY. <https://www.asprs.org/publications/proceedings/baltimore09/0110.pdf>
- Sedlazeck A, Koch R (2012) Perspective and non-perspective camera models in underwater imaging—overview and error analysis. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, Dellaert F, Frahm JM, Pollefeys M, Leal-Taixé L, Rosenhahn B (eds) *Outdoor and large-scale real-world scene analysis, lecture notes in computer science*, vol 7474. Springer, Berlin, pp 212–242. https://doi.org/10.1007/978-3-642-34091-8_10

- She M, Song Y, Mohrmann J, Köser K (2019) Adjustment and calibration of dome port camera systems for underwater vision. In: Fink GA, Frintrop S, Jiang X (eds) Pattern recognition, image processing, computer vision, pattern recognition, and graphics, vol 11824. Springer, Cham, pp 79–92. https://doi.org/10.1007/978-3-030-33676-9_6
- Shortis M (2015) Calibration techniques for accurate measurements by underwater camera systems. *Sensors* (Basel, Switzerland) 15(12):30810–30826. <https://doi.org/10.3390/s151229831>
- Shortis M (2019) Camera calibration techniques for accurate measurement underwater. In: Benjamin J, van Duivenvoorde W (eds) 3D recording and interpretation for maritime archaeology, Coastal Research Library, vol 31. Springer International Publishing, Cham, pp 11–27. https://doi.org/10.1007/978-3-030-03635-5_2
- Telem G, Filin S (2010) Photogrammetric modeling of underwater environments. *ISPRS J Photogramm Remote Sens* 65(5):433–444. <https://doi.org/10.1016/j.isprsjprs.2010.05.004>
- Treibitz T, Schechner Y, Kunz C, Singh H (2012) Flat refractive geometry. *IEEE Trans Pattern Anal Mach Intell* 34(1):51–65. <https://doi.org/10.1109/TPAMI.2011.105>
- Wolf PR, Dewitt BA, Wilkinson BE (2014) Elements of photogrammetry with applications in GIS, 4th edn. McGraw-Hill Education, New York. <https://doi.org/10.5555/556015>
- Yau T, Gong M, Yang YH (2013) Underwater camera calibration using wavelength triangulation. In: IEEE conference on computer vision and pattern recognition (CVPR), 2013, IEEE, Piscataway, NJ, pp 2499–2506. <https://doi.org/10.1109/CVPR.2013.323>