Refractive Calibration of Underwater Cameras

Anne Jordt-Sedlazeck* and Reinhard Koch

Department of Computer Science, Kiel University
Hermann-Rodewald-Str. 3, 24118 Kiel, Germany
{as,rk}@mip.informatik.uni-kiel.de
http://www.mip.informatik.uni-kiel.de

Abstract. In underwater computer vision, images are influenced by the
water in two different ways. First, while still traveling through the water,
light is absorbed and scattered, both of which are wavelength dependent,
thus create the typical green or blue hue in underwater images. Secondly,
when entering the underwater housing, the rays are refracted, affecting
image formation geometrically. When using underwater images in for ex-
ample Structure-from-Motion applications, both effects need to be taken
into account. Therefore, we present a novel method for calibrating the
parameters of an underwater camera housing. An evolutionary optimiza-
tion algorithm is coupled with an analysis-by-synthesis approach, which
allows to calibrate the parameters of a light propagation model for the
local water body. This leads to a highly accurate calibration method for
camera-glass distance and glass normal with respect to the optical axis.
In addition, a model for the distance dependent effect of water on light
propagation is parametrized and can be used for color correction.

1 Introduction

Research and exploration, oil production, rescue missions, and pipeline inspec-
tions are all possible applications of cameras in underwater housings, where
geometric information can be used in for example navigation, reconstruction,
or mosaicing algorithms. In order for these algorithms to reach the best pos-
sible accuracy, refraction at the underwater housings of the cameras needs to
be taken into account. This work describes a system for calibrating the un-
derwater housings of flat port, deep sea cameras with glass ports that can be
several centimeters thick. In order to be feasible in deep sea environments, the
approach must not require 3D calibration targets or even structured light, active
calibration targets, etc. We only use a 2D checkerboard pattern for calibration
that needs to be captured from different, but unknown camera poses. In case
of rigidly coupled multi-camera rigs, checkerboard images are recorded using all
cameras simultaneously and relative transformations between the cameras of the
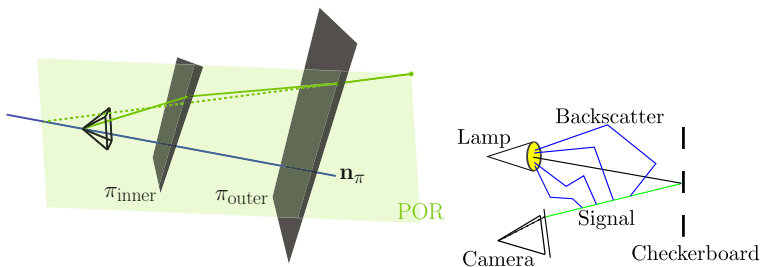rig are calibrated.

Fig. 1. Left: Refraction bends light rays at air-glass and glass-water interfaces prior to reaching the camera (solid green line). If ignoring the effect (dashed line), the rays do not meet in a common center of projection, but intersect an axis defined by the camera center and the interface normal (blue), hence the single-view-point assumption is invalid. All ray segments together with the normal $\mathbf{n}_\pi$ lie in a common plane, the plane of refraction (POR) [1]. Right: the light traveling towards the camera after re-flection at an object (green) is called the signal and is attenuated depending on the distance traveled. Due to multiple scattering events, a veiling light exists within the water (blue) that is scattered into the ray traveling towards the camera.

In case of flat-port underwater housings, the single-view-point (SVP) assumption for perspective cameras is invalid, as can be observed in figure 1 on the left when following the dashed line, i.e. ignoring refraction. Due to misalignment of the camera within the housing, the normal of the glass interface usually does not coincide with the optical axis. Agrawal et al. showed in [1] that underwater cameras in flat port housings are in fact axial cameras, i.e. all rays intersect a common axis, which is defined by the camera's center and the glass normal. Thus, the housing is parametrized by this normal and the distance between camera center and glass interface.

Figure 1 on the right shows how light is affected by water while still traveling through the water body. Depending on the light's wavelength, it is absorbed and scattered, thus the colors measured in the image are affected depending on the distance traveled through the water. Therefore, the proposed method additionally takes into account the effects on the scene colors, which create the typical green or blue hue. Based on the checkerboard images with known white and black areas, good calibration data for the water colors exists. After geometric calibration, all camera poses, i.e. all distances between checkerboard and camera are known, thus automatically calibrating water color parameters is very opportune at this point. The resulting radiometric parameters can be utilized for color correction in the texture images of the 3D-models as in [2].

In the following, our approach will be set into context with related work. Then, the models for underwater light propagation and refraction at the underwater housing will be introduced, followed by a description of the applied calibration approach. After that, experiments, conclusion, and outlook will conclude the article.

2   Related Work

Camera calibration has been an important topic in the computer vision community in the last decades. Especially for the perspective camera, a lot of different methods exist, for example [3], [4], or [5]. Even though the SVP-assumption is invalid for cameras in underwater housings, in the literature often one of these methods is used with the perspective camera model on underwater images ([6], [7]). Freyer and Fraser in [8] and Lavest et al. in [9] describe how focal length and radial distortion absorb the model error and how approximate, perspective calibrations for the underwater case can be computed based upon perspective calibrations using images captured in air.

However, due to the non-SVP nature of perspective cameras within underwater housings, these calibrations always have a systematic model error and measurements based on these calibrations tend to suffer from inaccuracies. Basically, three possibilities exist to solve the problem. The first one is to use a general camera model, i.e. calibrate a ray with starting point and direction for each pixel ([10], [11]). Secondly, a less general axial camera model [12] can be used. However, in case of cameras in underwater housings, a concise parametrization can be achieved, which leads to the third possibility, where refraction is incorporated explicitly by modeling the camera housing. Existing works are Treibitz et al. [13] assuming thin glass and interface-sensor parallelism. In the work of Telem et al. [14], 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. [15] assumes a camera looking onto a water surface, thus there is no glass involved. Kunz et al. [16] describe models for hemispherical and flat ports and use synthetic data to experiment with inaccuracies in Structure-from-Motion (SfM) based on the perspective model. The calibration approach is described, but not implemented and therefore not tested. However, in contrast to both other methods described above, the camera's image sensor and glass interface are not assumed to be parallel. Li et al. [17] develop a complete physical model and its calibration for a stereo rig using triangulated rays in water. The indices of refraction are assumed to be known. [18] describe a system for stereo rigs without having to use a calibration target. The most recent approach proposed by Agrawal et al. [1] showed that flat port underwater cameras are in fact axial cameras and the authors use this insight to develop a calibration method based on checkerboard corners. We use their method for initialization and the **contribution of our method** is to add an analysis-by-synthesis (AS) approach for geometric calibration, i.e. a rendered image of the checkerboard pattern is compared to the real one and the difference is minimized, thus eliminating errors introduced by inaccuracies in corner detection. AS has been successfully applied to perspective calibration beforehand [5], [19], a disadvantage is the increased run-time. After calibrating the geometric properties of the camera in its underwater housing, the parameters for a physical, distance dependent color model describing underwater light propagation are estimated. Thus effects of the local water body on the image colors [20], [2] can be removed. Our method works for both, monocular cameras and rigidly coupled multi-camera rigs. The proposed

method is more general than [13], less complicated when it comes to capturing images than [21] and [11], and works on both camera rigs and monocular cameras as opposed to [18] and [21], which require stereo rigs.

Even in established calibration methods [5], it is well known that an ambiguity between parameters may exist and can cause the optimization process to converge towards a local minimum in the error function. Therefore, instead of using local, derivative-based optimization, we opt for using CMA-ES, short for Covariance Matrix Adaptation Evolution-Strategy [22]. CMA-ES is a stochastic procedure that performs well on non-linear, non-convex error functions with noisy observations and has been used successfully for optimization in computer vision applications before in [23] or [24], even in calibration routines [25]. The next section will introduce the camera model used for calibration.

3   Physical Model and Optimization Strategy

In our method, a camera is described by its pose, parametrized by 3 parameters for translation and 3 for rotation (Euler angles). Inside the camera housing, we assume a calibrated, perspective camera with known focal length f, principal point (c_x, c_y), and two parameters each for radial and tangential lens distortion (r_1, r_2, t_1, t_2).

In the optimization process, there will be M cameras within a static rig, denoted by the index j. Note, that the monocular case is included here with $M = 1$. In general, the first camera is considered to be the master camera, determining the pose of the whole rig. The other cameras (slave cameras) are described by their relative poses with respect to the master camera. The rig will be used to capture a set of images from N different poses, denoted with the index i. A pixel in an image will be indexed by k, the overall number of pixels in an image is K. For the corner-based initialization steps, the detected 2D-corners of the checkerboard in the images will be denoted by $l < L$. Parameters include:

1. Geometric calibration:
   - N poses of the master camera in the world coordinate system ($6N$),
   - poses of slave cameras in the master camera coordinate system ($6(M-1)$),
   - interface distance and interface normals of all cameras in the rig ($3M$),
   - 2 parameters for each image determining gray-level offset and white balance ($2NM$).
2. Radiometric calibration:
   - 2 parameters for water color correction for each color channel (6), and
   - 2 parameters for each color channel and image parametrizing for gray-level offset and white balance ($3MN$).

In total, $6N + 6(M-1) + 3M + 2MN$ parameters must be estimated during geometric calibration and $6 + 3MN$ during radiometric calibration. In the following two sections, the effects of water on image formation will be modeled.

3.1   Radiometric Model

While still traveling through the water, the light is attenuated due to scattering and absorption (see also fig. 1 on the right). Part of the light is reflected by an object L_{obj}, here the checkerboard pattern, and directly travels toward the camera, denoted signal S. Depending on the distance z between checkerboard and camera, the light is attenuated [20], [2]:

$$S = L_{\text{obj}}\text{e}^{-\eta_\lambda z}, \tag{1}$$

where η_λ is the attenuation coefficient, a descriptor of the local water body that is wavelength dependent: $\lambda \in \{R, G, B\}$. R, G, B denote the three discretized color channels. Another proportion of the light is scattered at water molecules multiple times, thus creating a veiling light that exists within the water, colored by B_{∞_λ}. From the photons belonging to this group, by chance, some are scattered into the beam traveling from the checkerboard pattern towards the camera and add to the irradiation being incident upon a pixel of the image sensor, denoted backscatter B:

$$B = B_{\infty_\lambda}(1 - \text{e}^{-\eta_\lambda z}) \tag{2}$$

Hence, the backscatter proportion measured by the image sensor increases with increasing distance z between object and camera. The model for color in water used here is [20]:

$$I_\lambda = \alpha_{\lambda_{ij}}(S+B) + \beta_{\lambda_{ij}} = \alpha_{\lambda_{ij}}(L_{\text{obj}}\text{e}^{-\eta_\lambda z} + B_{\infty_\lambda}(1 - \text{e}^{-\eta_\lambda z})) + \beta_{\lambda_{ij}}, \quad \lambda \in R, G, B \tag{3}$$

where $\alpha_{\lambda_{ij}}$ and $\beta_{\lambda_{ij}}$ account for white balancing and offset on each of the color channels in image ij. Equation 3 is used for estimation of $\alpha_{\lambda_{ij}}, \beta_{\lambda_{ij}}, \eta_\lambda$, and B_{∞_λ}.

3.2   Geometric Model

When entering the underwater housing, the light is refracted due to entering a medium with different optical density. Therefore, depending on the incidence angle, light rays change their direction to some degree. We use the same model for the underwater housing and refraction computation as described in [18]. Refraction is computed explicitly (fig. 2) by intersecting the ray from the camera center through the 2D point with inner and outer interface plane and applying Snell's law at both intersections. If $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ are the normalized rays before and after refraction at medium interface i with refractive indices μ_i and μ_{i+1}, then [26]:

$$\mathbf{r}_{i+1} = \frac{\mu_i}{\mu_{i+1}}\mathbf{r}_i + \left(-\mathbf{r_i}\mu_{i+1}\mathbf{r}_i^{\text{T}}\mathbf{n}_\pi + \sqrt{1 - \frac{\mu_i}{\mu_{i+1}}\left(1 - (\mathbf{r}_i^{\text{T}}\mathbf{n}_\pi)^2\right)}\right)\mathbf{n}_\pi, \tag{4}$$

allowing to compute the rays in glass $\mathbf{r}_{\text{g}}$ and water $\mathbf{r}_{\text{w}}$ using the ray in air $\mathbf{r}_{\text{a}}$. For the calibration method described here, it is assumed that the interface thickness d_{g} is known as well as the indices of refraction for air, glass, and water ($\mu_{\text{a}}, \mu_{\text{g}}, \mu_{\text{w}}$
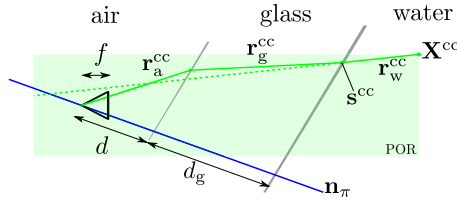
Fig. 2. Left: depicted is the POR from fig. 1. When back-projecting an image point into the scene, the ray $\mathbf{r}_a^{cc}$ runs from the camera through air intersecting the inner interface plane. The refracted ray in glass $\mathbf{r}_g^{cc}$ intersects the outer interface plane at $\mathbf{s}^{cc}$, is refracted again, yielding the ray in water $\mathbf{r}_w^{cc}$ on which the 3D point $\mathbf{X}^{cc}$ lies. Dashed line: classic, perspective projection.

respectively). The distance between camera center and glass d and the normal of the glass $\mathbf{n}_\pi$ need to be calibrated. Depending on d, d_g, $\mathbf{n}_\pi$, μ_a, μ_g, and μ_w, each pixel in the image $\mathbf{x}_{ijk} = (x_{ijk}, y_{ijk}), i < N, j < M, k < K$ can be transformed (eq. (4)) into a ray defined by a starting point $\mathbf{s}_{ijk}^{cc} = \frac{d}{\mathbf{r}_a^T \mathbf{n}_\pi}\mathbf{r}_a + \frac{d_g}{\mathbf{r}_g^T \mathbf{n}_\pi}\mathbf{r}_g$ and a direction $\mathbf{r}_{w_{ijk}}^{cc}$ in the local camera coordinate system, denoted by cc:

$$(\mathbf{s}_{ijk}^{cc}, \mathbf{r}_{w_{ijk}}^{cc}) = \text{Ray}(\mathbf{x}_{ijk}, d_j, \mathbf{n}_{\pi_j}). \tag{5}$$

In favor of readability, the transform from the relative pose of the slave cameras $j > 1$ to the absolute pose using the pose of the master camera is omitted, thus all following rotations and translations are already the absolute pose. Using the camera's rotation matrix $\mathbf{R}_{ij}$ and camera translation $\mathbf{C}_{ij}$, the ray can be transformed into the world coordinate system, denoted by wc:

$$(\mathbf{s}_{ijk}^{wc}, \mathbf{r}_{w_{ijk}}^{wc}) = (\mathbf{R}_{ij}\mathbf{s}_{ijk}^{cc} + \mathbf{C}_{ij}, \mathbf{R}_{ij}\mathbf{r}_{w_{ijk}}^{cc}). \tag{6}$$

Based on the ray in world coordinates, $z_{ijk} \in \mathbb{R}$ can be computed such that:

$$\mathbf{X}_{ijk} = \mathbf{s}_{ijk}^{wc} + z_{ijk}\mathbf{r}_{w_{ijk}}^{wc} = \mathbf{R}_{ij}\mathbf{s}_{ijk}^{cc} + \mathbf{C}_{ij} + z_{ijk}\mathbf{R}_{ij}\mathbf{r}_{w_{ijk}}^{cc} \tag{7}$$

is the intersection of the ray with the xy-plane of the world coordinate system. In our calibration approach, the checkerboard is set to be fixed in that plane, with known corners $\hat{\mathbf{X}}_l, l < L$, thus all z-coordinates of the checkerboard points are zero.

Note that typical calibration routines minimize the reprojection error of 3D points projected into images. However, in case of refractive projection, this requires solving a 12th degree polynomial [1], hence it is convenient that the proposed AS approach only needs the back-projection of a pixel onto its ray.

3.3 Optimization Strategy

The optimization strategy used in our method is CMA-ES. The main reason for this is the ability of finding the global minimum even in presence of local minima.
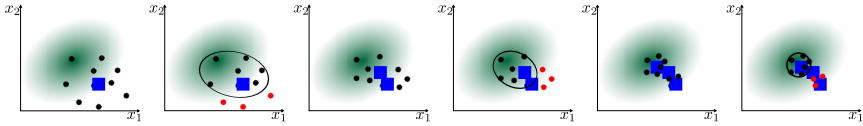
Fig. 3. CMA-ES iterations. Black dots are the samples drawn in each iteration, determined by mean (blue rectangle) and current covariance (black ellipse). The worst samples from each iteration are discarded (red dots), the remaining samples are used to update covariance and mean.

Figure 3 gives an idea of the method based on an exemplary two-dimensional error function (green ellipse). The algorithm is initialized with starting parameters (blue rectangle) and an initial parameter covariance. For each generation, a set of random samples is drawn based on current mean and covariance. Each sample is evaluated using the fitness function and the worst samples are eliminated (red dots, second image in top row). Using the remaining samples, the covariance is updated (black circle) and a new mean is computed. Those steps are repeated until mean and covariance converge toward the global minimum. While doing so, the covariance of the error function is adapted iteratively based on an update function that uses the covariance of the last step and the new samples. Note, that for this approach only the number of samples to be drawn in each generation γ_{CMA-ES} and an initial covariance need to be set. For small initial deviations and small γ_{CMA-ES}, the optimization is local, while for large deviations and γ_{CMA-ES}, a global search is conducted. The following section describes how model and optimization strategy are used in the presented calibration algorithm.

4   Calibration Algorithm

The calibration algorithm described here can be separated into initialization, geometric and radiometric optimization. During the initialization phase, the poses of the cameras with respect to the checkerboard and initial housing parameters are computed individually. The second step is an optimization over all cameras using an analysis-by-synthesis approach within an CMA-ES optimization routine. Afterward, the parameters describing the local water body are determined.

4.1   Corner-Based Initialization of Geometry

After taking a set of checkerboard images using the camera underwater, the calibration process is initialized by linearly determining camera poses, interface distances, and normals for each camera individually using the method described in [1]. Initialization is followed by a corner-based, non-linear optimization of the camera poses using common housing parameters for all cameras. In order not to optimize the 2D reprojection error in the images (section 3.2), the image points are back-projected onto the corresponding 3D-ray and then intersected with the
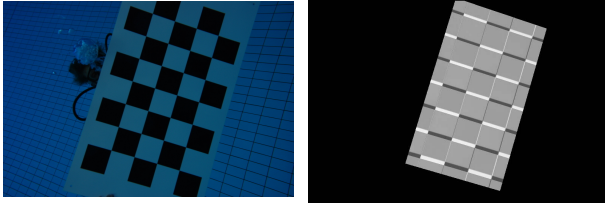
Fig. 4. Analysis-by-synthesis images. Left: original image. Right: pixel-wise difference between original image and rendered image, note that the squares of these differences are summed up in the error function.

checkerboard $\mathbf{X}_{ijl}$ (eq. (4) - (7)). This allows to compute a 2D error on the checkerboard, by using the known corners $\hat{\mathbf{X}}_l, l < L$. Then,

$$E_{\text{geometric}_{ij}} = \sum_{l<L} ||\mathbf{X}_{ijl} - \hat{\mathbf{X}}_l||_2^2 \tag{8}$$

is the error function that is used with CMA-ES to improve the initial camera poses, with 6 parameters per pose. Note, that the rotation matrix is represented by its incremental Euler-angles ϕ_x, ϕ_y, and ϕ_z. During the optimization, the small changes in the rotation matrix are approximated linearly [27]:

$$\mathbf{R}_{ij} \approx \mathbf{R}_{\text{init}_{ij}} + \begin{bmatrix} 0 & -\phi_{z_{ij}} & \phi_{y_{ij}} \\ \phi_{z_{ij}} & 0 & -\phi_{x_{ij}} \\ -\phi_{y_{ij}} & \phi_{x_{ij}} & 0 \end{bmatrix} \tag{9}$$

$\mathbf{R}_{ij}$ is projected to the closest valid rotation matrix.

4.2 Analysis-by-Synthesis Optimization of Geometry

During the analysis-by-synthesis stage of the algorithm, a 3D model of the checkerboard is rendered into the camera views, exploiting all image pixels simultaneously. The sum of squared differences between rendered and measured gray level values is the error function to be minimized. This allows to calibrate the parameters for the camera poses, the interface distance, and the interface normal directly from all pixels of the image data. In each iteration, a set of parameters is used to compute the ray for each point $\mathbf{x}_{ijk}$, known to lie on the checkerboard in an image:

$$(\mathbf{s}_{ijk}^{\text{cc}}, \mathbf{r}_{\text{w}_{ijk}}^{\text{cc}}) = \text{Ray}(\mathbf{x}_{ijk}, d_j, \mathbf{n}_{\pi_j}), \tag{10}$$

Using equations (6) and (7), the ray $(\mathbf{s}_{ijk}^{\text{cc}}, \mathbf{r}_{\text{w}_{ijk}}^{\text{cc}})$ is first transformed into the world camera coordinate system and then intersected with the checkerboard plane (xy-plane of world coordinate system) $\mathbf{X}_{ijk} = (X_{ijk}, Y_{ijk}, 0)$. Based on the coordinates $\mathbf{X}_{ijk}$ on the plane, the checkerboard gray value $I_{\text{check}_{ijk}}$ is determined and after applying α_{ij} and β_{ij} for gray-level contrast and offset, the rendered

gray-level value $I_{\mathrm{ren}}(\mathbf{x}_{ijk}) = \alpha_{ij} I_{\mathrm{check}_{ijk}} + \beta_{ij}$ is used in the error function to be minimized by CMA-ES:

$$E_{\mathrm{AbyS}} = \sum_{i<M} \sum_{j<N} \sum_{k<K} ||I_{\mathrm{ren}}(\mathbf{x}_{ijk}) - I(\mathbf{x}_{ijk})||_2^2. \qquad (11)$$

4.3   Radiometric Calibration

After geometric calibration, camera geometry is utilized to estimate the water parameters. First, a linear initialization is computed. When considering colors on the checkerboard, one knows that L_{obj} is either black or white, thus $L_{\mathrm{obj}} = 0$ or $L_{\mathrm{obj}} = 1$ respectively. When averaging the distance between camera and checkerboard for all pixels of the checkerboard $\bar{z}_{ij} = \frac{1}{K} \sum_{k<K} z_{ijk}$ and averaging the black and white observations, $\bar{I}_{\mathrm{b}\lambda_{ij}} = \frac{1}{K_{\mathrm{b}}} \sum_{k<K_{\mathrm{b}}} I_{\mathrm{b}\lambda_{ijk}}$ and $\bar{I}_{\mathrm{w}\lambda_{ij}} = \frac{1}{K_{\mathrm{w}}} \sum_{k<K_{\mathrm{w}}} I_{\mathrm{w}\lambda_{ijk}}$ respectively, a linearly estimated $\eta_{\lambda_{ij}}$ and $B_{\infty\lambda_{ij}}$ for each color channel $\lambda \in \{R, G, B\}$ can be determined for $\alpha_{\lambda_{ij}} = 1$, $\beta_{\lambda_{ij}} = 0$:

$$\eta_{\lambda_{ij}} = -\frac{\ln\left(\frac{\bar{I}_{\mathrm{w}\lambda_{ij}} - \bar{I}_{\mathrm{b}\lambda_{ij}}}{\alpha_{\lambda_{ij}}} - \beta_{\lambda_{ij}}\right)}{\bar{z}_{ij}}, \quad B_{\infty\lambda_{ij}} = \frac{\bar{I}_{\mathrm{b}\lambda_{ij}} - \beta_{\lambda_{ij}}}{\alpha_{\lambda_{ij}}(1 - \mathrm{e}^{-\eta_{\lambda_{ij}}\bar{z}_{ij}})} \qquad (12)$$

The initialization for η_λ and $B_{\infty\lambda}$ is then the mean from all images:

$$\eta_\lambda = \frac{1}{MN} \sum_{j<M} \sum_{i<N} \eta_{\lambda ij}, \quad B_{\infty\lambda} = \frac{1}{MN} \sum_{j<M} \sum_{i<N} B_{\infty\lambda_{ij}} \qquad (13)$$

Afterward, equation (3) is used in a non-linear Levenberg-Marquardt routine to optimize η_λ and $B_{\infty\lambda}$ for the whole scene and $\alpha_{ij\lambda}$ and $\beta_{ij\lambda}$ as additional parameters for white balance and offset for each image and color channel.

5   Experiments

CMA-ES[1] generates individuals of a population during each generation of the genetic algorithm. The evaluation of all individuals of a generation in CMA-ES can easily by parallelized using OpenMP[2]. Note that only parameters for the initialization, but not for the optimizer were changed throughout our experiments. As a rule of thump, the initial deviations for the parameters were set to approximately half of the maximal expected distance, e.g. the interface distance was expected to change up to 20 mm from the initialization, while the rotation angles were expected to change only slightly. The population size $\gamma_{\mathrm{CMA-ES}}$ was set to equal the number of parameters.

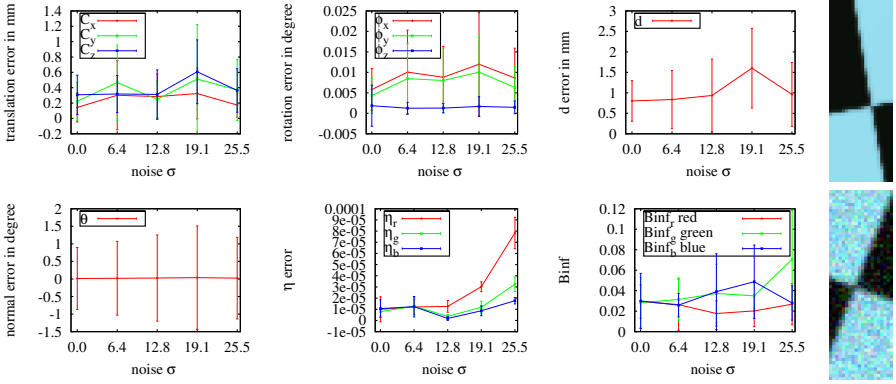[1] http://www.lri.fr/~hansen/cmaesintro.html
[2] http://openmp.org/wp/

Fig. 5. Synthetic images, accuracy of the different parameters in presence of growing noise in images. For rendered image intensity values $I \in [0, 255]$, normal distributed noise was added: $I_N = I + N(0, \sigma)$, with a cut-off for $I_N < 0$ or $I_N > 255$. On the right: exemplary image cut-outs with zero noise (top) and highest noise level (bottom).
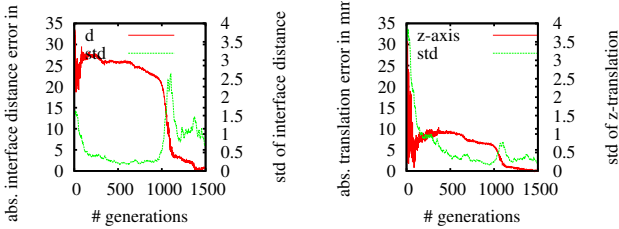

Fig. 6. Estimation path for an exemplary camera. Left: interface distance error (solid line) and standard deviation from CMA-ES adaptation (dashed line). Right: camera translation error in z-direction (solid line) and standard deviation from CMA-ES adaptation (dashed line).

5.1 Synthetic Data

In order to evaluate the accuracy of our method, synthetic checkerboard images were rendered. The checkerboard square size was $100\,mm$, with 10×8 squares. The camera had a resolution of 800×600 pixels and a focal length of 800 pixels. The distance between camera and checkerboard was between 1500 and $4000\,mm$. For 5 noise levels, we rendered 8 sets of images each, automatically detected checkerboard corners, and plotted the results in fig. 5. Note that the normal error depicted is the angle between the true normal and the computed normal. As can be seen, the parameters are computed with high accuracy even though a considerable amount of noise is added to the images. Since we use CMA-ES for optimization, it is possible to visualize the adaptation process of the uncertainty for the individual parameters. Expected are correlations between rotation and translation of the camera pose, interface distance and camera translation in z-direction, interface normal in x-direction and camera rotation around the y-axis,

|	Ground Truth	Result Left	Result Right
Perspective Calibration			
f	800	1059.39	1055.79
c_x, c_y	399.5, 299.5	404.47, 305.90	389.21, 300.78
$r_1, r_2,$	0, 0, 0, 0	0.382, 0.273,	0.379, 0.318,
t_1, t_2		0.005, 0.003	0.0002, -0.011
$\mathbf{C}_{\text{rig}}$	$(200, 0, 0)^{\text{T}}$	$(201.73, -1.60, 14.62)^{\text{T}}$	
Refractive Calibration			
d	10	11.57	10.62
$\mathbf{n}_\pi$	$\begin{pmatrix} 0.0076 \\ 0.0044 \\ 0.9999 \end{pmatrix}$	$\begin{pmatrix} 0.0075 \\ 0.0046 \\ 0.9999 \end{pmatrix}$	$\begin{pmatrix} 0.0074 \\ 0.0042 \\ 0.9999 \end{pmatrix}$
$\mathbf{C}_{\text{rig}}$	$(200, 0, 0)^{\text{T}}$	$(199.89, 0.02, -0.31)^{\text{T}}$	
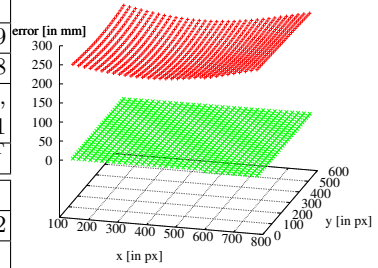
Fig. 7. Left: the table shows calibration results for a stereo rig using perspective calibration on underwater images (top 4 rows) and using the proposed method (bottom 3 rows). Right: the resulting calibrations are used to triangulate a plane up to 4000 mm away using the stereo rig (green: proposed method, red: perspective calibration). The presented errors arise depending on the pixel position in the left image and are due to error compensation by camera extrinsics during calibration. Note, that the calibration distance was up to 3500 mm.

and interface normal in y-direction and camera rotation around the x-axis. Fig. 6 shows the estimation path of the camera translation in z-direction and the interface distance using an exemplary camera set-up. As can be seen on the left, both the interface distance error and the standard deviation decrease with improving estimation results over the generations. Since interface distance and z-translation correlate, the uncertainty for the translation in z-direction increases after approximately 1000 generations.

In addition to evaluating the accuracy of our approach, we compare it to a perspective calibration based on images captured below water. Calibrating perspectively on underwater images causes the focal length and radial distortion to approximate the refractive effect. Therefore, we get different radial distortion parameters and the true focal length is multiplied by the index of refraction of water. In fig. 7 the results of calibrating a stereo rig are shown and for the perspective calibration the results differ strongly compared to the ground truth data. In case of radial distortion and focal length, this fits the results of [8] and [9]. However, in our experiments, part of the error is also absorbed by erroneous camera poses, especially translation: the average translation error in mm in the underwater case with the proposed method is $(0.25, 0.45, 0.48)^{\text{T}}$, while the average translation error in the perspective calibration is $(8.08, 10.96, 22.46)^{\text{T}}$. Figure 7 summarizes calibration results and compares the accuracy of the resulting calibrations, when applied to triangulation using the calibrated stereo rig. The novel camera model computes the distance error-free, while for the perspective camera model an error of up to 300 mm was introduced.

Table 1. Results of small scale test calibration. Note that the ground truth interface distance is unknown, but the camera rig was moved backwards in 10 mm steps, i. e. $dist_2 = dist_1 + 10$ mm and $dist_3 = dist_1 + 20$ mm. The angle θ is the measured angle between optical axis and interface normal. The second camera in the rig was fixed 50 mm to the right of the first one. In the experiments denoted with 'monocular' both cameras of the rig were calibrated individually, in the run denoted with 'stereo' they were calibrated as a rig.

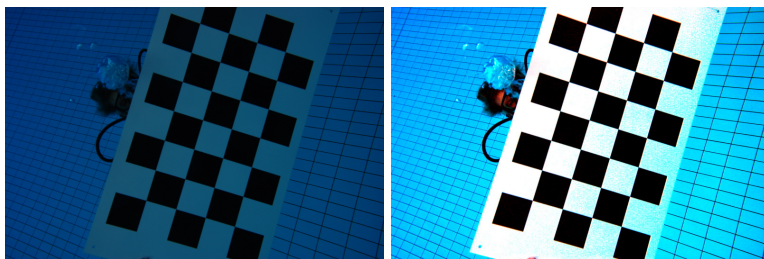|	d_{left} in mm	d_{right} in mm	θ_{left} in °	θ_{right} in °	translation of slave camera in mm
$dist_1$, monocular	88.86	89.87	0.59	0.84	
$dist_1$, stereo	88.09	87.63	0.86	1.38	$\mathbf{C} = (49.97, -0.02, -2.73)^{\text{T}}$
$dist_2$, monocular	99.53	102.96	0.81	0.77	
$dist_2$, stereo	103.38	111.18	1.18	1.59	$\mathbf{C} = (50.79, -0.29, -0.02)^{\text{T}}$
$dist_3$, monocular	122.56	115.82	1.49	2.56	
$dist_3$, stereo	122.40	114.99	0.89	1.71	$\mathbf{C} = (49.00, -0.22, -4.94)^{\text{T}}$

Fig. 8. Left: original color image of checkerboard with strong green hue. Right: corrected colors of the same image. Note that only the colors on the checkerboard are can be corrected because they lie on the xy-plane of the world coordinate system for which camera-object distance is known after calibration.

5.2 Real Data

The method was tested on two different real cameras. In the first scenario, two different camera-interface distances were tested in a small-scale camera setting in the lab. The results are summarized in table 1 and show that the method works well on these image sets.

In the second scenario, a camera was calibrated based on images captured in a pool with a strong blue hue (fig. 8, left). The estimated interface distance in this case was 74 mm, the angle θ between optical axis and interface normal was estimated to be 1.39°. Figure 8 shows an exemplary input image on the left and the resulting color corrected image on the right. The colors on the checkerboard are corrected with the estimated distance between checkerboard and camera. Near the bottom of the checkerboard, the red channel cannot be corrected due to the red light being almost completely absorbed by the water.

6   Conclusion and Future Work

We have presented a novel method for calibrating a camera including a parametrization of a flat port underwater housing. The method uses an analysis-by-synthesis approach in order to become independent from errors in corner detection for the geometric parameters. In contrast to well-known methods for perspective calibration, which use local derivative-based methods, our method applies CMA-ES in the optimization step. Although being time-consuming, this approach is less prone to get caught in local minima of the error function. This is of great advantage since even the simpler problem of pose estimation based on planar calibration targets is known to have ambiguities between rotation and scale. The resulting parameters describing the camera underwater housing can be utilized in applications based on the intrinsic geometric information in images such as SfM, mosaicing, and SLAM. In a second step, parameters for water color are calibrated and can be used for texture color correction for models or mosaics.

In the future, we plan to utilize the results of our method in order to develop a SfM system with explicit consideration of refraction. In addition, it would be interesting to compare a CMA-ES-based perspective calibration algorithm to see whether it outperforms local derivative optimization schemes like the Levenberg-Marquardt algorithm. Other than that, we plan to improve the run-time of the proposed method by using the GPU to render the images of the checkerboard.

References

1. Agrawal, A., Ramalingam, S., Taguchi, Y., Chari, V.: A theory of multi-layer flat refractive geometry. In: CVPR (2012)
2. Sedlazeck, A., Koser, K., Koch, R.: 3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions. In: Proc. OCEANS 2009. OCEANS 2009-EUROPE, pp. 1–10 (2009)
3. 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, 323–344 (1987)
4. Bouguet, J.Y.: Visual methods for three-dimensional modelling. PhD thesis, California Institute of Technology, California Institute of Technology Pasadena, CA, USA (1999)
5. 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. XXXVII. Part B3a, pp. 297–302 XXI. ISPRS Congress (2008)
6. 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 (2000)
7. Pessel, N., Opderbecke, J., Aldon, M.-J.: 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 (2003)
8. Fryer, J.G., Fraser, C.S.: On the calibration of underwater cameras. The Photogrammetric Record 12, 73–85 (1986)

9. Lavest, J.-M., Rives, G., Lapresté, J.T.: Underwater Camera Calibration. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 654–668. Springer, Heidelberg (2000)
10. Grossberg, M.D., Nayar, S.K.: The raxel imaging model and ray-based calibration. International Journal of Computer Vision 61(2), 119–137 (2005)
11. Narasimhan, S.G., Nayar, S.K., Sun, B., Koppal, S.J.: Structured Light in Scattering Media. In: IEEE International Conference on Computer Vision (ICCV), vol. I, pp. 420–427 (2005)
12. Ramalingam, S., Sturm, P., Lodha, S.: Theory and calibration algorithms for axial cameras. In: Proceedings of the Asian Conference on Computer Vision, Hyderabad, India, vol. I, pp. 704–713 (2006)
13. Treibitz, T., Schechner, Y.Y., Singh, H.: Flat refractive geometry. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8 (2008)
14. Telem, G., Filin, S.: Photogrammetric modeling of underwater environments. ISPRS Journal of Photogrammetry and Remote Sensing 65(5), 433 (2010)
15. Chari, V., Sturm, P.: Multiple-view geometry of the refractive plane. In: Proceedings of the 20th British Machine Vision Conference, London, UK (2009)
16. Kunz, C., Singh, H.: Hemispherical refraction and camera calibration in underwater vision. In: OCEANS 2008, pp. 1–7 (2008)
17. Li, R., Li, H., Zou, W., Smith, R., Curran, T.: Quantitative photogrammetric analysis of digital underwater video imagery. IEEE Journal of Oceanic Engineering 22, 364–375 (1997)
18. Sedlazeck, A., Koch, R.: Calibration of housing parameters for underwater stereo-camera rigs. In: Proceedings of the British Machine Vision Conference, pp. 118.1–118.11 (2011)
19. Eisert, P.: Model-based camera calibration using analysis by synthesis techniques. In: Proc. Vision, Modeling, and Visualization, VMV 2002, pp. 307–314 (2002)
20. Schechner, Y.Y., Karpel, N.: Clear underwater vision. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 1, I-536–I-543 (2004)
21. Li, R., Tao, C., Zou, W.: An underwater digital photogrammetric system for fishery geomatics. In: Intl. Archives of PRS, vol. XXXI, pp. 319–323 (1996)
22. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9(2), 159–195 (2001)
23. Stoll, C., Gall, J., de Aguiar, E., Thrun, S., Theobalt, C.: Video-based reconstruction of animatable human characters. ACM Transactions on Graphics (Proc. SIGGRAPH ASIA 2010) 29(6), 139–149 (2010)
24. Jordt, A., Koch, R.: Fast tracking of deformable objects in depth and colour video. In: Proceedings of the British Machine Vision Conference, BMVC 2011 (2011)
25. Jordt, A., Siebel, N.T., Sommer, G.: Automatic high-precision self-calibration of camera-robot systems. In: Proceedings of 2009 IEEE International Conference on Robotics and Automation (ICRA 2009), Kobe, Japan, pp. 1244–1249 (2009)
26. Glassner, A.S.: Principles of Digital Image Synthesis. Morgan Kaufmann Publishers Inc., San Francisco (1994)
27. McGlone, J.C. (ed.): Manual of Photogrammetry, 5th edn. ASPRS (2004)