Flat Refractive Geometry

Tali Treibitz , Yoav Y. Schechner
Department of Electrical Engineering
Technion - Israel Inst. Technology, Haifa 32000, Israel
ttali@tx.technion.ac.il, yoav@ee.technion.ac.il

Hanumant Singh
Woods Hole Oceanographic Institution
Woods Hole, MA 02543, USA
hsingh@whoi.edu

Abstract

While the study of geometry has mainly concentrated on single-viewpoint (SVP) cameras, there is growing attention to more general non-SVP systems. Here we study an important class of systems that inherently have a non-SVP: a perspective camera imaging through an interface into a medium. Such systems are ubiquitous: they are common when looking into water-based environments. The paper analyzes the common flat-interface class of systems. It characterizes the locus of the viewpoints (caustic) of this class, and proves that the SVP model is invalid in it. This may explain geometrical errors encountered in prior studies. Our physics-based model is parameterized by the distance of the lens from the medium interface, beside the focal length. The physical parameters are calibrated by a simple approach that can be based on a single-frame. This directly determines the system geometry. The calibration is then used to compensate for modeled system distortion. Based on this model, geometrical measurements of objects are significantly more accurate, than if based on an SVP model. This is demonstrated in real-world experiments.

1. Introduction

There is a growing interest in imaging systems that defy the single viewpoint (SVP) assumption. Studies have analyzed general imaging models [9, 28], or special cameras that do not have an SVP [24, 34]. In this work we analyze a common and important class of non-SVP systems. They are made of a standard perspective camera looking into a refractive medium through a flat interface. These systems are commonly used to look at objects in water. In aquaria, the water body is *embedded in the air-based* environment containing the camera. Alternatively, an air chamber enclosing the camera may be *embedded in the water*, as experienced by the eyes of human scuba-divers (Fig. 1a) or underwater robots. Both cases are equivalent as an imaging model.

The importance of analyzing this system class stems both from its wide relevance in a range of fields and from the vision challenges it poses. Beside computer vision, this system class further affects fields such as oceanic engineering [25, 26], psychology[19], agriculture [39], biology [32], geology [30] and archaeology. Starting with a motivation
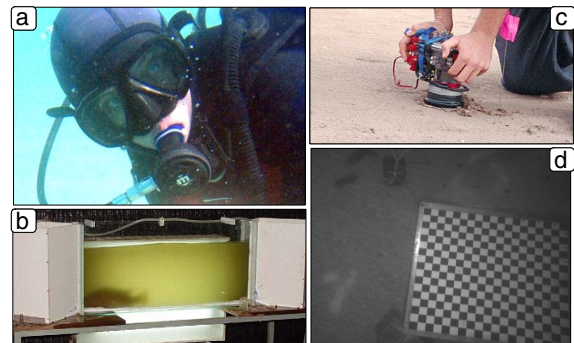

Figure 1. Vision through flat refractive interfaces. (a) A diver mask. (b) Fish moving through a transparent tunnel for online agricultural fish sorting, courtesy of B. Zion. (c) An underwater microscope for riverbed studies, courtesy of H. Chezar. (d) An underwater frame. Distortions increase towards the corner.

of human vision, distortions experienced by a lay person looking into a fish tank become critical for divers (Fig. 1a), as they use a face mask. As reported by [19], the distortions cause sever problems of distance perception, that were yet to be explained. Flat interfaces in aquaria were used in computer vision studies that developed methods of stereo [27], three dimensional (3D) recovery by structured light [10], classification [6], motion analysis [14] and visibility recovery [16]. In Ref. [39], a flat interface is used by a computer vision module in a system that sorts living edible fish in agricultural ponds (Fig. 1b). In field operations, extensive studies deal with stereo scene recovery [8, 22, 23, 25, 26, 32] by remotely-operated vehicles, which commonly use of a *flat port* (window). This interface is also used by *in-situ* microscopy [30] of river-beds (Fig. 1c).

The non-SVP nature of flat interface systems has not been pointed out and analyzed, to the best of our knowledge. The related literature treated these systems as perspective [8, 22, 23, 25, 26, 32]. As we show in this paper, the SVP assumption is *very erroneous*, in general, in this system class. Some studies regarded refraction as yielding a mere radial distortion [26, 32] in an SVP system. This coincided with reports [32] of unexpected and consistent errors when fitting an SVP model. We believe that such errors stem from the non-SVP nature of the system, as we
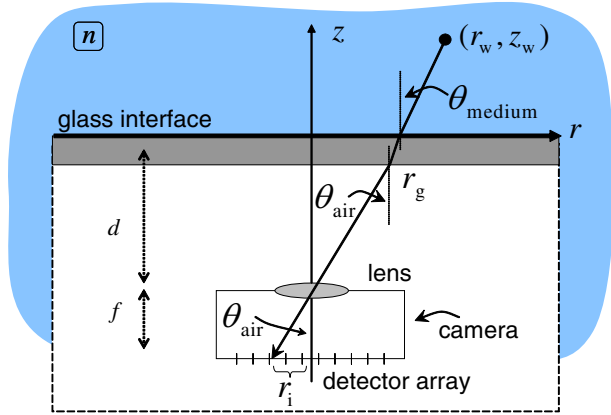
Figure 2. A ray from an object at $(r_{\rm w}, z_{\rm w})$ in the medium intersects a flat interface at the point $r_{\rm g}$. The ray continues through the air chamber until it reaches a perspective lens, at distance d from the interface. The image coordinate of the ray is $r_{\rm i}$.

show in this paper. In Fig. 1d, the distortions are seen in a two dimensional (2D) object. However, the non-SVP nature induces worse, 3D distortions, which are not a 2D (radial) mapping of coordinates: objects occluded under SVP may appear unoccluded by the system, and vice versa. Recently Ref. [14] concluded that a refraction model is needed for calibration and scene reconstruction.[1]

We model this system class, and show that it has a non-SVP. The caustic surface is derived in closed form and shown to have significant dimensions. We show that applying an SVP model can lead to errors in the order of dozens of pixels. To counter this, we suggest a physics-based calibration scheme for the pixels' ray map. It is easily applied in the field and can be based on a single-frame. It allows changes of zoom and focus settings *in situ*. Based on the ray map, geometrical tasks can be performed [33]. We demonstrate this in underwater experiments, by scuba diving in the ocean, using different lens settings. Our method significantly improves the accuracy of geometrical measurements.

2. Background

Consider a ray passing through a medium and a flat interface, as in Fig. 2. The setup has radial symmetry around the optical axis z, and r denotes the radial distance from the optical axis. The ray undergoes refraction when passing from the medium to a glass interface and again when passing from the glass to the air, where the camera resides. According to Snell's law [3]

$$n \sin \theta_{\rm medium} = n_{\rm glass} \sin \theta_{\rm glass} = \sin \theta_{\rm air} \ , \qquad (1)$$

where n and $n_{\rm glass}$ are respectively the refractive indices of the medium (e.g. water) and glass, relative to air. Here, $\theta_{\rm medium}, \theta_{\rm glass}, \theta_{\rm air}$ are the angles of the ray (relative to z) in the corresponding media. According to Eq. (1) the glass

---
[1]Ref. [21] used a refractive model to recover a non-flat water-interface.

interface (its index of refraction) does not change the refraction angle $\theta_{\rm air}$, for a given $\theta_{\rm medium}$.

While it does not change $\theta_{\rm air}$, the glass slightly shifts radially the ray's point of exit from the interface, since $\theta_{\rm medium} \neq \theta_{\rm glass}$. However, this shift is typically negligible relative to the other distances in the system,[2] because $n \sim n_{\rm glass}$. Hence, there is practically no need to correct for the effect inside the glass, and we may focus on the effects created by the bulk medium (e.g., water).

In Fig. 2, the distance z is measured from the air interface. The world coordinates of an object embedded in the medium are $(r_{\rm w}, z_{\rm w})$. A ray from this object is imaged to a point on the detector array. In this trajectory, the ray passes through a point on the interface, at a distance $r_{\rm g}$ from the axis. The value of $r_{\rm g}$ can be calculated using Fermat's principle: the ray path between two points is the one that is traversed in the least time. Accounting for the slower speed of light in the medium due to n, the optical path length is

$$L = n\sqrt{(r_{\rm w} - r_{\rm g})^2 + z^2} + \sqrt{r_{\rm g}^2 + d^2} \ , \qquad (2)$$

where d is the distance between the interface and the center of projection of the lens (at the plane of its entrance pupil). Thus, similarly to [7], $r_{\rm g}$ should satisfy the condition[3]

$$0 = \frac{\partial L}{\partial r_{\rm g}} = n\frac{r_{\rm g} - r_{\rm w}}{\sqrt{(r_{\rm g} - r_{\rm w})^2 + z^2}} + \frac{r_{\rm g}}{\sqrt{r_{\rm g}^2 + d^2}} \ . \qquad (3)$$

3. Modeling a Flat-Interface-Medium System

In this paper we look at a *system* defined by the combined effect of several elements, following the ray trajectory: `medium → interface → air → perspective camera`. Based on the simple principles described in Sec. 2, we now model the geometry sensed by this system. Note that in the heart of the system is an SVP camera. Thus, unless stated otherwise, the terms *focal length*, *center of projection* and *entrance/exit pupil* refer only to this internal camera in air. As we shall show, the *system* as a whole does *not* have a center of projection,[4] in general. Intuition into this can be gained by Fig. 3: rays coming from different objects appear as if imaged from different points of view.

3.1. Entrance Pupil in the Air

We seek the image coordinates corresponding to an object. We thus study the chief ray from the object to the pixel. After passing the interface to the air, the ray enters the camera lens. Consider first Fig. 2. For a perspective lens,

$$r_{\rm g} = dr_{\rm i}/f \ , \qquad (4)$$

---
[2]In a $5mm$ thick glass and $\theta_{\rm medium} = 20°$, the shift is $\sim 0.28mm$.
[3]The maximum of L is ∞. Therefore, the finite extremum of L yields the minimum path length.
[4]These effects are lowered by using a dome-shaped interface [31, 35, 36] or corrective optical modules [13, 29]. However, they require precise alignment to the camera center of projection, in tight tolerances [13]. This rules out changes of zoom or focus by the lens during the work, *in-situ*.
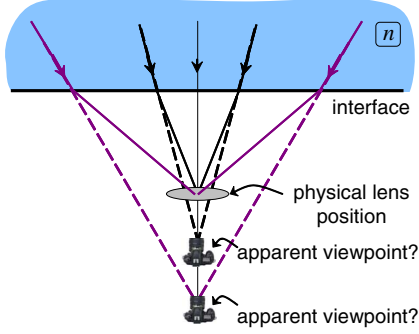
Figure 3. Looking through a flat interface into a medium yields a non-SVP system despite the use of a perspective camera.

where r_i is the radial coordinate of the image pixel corresponding to the ray. Here f is the focal length[5] of the camera in air. Using Eq. (4) in Eq. (3) yields

$$\left(r_w - \frac{d}{f}r_i\right)^2 \left[\left(\frac{fn}{r_i}\right)^2 + (n^2-1)\right] = z_w^2 \quad, \tag{5}$$

which relates the world coordinates (r_w, z_w) to the image coordinate r_i, as a function of d, f and n.

The setup in Fig. 2 is generalized in Fig. 4a: rather than a thin lens, it represents a camera having a compound lens. The chief ray is defined as a ray that passes though the center of the lens aperture (iris). The compound lens is considered as a black box, characterized by two planes: the *entrance pupil* and the *exit pupil*. The chief ray always enters the lens-module as if it crosses the axis point in the entrance pupil. It then emerges from the lens towards the detector array as if it crossed the axis point in the exit pupil. According to [1], the *center of projection of the camera* is at the center of the entrance pupil. Hence d is measured from the glass interface to the entrance pupil. Thus, Eq. (5) applies to compound lenses, with these definitions.

3.2. Entrance Pupil in the Glass

In the setup illustrated in Fig. 4b, the entrance pupil of the camera lies directly on the interface, i.e., $d = 0$. This special case is the only one in which the *system* as a whole has an SVP. The viewpoint is maintained at the entrance pupil (in the flat glass). However, the rays passing through this center of projection change direction due to refraction, as illustrated in Fig. 4b. This model was analyzed in Ref. [15].

At small angles, $\theta_{air} \ll 1$, Eq. (1) means that $\theta_{medium} \approx \theta_{air}/n$. Since $n > 1$, the angles are smaller in the medium. Hence the *system* (rather than the sole camera) behaves as if it has a longer effective focal length

$$f_{medium}^{effective} = nf \quad. \tag{6}$$

---
[5]The term "focal length," as commonly used in the computer vision literature, signifies the distance between the sensor and the lens' exit pupil.

Hence, if the camera is perspective, then the whole system is perspective, satisfying

$$r_i^{persp} = r_w \frac{f_{medium}^{effective}}{z_w} \quad. \tag{7}$$

The linear approximation in Eqs. (6,7) breaks down as θ_{air} increases. The nonlinear relation (Eq. 1) between the ray angles can be considered as a radial lens distortion. This was considered as a phenomenon to be modeled by a polynomial function of r_i in previous studies [26, 32]. This numerical approximation required empirical calibration of the polynomial coefficients.

In contrast, now we give an explicit, closed-form expression for the distortion created by refraction. Using $d = 0$ in Eq. (5) directly yields the relation between r_i and r_w

$$r_i = fn \left[(z_w/r_w)^2 - (n^2-1)\right]^{-\frac{1}{2}} \quad. \tag{8}$$

Using Eqs. (7,8), the distortion correction function is

$$r_i^{persp} = f_{medium}^{effective} \left[n^2 - 1 + \left(f_{medium}^{effective}/r_i\right)^2\right]^{-\frac{1}{2}} \quad. \tag{9}$$

Suppose that standard calibration yields $f_{medium}^{effective}$ at small angles around the axis *in-situ*. Then, Eq. (9) directly corrects for the nonlinear radial distortion, if the medium refractive index n is known. This alleviates the need for empirical polynomial calibration.

It must be stressed again that only at $d = 0$ can such distortions be modeled as a mere radial pixel shift in an SVP system. In all other cases, the model and calibration are more elaborate, as we describe.

3.3. Entrance Pupil in the Medium

It is important to understand that in some cases the effective center of projection of the camera can lie *outside* the air chamber, i.e., inside the medium. This occurs despite having all of the lens elements inside the air chamber. For instance, Ref. [1] describes a commercial lens module whose entrance pupil is $6cm$ in front of the physical lens barrel. In such cases, if the lens is adjacent to the interface, the entrance pupil (thus the center or projection) is effectively located in the medium. Such a case is depicted in Fig. 4c. In this case, Eq. (5) still applies, but here d is negative.

4. Caustics

Since generally the system cannot be described using the SVP model, we aim to characterize it through a ray map of the pixels. A ray is a parametric line[6] whose world coordinates are (R, Z). This ray is projected to a specific pixel.

---
[6]The azimuthal coordinate is not needed, since all the chief rays are meridional rays, as defined in [3]. This is true for a camera in air. Since the interface is perpendicular to the optical axis, the chief rays remain meridional.
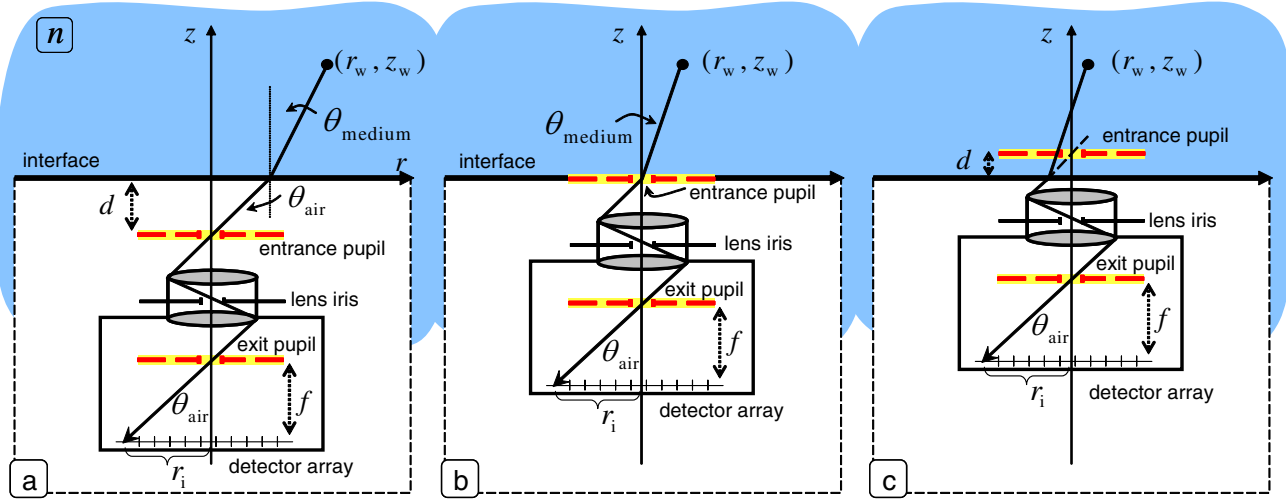
Figure 4. A lens module looking through a flat interface into a refractive medium. The entrance pupil can lie in the air (a), on the interface (b) or in the medium (c). In all subfigures the ray is imaged to the *same* pixel r_i, through a fixed focal length f. Moreover, the object distance z_w from the interface is the same. Despite the same r_i, f, z_w, the radial coordinate of the object r_w is different in each subfigure. Note that in (c), the lens images the ray as if the ray comes from the entrance pupil, although in reality the ray comes from a different direction and is refracted. Estimating an effective focal length $\hat{f}_{\text{medium}}^{\text{effective}}$ in the medium yields values that increase from (a) to (c).

Thus each pixel corresponds to a line in (R, Z). A parameter α determines the location along the ray,

$$\begin{bmatrix} R(r_i, \alpha) \\ Z(r_i, \alpha) \end{bmatrix} = \mathbf{p}(r_i) + \alpha \mathbf{q}(r_i) \ , \tag{10}$$

where $\mathbf{p}(r_i) = [p_R, p_Z]^T$ are the coordinates of the interface surface, $\mathbf{q}(r_i) = [q_R, q_Z]^T$ is the ray's direction vector and T denotes transposition. From Figs. 2 and 4, note that

$$\sin \theta_{\text{air}}(r_i) = \frac{r_i}{\sqrt{f^2 + r_i^2}} \ . \tag{11}$$

Then,
$$p_R = r_g \quad , \quad p_Z = 0 \ , \tag{12}$$

$$q_R = \frac{\sin \theta_{\text{air}}(r_i)}{n} \ , \quad q_Z = \cos \left\{ \sin^{-1} \left[\frac{\sin \theta_{\text{air}}(r_i)}{n} \right] \right\}. \tag{13}$$

Thus, the parametric ray representation is

$$R(r_i, \alpha) = \frac{d}{f} r_i + \alpha \frac{1}{n} \sin \theta_{\text{air}}(r_i) \tag{14}$$

$$Z(r_i, \alpha) = \alpha \cos \left\{ \sin^{-1} \left[\frac{1}{n} \sin \theta_{\text{air}}(r_i) \right] \right\} \ . \tag{15}$$

The differential change in coordinates from (r_i, α) to (R, Z) is expressed by the Jacobian matrix

$$\mathbf{J} = \begin{bmatrix} \left(\frac{\partial p_R}{\partial r_i} + \alpha \frac{\partial q_R}{\partial r_i} \right) & q_R \\ \left(\frac{\partial p_Z}{\partial r_i} + \alpha \frac{\partial q_Z}{\partial r_i} \right) & q_Z \end{bmatrix} \ . \tag{16}$$

The locus of the singularities in $\mathbf{J}$ represents a surface [5, 9], to which all chief rays are tangent. This is the *caustic* [3]. In the context of imaging, the caustic is regarded as the locus of all the focal points, i.e., the viewpoints of the system. For example, in a perspective system, the caustic is a single point. To find the caustic surface, we find the points where $|\mathbf{J}| = 0$. Applying this condition on Eq. (16) yields

$$\alpha = \frac{\left(q_R \frac{\partial p_Z}{\partial r_i} - q_Z \frac{\partial p_R}{\partial r_i} \right)}{\left(q_Z \frac{\partial q_R}{\partial r_i} - q_R \frac{\partial q_Z}{\partial r_i} \right)} \ . \tag{17}$$

Using Eq. (17) in Eqs. (10-13) yields the caustic coordinates

$$R_{\text{caustic}} = d\chi(r_i) \ , \quad Z_{\text{caustic}} = -dn \left[1 + \chi(r_i) \right]^{1.5} , \tag{18}$$

where

$$\chi(r_i) = \left(1 - \frac{1}{n^2} \right) \frac{r_i^2}{f^2} \ . \tag{19}$$

Obviously, following Eq. (18), the caustic is not a point (unless $d = 0$). Therefore, the system does not have an SVP.

Fig. 5 depicts the caustic in a field of view (FOV) for which $\max(\theta_{\text{air}}) = 50°$. From Eq. (18), both R_{caustic} and Z_{caustic} depend linearly on d. Therefore, the dimensions in Fig. 5 are normalized by d. An immediate conclusion is that one should place the camera such that d is as small as possible, in order to make a perspective model more valid.

Note that the extent of the caustic in Fig. 5 is larger than d. For $d = 2cm$, for example, the viewpoint in the image center is $\sim 2cm$ apart from the viewpoint at the image corners. Furthermore, when looking at an aquarium, d is in the order of tens of centimeters, leading to a similarly large spreading of the viewpoint locus. Suppose that one selects a single point in the caustic to be a representative viewpoint
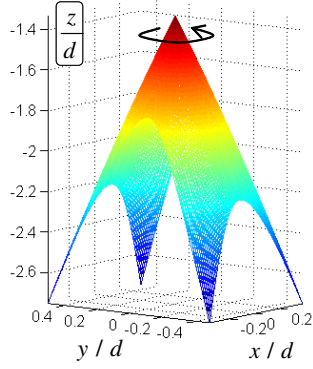
Figure 5. Caustic of a system having a flat interface with water. The camera has an FOV of $\max(\theta_{\text{air}}) = 50°$. The caustic has radial symmetry which is violated towards the boundaries of the FOV due to the rectangular shape of the sensor. The extent of the caustics is $\mathcal{O}(d)$, and can often reach centimeters or decimeters.
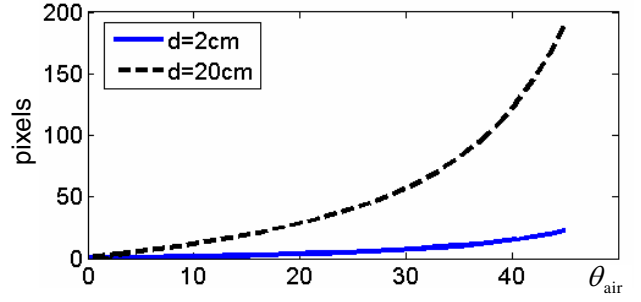

Figure 6. The error e_{SVP} caused by an SVP model. It is based on two objects imaged to the same pixel via a flat interface system. The distances of the objects are $z_{\text{w},1} = 1m$ and $z_{\text{w},2} = 1.5m$.

for an SVP model. The large spread of the caustic prevents any such point to yield a good approximation. This large spread is significant especially when imaging short ranges inside the medium, as done in turbid environments. Therefore, approximating the model by an SVP is clearly wrong.

When the entrance pupil is outside the lens (Sec. 3.3), the caustic shape is flipped upside down. Thus in this case, the entire wide set of viewpoints is in the medium.

5. When does the Perspective Model Apply?

Sometimes, cases described in Secs. 3.1 and 3.3 can be practically assumed to have an SVP. This section discusses when this can be done. Consider two object points $(r_{\text{w},1}, z_{\text{w},1})$ and $(r_{\text{w},2}, z_{\text{w},2})$ that lie on the same chief ray. According to the correct model (Sec. 3.1 or Sec. 3.3) they are imaged to the same point $r_{\text{i},1} = r_{\text{i},2}$. On the other hand, use of an SVP model would yield projected image points according to Eqs. (7-9). In this incorrect model, the image points would be $r_{\text{i},1}^{\text{persp}} \neq r_{\text{i},2}^{\text{persp}}$ when $z_{\text{w},1} \neq z_{\text{w},2}$. Therefore, the criterion

$$e_{\text{SVP}} \equiv r_{\text{i},1}^{\text{persp}} - r_{\text{i},2}^{\text{persp}} \qquad (20)$$

can indicate the error caused by assuming an SVP model.

Fig. 6 presents e_{SVP} as a function of the location in the FOV. Here $f = 20mm$, $z_{\text{w},1} = 1m$ and $z_{\text{w},2} = 1.5m$. A pixel is $5\mu m$ wide. The value $d = 2cm$ is typical to underwater housings, while $d = 20cm$ characterizes cases of aquaria. In Fig. 6, e_{SVP} is small only for very small angles or for a very small d. In these cases, the perspective model is valid. If θ_{air} and d are not very small, the SVP assumption yields very large errors, that can reach tens of pixels.

6. Calibration

Calibrating a camera having a non-central viewpoint involves calibrating a *ray map*, i.e., the trajectory of the light ray that each pixel captures. There are non-parametric methods for calibration of such non-SVP systems [9, 28]. However, Ref. [28] mentions stability difficulties, while the method in [9] may be complex to implement in a hostile environment in the sea. Both methods in [9, 28] require multiple frames. Fortunately, here the imaging model is parameterized thanks to its physics-based root. Thus calibration can focus simply on the unknown parameters.[7] Furthermore, the calibration can be based on a single frame.

Based on Sec. 3, we develop a parametric calibration process for the flat interface medium system. The index n is assumed to be known since according to [20], n varies by less than 3% over the entire range relevant to hydrologic optics, where $n \approx 4/3$. Nevertheless, n can be a free parameter that is involved in the calibration process as in [21]. The camera itself, irrespective of the medium of interest, has a center of projection. The image coordinates of this center are represented by $\mathbf{c} = [c_x, c_y]$.

To obtain intuition, consider an object at $(r_{\text{w}}^{\text{obj}}, z_{\text{w}}^{\text{obj}})$, which is imaged into the image coordinates $(x_{\text{i}}^{\text{obj}}, y_{\text{i}}^{\text{obj}})$. The radial image coordinate is

$$r_{\text{i}}^{\text{obj}} = \sqrt{(x_{\text{i}}^{\text{obj}} - c_x)^2 + (y_{\text{i}}^{\text{obj}} - c_y)^2} \ . \qquad (21)$$

Suppose for the moment that the camera had been calibrated in air, i.e., the inherent properties f and $\mathbf{c}$ of the perspective camera were measured, as in [11, 12, 38]. Using Eq. (5), the remaining unknown parameter is directly estimated:

$$\hat{d} = f \frac{r_{\text{w}}^{\text{obj}}}{r_{\text{i}}^{\text{obj}}} - z_{\text{w}}^{\text{obj}} \left[n^2 + \left(r_{\text{i}}^{\text{obj}}/f \right)^2 (n^2 - 1) \right]^{-\frac{1}{2}} \ . \qquad (22)$$

Eq. (22), however, is impractical. It requires the object's off axis position r_{w} to be known, which is difficult to achieve. Moreover, Eq. (22) requires prior calibration in air. This is inconvenient, especially if zoom or focus settings are changed *in-situ*. We thus present here a simpler way that works well *in-situ*, as we demonstrate in field experiments.

---

[7]A physics-based parameter calibration was applied on non-SVP catadioptric cameras [34] in air.
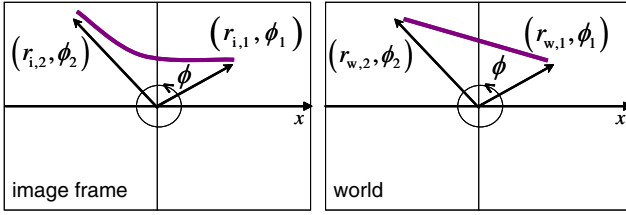
Figure 7. A straight object is imaged for calibration purposes. The object may appear bent in the raw frame. However, the world coordinates estimated by the model are unaffected.

The proposed method only requires imaging of an object of a known length s_{known} and distance $z_{\text{w}}^{\text{obj}}$. Identifying the object's extremities, we index these points as $m = 1, 2$. Their corresponding image coordinates are $(r_{\text{i},1}, \phi_1)$ and $(r_{\text{i},2}, \phi_2)$, where ϕ_m denotes the azimuthal coordinate of a point (Fig. 7). The object may appear bent in the frame due to the distortions. In the world coordinates, it maintains its straight form, though. Based on Eq. (5) the corresponding world coordinates $r_{\text{w},1}$ and $r_{\text{w},2}$ should satisfy

$$\hat{r}_{\text{w},m} = \frac{d}{f}r_{\text{i},m} + \frac{z_{\text{w}}^{\text{obj}}}{\sqrt{(fn/r_{\text{i},m})^2 + n^2 - 1}} \ . \quad (23)$$

Based on the law of cosines, the object length should satisfy

$$\hat{s} = \sqrt{(\hat{r}_{\text{w},1})^2 + (\hat{r}_{\text{w},2})^2 - 2\hat{r}_{\text{w},1}\hat{r}_{\text{w},2}\cos|\phi_1 - \phi_2|} \ . \quad (24)$$

Following Eqs. (21,23,24) the calibrated values $\hat{d}, \hat{f}, \hat{\mathbf{c}}$ are the ones that satisfy

$$\hat{s}(\hat{d}, \hat{f}, \hat{\mathbf{c}}) = s_{\text{known}} \ . \quad (25)$$

Hence $\hat{d}, \hat{f}$ and $\hat{\mathbf{c}}$ are parameters[8] of the image-formation model that are set to fit the known data. This fitting is easily generalized to work on a set of several measurements (e.g., a least-squares fit). Recall that if $\hat{d} < 0$, then the entrance pupil is in the medium.

Eqs. (21,23) assume that the camera in the air chamber does not have inherent radial distortions. However, if it does, then the parameters of these camera distortions can be incorporated into the fitting model. Now,

$$r_{\text{i}}^{\text{distorted}} = \sqrt{(x_{\text{i}} - c_x)^2 + (y_{\text{i}} - c_y)^2} \ , \quad (26)$$

since (x_i, y_i) is a distorted projection. The distorted camera projection is corrected (rectified) by a parametric model,[9] e.g.,

$$r_{\text{i}} = r_{\text{i}}^{\text{distorted}} + k(r_{\text{i}}^{\text{distorted}})^3 + \dots \quad (27)$$

The value of r_{i} resulting from Eqs. (26,27) is used in Eq. (23). Thus, the parameter k becomes part of the fitting degrees of freedom in Eq. (25).

[8]If f and $\mathbf{c}$ are known a-priori, (e.g., by prior calibration in air), then in this case the sole fitting parameter is $\hat{d}$.

[9]As in [37], here we do not need to refer to tangential distortion.

Rule of Thumb

Assume we calibrate the system using a perspective model in the medium. Then, we get an estimated $\hat{f}_{\text{medium}}^{\text{effective}}$, as described in Sec. 3.2. If $\hat{f}_{\text{medium}}^{\text{effective}} \approx nf$, as in Eq. (6), then $d \sim 0$. In this case, the errors induced by the SVP model might be negligible. If $|\hat{f}_{\text{medium}}^{\text{effective}} - nf| > 0$, then d is significant and our model has to be applied. Specifically, $\hat{f}_{\text{medium}}^{\text{effective}} < nf \iff d > 0$ and the opposite for $d < 0$.

7. Experiments

7.1. Calibration

We used a Nikon D100 camera and a Nikkor $24-85mm$ lens. They were housed in a Sealux underwater housing having a flat port PN94. We worked underwater in the ocean. The setup is shown in Fig. 8a. We used a checkerboard pattern as a calibration target (Fig. 8b). Each square is $27.5mm \times 27.5mm$. The known sizes enabled us to select, after image capture, a few lines of known length. Then, we applied Eqs. (23-25) on them. We show results of our method in two sessions. In each session we used the lens in different zoom and focus settings. In session 1, $z_{\text{w}}^{\text{obj}} = 48cm$, and $[\hat{d}, \hat{f}] = [7.9cm, 24.3mm]$. In session 2, $z_{\text{w}}^{\text{obj}} = 78cm$, and $[\hat{d}, \hat{f}] = [4.2cm, 58mm]$. We conducted similar experiments in a swimming pool, in which $[\hat{d}, \hat{f}] = [7.4cm, 26mm]$. Note that the value of d changes significantly when changing the lens setting f.

7.2. Validation

The calibration result $\hat{d}$ can now be applied for measuring unknown objects that are placed in distances that are generally different from the calibration distance. For validation we applied this measurement method on objects that we later measured independently by a ruler. Fig. 8c shows examples of such objects that we measured. On each imaged object we applied Eqs. (21,23,24), using the values of $\hat{d}, \hat{f}$ and $\hat{\mathbf{c}}$ calibrated in Sec. 7.1. To show the generalization of the model, about half of the measured validation objects are at a distance $z_{\text{w}}^{\text{obj}}$ that is different than the one used during calibration. The estimated sizes of the validation objects are $\hat{s}(\hat{d}, \hat{f}, \hat{\mathbf{c}})$. Table 1 compares $\hat{s}(\hat{d}, \hat{f}, \hat{\mathbf{c}})$ to the ground truth length s_{known} of each object.

7.3. Comparison to the Known Art

To demonstrate the significance of our approach, we compare it to the current practice. The current art [8, 23, 25, 26, 32] of vision through a flat interface has relied on the perspective model, regarding the refraction effects as a radial polynomial lens distortion. To emulate this, we used an off-the-shelf camera calibration toolbox [4]. This toolbox is based on [12]. It uses calibration images of a checkerboard pattern taken in different positions. We took such images underwater with our system. Then, the standard pro-
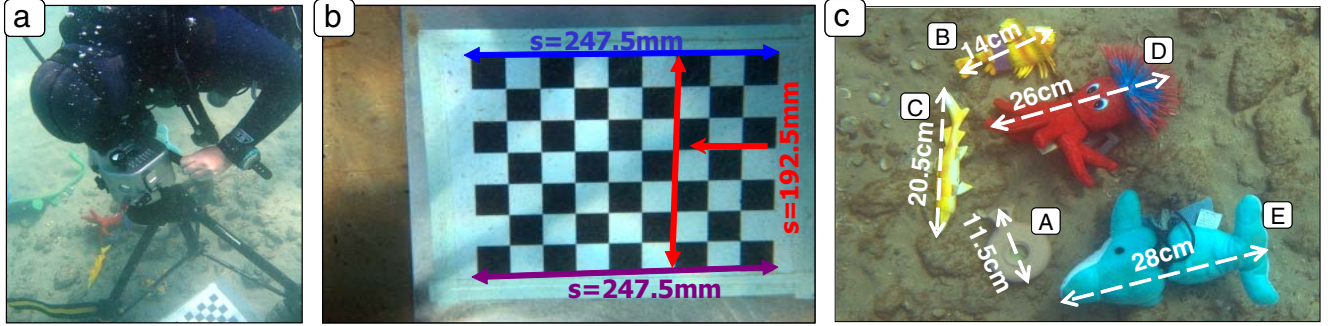
Figure 8. Calibration and validation experiments done in the ocean while scuba diving. (a) The imaging setup. (b) The checkerboard chart used for *in-situ* calibration. Specific line lengths are marked. (c) Underwater objects are used for validation of the calibration. They were measured independently.

session	object	$z_{\mathrm{w}}^{\mathrm{obj}}$	s_{known}	$\hat{s}(\hat{d}, \hat{f}, \hat{\mathbf{c}})$	$\hat{s}_{\mathrm{persp}}$
	A	48	11.5	11.5	8.9
	B	48	14.0	13.3	10.7
	C	48	20.5	20.0	15.7
1	A	134	11.5	11.1	10.1
	B	134	14.0	14.0	12.8
	C	134	20.5	20.6	18.6
	D	134	26.0	26.7	24.3
	E	134	28.0	29.2	26.8
	A	153	11.5	11.6	10.4
	B	153	14.0	14.3	12.7
2	C	153	20.5	20.5	18.7
	A	78	11.5	11.1	9.7
	B	78	14.0	13.6	11.9
	C	78	20.5	19.3	17.0

Table 1. Results of the validation experiments. Units are *cm*.

cess [12] obtained the camera parameters of the perspective model, particularly $\hat{f}_{\mathrm{medium}}^{\mathrm{effective}}$ (the focal length in water) and the presumed radial distortion parameters. We compensated the images for these estimated radial distortions. Then, we used $\hat{f}_{\mathrm{medium}}^{\mathrm{effective}}$ to find the objects' world coordinates

$$r_{\mathrm{w}}^{\mathrm{persp}} = r_{\mathrm{i}} z_{\mathrm{w}}^{\mathrm{obj}} / \hat{f}_{\mathrm{medium}}^{\mathrm{effective}} \quad . \tag{28}$$

Using Eq. (28) in Eq. (24) yielded an estimate for the object length $\hat{s}_{\mathrm{persp}}$, which appears in Table 1 for comparison. Our physics-based model, $\hat{s}(\hat{d}, \hat{f}, \hat{\mathbf{c}})$ which accounts for the non-SVP nature of the system is significantly more accurate: $\hat{s}(\hat{d}, \hat{f}, \hat{\mathbf{c}})$ fluctuates by $\approx 1 - 2\%$ around the correct value. In contrast, the error in $\hat{s}_{\mathrm{persp}}$ is $\approx 10\% - 15\%$. Moreover, $\hat{s}_{\mathrm{persp}}$ is *biased*. Similar results were obtained in the swimming pool.

8. Discussion

We suggested here a physics-based model for an imaging system having a flat refractive interface. The paper reveals that such a system does not have an SVP. The caustic extent is $\mathcal{O}(d)$ and thus can be significant. This yields significant 3D distortions. For calibration of a flat-refractive interface

system, we presented a method that can be applied *in-situ*, even if the lens settings are changed. It can be based on a single frame. In real experiments our method yielded results that are significantly more accurate than the SVP model.

These results have implications in the wide range of fields that use flat interface systems (see Sec. 1). Some system aspects may exist in microscopy. There, a liquid-based biological specimen or tissue is typically covered by a flat cover-glass, through which it is viewed. In other cases, the specimen is in a flask or Petri dish, and is viewed underneath, via the flat transparent dish. Thus, unless the objective lens used has an orthographic projection as described in Ref. [17], the system is prone to distortions as we describe.

When attempting underwater 3D stereo reconstruction with a flat port, assuming an SVP model is likely to yield significant errors. For stereo, there is need to use the *ray-map* which is calibrated in this paper. Closed-form stereo in this system requires additional theoretical work. Other possible extensions include self-calibration methods, to alleviate the need for a known calibration object. In addition, we plan to examine the sensitivity of the fitting problem (25) to the imaging parameters and data. It is further worth modeling interfaces that are not perpendicular to the optical axis, or even curved [18]. The physics-based geometric model can be expanded into a radiometric falloff model, based on Fresnel's laws of refraction. Defocus analysis, as in [2] is also beneficial.

Acknowledgements

References

[1] M. Aggarwal and N. Ahuja. A pupil-centric model of image formation. *Int. J. Comp. Vision*, 48:195–214, 2002.

[2] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *Int. J. Comp. Vision*, 35:175–196, 1999.

[3] M. Born and E. Wolf. *Principles of Optics*, chapter 3,4. Cambridge University Press, 7 edition, 1999.

[4] J.-Y. Bouguet. Camera calibration toolbox for matlab. www.vision.caltech.edu/bouguetj/calib_doc.

[5] D. G. Burkhard and D. L. Shealy. Flux density for ray propagation in geometrical optics. *JOSA*, 63:299–304, 1973.

[6] D. R. Edgington, I. Kerkez, D. E. Cline, J. Mariette, M. Ranzato, and P. Perona. Detecting, tracking and classifying animals in underwater video. In *IEEE CVPR, demo papers*, 2006.

[7] G. Glaeser and H. P. Schröcker. Reflections on refractions. *J. Geometry & Graphics*, 4:1–18, 2000.

[8] N. Gracias and J. Santos-Victor. Underwater video mosaics as visual navigation maps. *CVIU*, 79:66–91, 2000.

[9] M. D. Grossberg and S. K. Nayar. The raxel imaging model and ray-based calibration. *Int. J. Comp. Vision*, 61:119–137, 2005.

[10] M. Gupta, S. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. *IEEE CVPR*, 2008.

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, chapter 6. Cambridge University Press, New York, NY, USA, 2nd edition, 2003.

[12] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *Proc. IEEE CVPR*, pages 1106–1112, 1997.

[13] A. Ivanoff and P. Cherney. Correcting lenses for underwater use. *J. Soc. Mot. Pict. & Televis. Eng.*, 69:264–266, 1960.

[14] Y. H. Kwon and J. B. Casebolt. Effects of light refraction on the accuracy of camera calibration and reconstruction in underwater motion analysis. *Sports Biomech.*, 5:315–340, 2006.

[15] J. M. Lavest, G. Rives, and J. T. Lapresté. Underwater camera calibration. In *Proc. ECCV*, pages 654 – 668, 2000.

[16] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas. Synthetic aperture confocal imaging. *ACM TOG*, 23:825–834, 2004.

[17] M. Levoy, R. Ng, A. Adams, , M. Footer, and M. Horowitz. Light field microscopy. *ACM TOG*, 25:924–934, 2006.

[18] R. Li, H. Li, W. Zou, R. G. Smith, and T. A. Curran. Quantitative photogrammetric analysis of digital underwater video imagery. *IEEE J. Oceanic Eng.*, 2:364–375, 1997.

[19] S. M. Luria and J. A. S. Kinney. Underwater vision. *Science*, 167:1454–1461, 1970.

[20] C. D. Mobley. *Light and Water: Radiative Transfer in Natural Waters*, chapter 3.6. San-Diego: Acad. Press, 1994.

[21] N. J. W. Morris and K. N. Kutulakos. Dynamic refraction stereo. In *Proc. IEEE ICCV*, pages 1573 – 1580, 2005.

[22] S. Negahdaripour, C. Barufaldi, and A. Khamene. Integrated system for robust 6-DOF positioning utilizing new closed-form visual motion estimation methods in planar terrains. *IEEE J. Oceanic Eng.*, 31:533–550, 2006.

[23] S. Negahdaripour, H. Sekkati, and H. Pirsiavash. Opti-acoustic stereo imaging, system calibration and 3-d reconstruction. In *Proc. IEEE Beyond Multiview Geometry*, 2007.

[24] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: panoramic stereo imaging. *IEEE Trans. PAMI*, 23:279–290, 2001.

[25] N. Pessel, J. Opderbecke, and M. Aldon. Camera self-calibration in underwater environment. In *Int. Conf. in Cent. Eur. Comp. Graphics, Visualiz. & Comp. Vision*, 2003.

[26] O. Pizarro, R. Eustice, and H. Singh. Relative pose estimation for instrumented, calibrated imaging platforms. In *proc. VIIth Digital Imaging Comp., Tech. and Applications Conf.*, pages 601–612, 2003.

[27] J. P. Queiroz-Neto, R. Carceroni, W. Barros, and M. Campos. Underwater stereo. In *Brazilian Symp. Comp. Graphics Image Processing*, pages 170– 177, 2004.

[28] S. Ramalingam, P. Sturm, and S. Lodha. Towards complete generic camera calibration. In *Proc. IEEE CVPR*, volume 1, pages 1093–1098, 2005.

[29] S. Ray. *Applied Photographic Optics*, chapter 43, pages 349–352. Focal Press, 3rd edition, 2002.

[30] D. M. Rubin, H. Chezar, J. N. Harney, D. J. Topping, T. S. Melis, and C. R. Sherwood. Underwater microscope for measuring spatial and temporal changes in bed-sediment grain size. *Sedimentary Geology*, 202:402–408, 2007.

[31] Y. Y. Schechner and N. Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE J. Oceanic Eng.*, 30:570–587, 2005.

[32] M. R. Shortis and E. S. Harvey. Design and calibration of an underwater stereo-video system for the monitoring of marine fauna populations. *Int. Archives Photogrammetry & Remote Sensing*, 32:792–799, 1998.

[33] P. Sturm, S. Ramalingam, and S. Lodha. On calibration, structure from motion and multi-view geometry for generic camera models. In *Imaging Beyond the Pinhole Camera*, pages 87–105. Springer, 2006.

[34] R. Swaminathan, M. D. Grossberg, and S. K. Nayar. Non-single viewpoint catadioptric cameras: Geometry and analysis. *Int. J. Comp. Vision*, 66:211–229, 2006.

[35] T. Treibitz and Y. Y. Schechner. Instant 3Descatter. In *Proc. IEEE CVPR*, pages 1861–1868, 2006.

[36] T. Treibitz and Y. Y. Schechner. Active polarization descattering. *To appear in IEEE PAMI*, 2008.

[37] G. Q. Wei and S. D. Ma. Implicit and explicit camera calibration: theory and experiments. *IEEE Trans. PAMI*, 16:469–480, 1994.

[38] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. PAMI*, 22:1330–1334, 2000.

[39] B. Zion, V. Alchanatis, V. Ostrovsky, A. Barki, and I. Karplus. Real-time underwater sorting of edible fish species. *J. Comp. Electr. in Agriculture*, 56:34–45, 2007.