Flat Refractive Geometry

Tali Treibitz, *Member*, *IEEE*, Yoav Y. Schechner, *Member*, *IEEE*,
Clayton Kunz, and Hanumant Singh, *Senior Member*, *IEEE*

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, besides 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. In addition, we examine by simulation the errors expected by using the SVP model. We show that when working at a constant range, the SVP model can be a good approximation.

Index Terms—Computer vision, vision and scene understanding, 3D/stereo scene analysis, camera calibration, imaging geometry.

1 Introduction

There is a growing interest in imaging systems that defy the single viewpoint (SVP) assumption. Studies have analyzed general imaging models [12], [34] or special cameras that do not have an SVP [25], [30], [42]. 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 cameras on 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. Besides computer vision, this system class further affects fields such as oceanic engineering [31], [32], psychology [23], agriculture [51], biology [39], geology [36], and archeology. Starting with a motivation 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 in [23], the distortions cause severe problems of distance perception that have yet to be explained. Flat interfaces in aquaria were used in computer vision studies that developed methods of stereo [33], [49], three-dimensional (3D) recovery by structured light [13], classification [8], motion analysis [19], and visibility recovery [21]. In [51], 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 [11], [28], [29], [31], [32], [39] by remotely operated vehicles, which commonly use a *flat port* (window). This interface is also used by *in-situ* microscopy [36] of riverbeds (Fig. 1c).

The related literature has mostly treated flat-interface systems as perspective [11], [28], [29], [31], [32], [39]. Photogrammetry studies [22], [24], [44] calibrate for system parameters directly. As we show in this paper, the SVP assumption is *significantly erroneous* in general in this system class. Some studies regarded refraction as yielding a mere transversal distortion [32], [39] in an SVP system. This coincided with reports [39] 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 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, Kwon and Casebolt [19] 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. Thus, we examine by simulation the expected errors from using an SVP

- *T. Treibitz is with the Department of Computer Science and Engineering, University of California, San Diego, 9500 Gilman Drive, Mail Code 0404, La Jolla, CA 92093-0404. E-mail: tali@cs.ucsd.edu.*
- *Y.Y. Schechner is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel. E-mail: yoav@ee.technion.ac.il.*
- *C. Kunz is with the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Blake 211, Woods Hole, MA 02543. E-mail: ckunz@whoi.edu.*
- *H. Singh is with the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Blake 211, Woods Hole, MA 02543. E-mail: hsingh@whoi.edu.*

Manuscript received 13 Dec. 2009; revised 3 Jan. 2011; accepted 17 Mar. 2011; published online 13 May 2011.
Recommended for acceptance by A. Fitzgibbon.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-12-0815.
Digital Object Identifier no. 10.1109/TPAMI.2011.105.

1. Morris and Kutulakos [27] used a refractive model to recover a nonflat water interface.

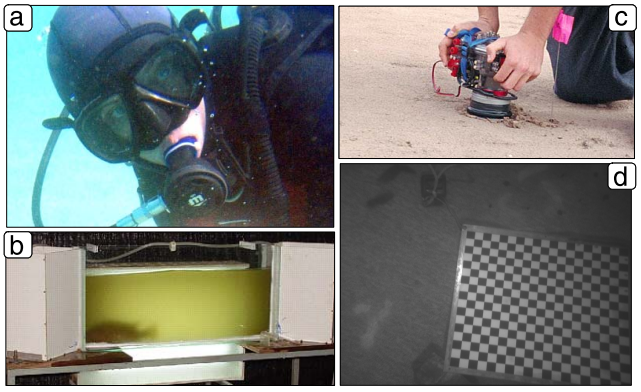


Fig. 1. Vision through flat refractive interfaces. (a) A diver mask. (b) Fish moving through a transparent tunnel for online agricultural fish sorting, courtesy of B. Zion. (c) An underwater microscope for riverbed studies, courtesy of H. Chezar. (d) An underwater frame. Distortions increase toward the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

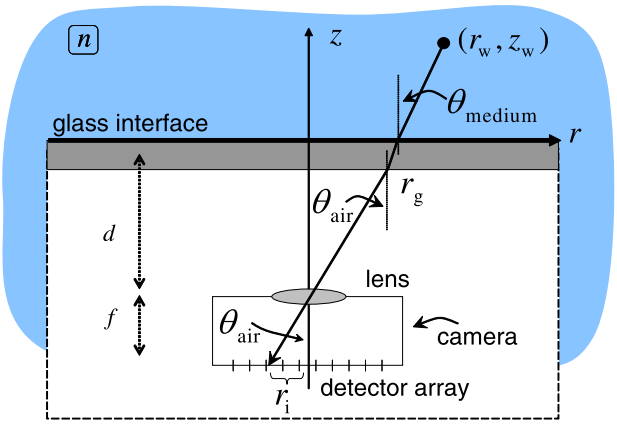


Fig. 2. A ray from an object at (r_w, z_w) in the medium intersects a flat interface at point r_g. The ray continues through the air chamber until it reaches a perspective lens, at distance d from the interface. The image coordinate of the ray is r_i.

approximation for a flat-interface system. We show that in a constant range, SVP camera models can be used. In addition, 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 [41]. 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. Partial results were presented in [48].

2 Background

2.1 Refraction

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. We assume that the optical axis is perpendicular to the interface. 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 [4],

$$n \sin\theta_{\text{medium}} = n_{\text{glass}} \sin\theta_{\text{glass}} = \sin\theta_{\text{air}}, \tag{1}$$

where n and n_{glass} are respectively the refractive indices of the medium (e.g., water) and glass, relative to air. Here, $\theta_{\text{medium}}, \theta_{\text{glass}}, \theta_{\text{air}}$ are the angles of the ray (relative to z) in the corresponding media. According to (1), the glass interface (its index of refraction) does not change the refraction angle θ_{air}, for a given θ_{medium}.

While it does not change θ_{air}, the glass slightly shifts radially the ray's point of exit from the interface, since $\theta_{\text{medium}} \neq \theta_{\text{glass}}$. The effect of this shift is smaller in magnitude than the effect associated with angular refraction. For example, for a 5 mm thick glass, $\theta_{\text{medium}} = 20°$, $f = 20$ mm, and $d = 20$ mm, the pixel shift caused by angular refraction is 10 times the lateral shift caused by the glass. Hence, in this paper we 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_w, z_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_g from the axis. The value of r_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

$$L = n\sqrt{(r_w - r_g)^2 + z_w^2} + \sqrt{r_g^2 + d^2}, \tag{2}$$

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 [10], r_g should satisfy the condition[2]

$$0 = \frac{\partial L}{\partial r_g} = n\frac{r_g - r_w}{\sqrt{(r_g - r_w)^2 + z_w^2}} + \frac{r_g}{\sqrt{r_g^2 + d^2}}. \tag{3}$$

2.2 The SVP Camera Model

The camera behind the interface is perspective. Therefore, here we shortly review the model that describes it, following [15]. Assuming the world coordinates are given in the camera reference frame, the two sets of coordinates are related via a 3×3 matrix $\mathbf{K}$,

$$\tilde{\mathbf{x}}_i = \mathbf{K}\mathbf{x}_w. \tag{4}$$

Here, the world coordinate vector is $\tilde{\mathbf{x}}_w = [x_w, y_w, z_w]^T$, where ^{T} denotes transposition. The image homogenous coordinate vector is $\tilde{\mathbf{x}}_i = [\tilde{x}_i, \tilde{y}_i, \tilde{w}_i]^T$. Image coordinates (x_i, y_i) are related to the homogeneous coordinates by $[x_i, y_i] = (1/\tilde{w}_i)[\tilde{x}_i, \tilde{y}_i]$. The matrix $\mathbf{K}$ encapsulates the intrinsic parameters of the system

$$\mathbf{K} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

Here, f is the focal length, expressing the distance between the sensor and the lens' exit pupil. As our problem has radial symmetry, we assume a uniform focal length for both x and y-axes. The coordinates (c_x, c_y) are the coordinates of the "principal point" where the optical axis intersects the image

2. The maximum of L is ∞. Therefore, the finite extremum of L yields the minimum path length.

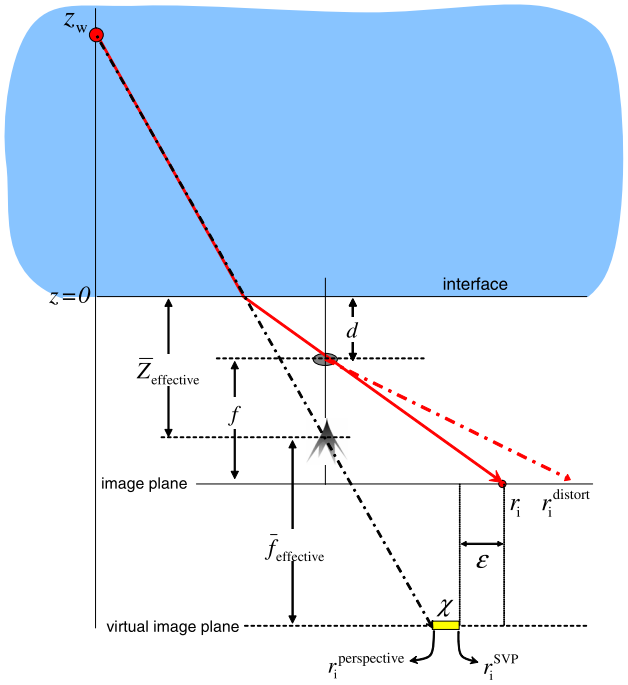


Fig. 3. Notations for radial coordinates used in the paper. A ray from the object is refracted by the interface, passes the camera pinhole, and projected to the coordinate r_i. If the camera lens introduces distortion, the distorted image coordinate is r_i^{distort}. As explained in Section 7.1, assuming the system is perspective, the object can be projected using the effective focal length $f_{\text{effective}}$, to the coordinate $r_i^{\text{perspective}}$. If the system is calibrated using the model in (4) and (6), radial distortion extent χ is added to the coordinate $r_i^{\text{perspective}}$. This results in the coordinate r_i^{SVP}. The difference between r_i and r_i^{SVP} is the SVP approximation error ε. Note that the refraction angle is exaggerated for purposes of demonstration.

plane. We assume a high quality camera and therefore we do not account for skew between the x and y-axes.

In addition to the linear operation of projection, the perspective model also includes nonlinear terms to capture lens distortion. As in [5] and [15], one may use a model which includes two parameters each for radial transversal distortion, κ_1 and κ_2. Usually, tangential distortion can be neglected [50]. In this model, observed (distorted) pixels $(x_i^{\text{distort}}, y_i^{\text{distort}})$ are related to ideal pixels (x_i, y_i) by

$$x_i^{\text{distort}} = x_i + (x_i - c_x)\left[\kappa_1 r_i^2 + \kappa_2 r_i^4\right], \tag{6}$$

$$y_i^{\text{distort}} = y_i + (y_i - c_y)\left[\kappa_1 r_i^2 + \kappa_2 r_i^4\right], \tag{7}$$

where $r_i^2 = \bar{x}_i^2 + \bar{y}_i^2$, while $\bar{x}_i = (x_i - c_x)/f, \bar{y}_i = (y_i - c_y)/f$. The radial coordinate

$$r_i^{\text{distort}} = \sqrt{\left(x_i^{\text{distort}} - c_x\right)^2 + \left(y_i^{\text{distort}} - c_y\right)^2}$$

is depicted in Fig. 3. We define

$$\chi = r_i^{\text{distort}} - r_i \tag{8}$$

as the magnitude of the radial lens distortion.

3 Modeling a Flat-Interface-Medium System

In this paper, we study a *system* defined by the combined effect of several elements, following the ray trajectory:

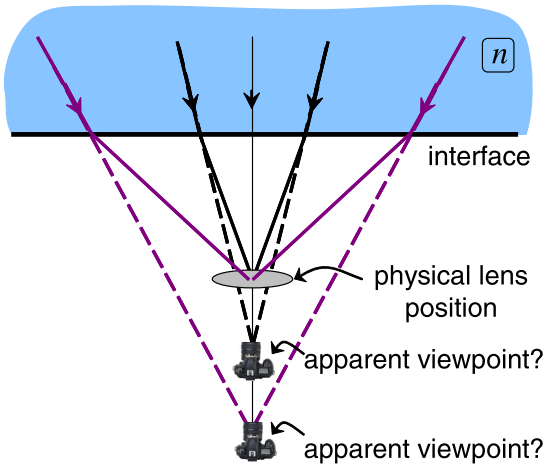


Fig. 4. Looking through a flat interface into a medium yields a non-SVP system despite the use of a perspective camera.

`medium` → `interface` → `air` → `perspective camera`. Based on the simple principles described in Section 2.1, 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 in general. Intuition into this can be gained by Fig. 4: Rays coming from different objects appear as if imaged from different points of view. These effects are lowered by using a dome-shaped interface [37], [46], [47] or corrective optical modules [17], [35]. However, they require precise alignment to the camera center of projection, in tight tolerances [17], [18]. This rules out changes of zoom or focus by the lens during the work *in situ*.

Another interesting phenomenon is illustrated in Fig. 5. Both the square and the round objects are visible when viewed in air. However, in water they are projected into the same pixel. Thus, the squared object occludes the round one. In addition, Fig. 5 shows that the pixel shift along the radial axis caused by refraction depends on the object distance. We elaborate on this in Section 7.1. Thus, from now on we refer to the following terms:

Definition 1. Radial distortion *is a pixel shift along the radial axis.*

Definition 2. Transversal distortion *is a radial distortion whose magnitude depends solely on the radial coordinate.*

Definition 3. Nontransversal distortion *is a radial distortion with magnitude that* depends on the object distance *in addition to the radial coordinate.*

Occlusions are also demonstrated in Fig. 6. Here, images were taken by a camera looking into an aquarium. The top image shows the scene when the aquarium was empty. The bottom image shows the scene when the aquarium was full of water. On the right side of the scene, there is a red stick visible in the air. In the water, this object suddenly appears as a fence. On the left side of the scene, the introduction of water suddenly introduces an image of a shark. These effects could have been imitated by an in-air SVP system by moving the

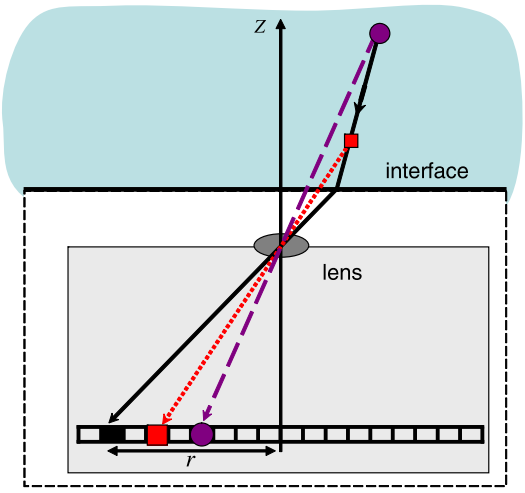


Fig. 5. The distortion caused by refraction is nontransversal and depends on the object distance. When embedded in water, both the square and the round objects are projected into the same coordinate. However, in air (the undistorted projection), the objects are projected into different coordinates. Thus, calibrating radial distortion according to one of the objects results in an error when rectifying the other. This happens because the distortion is nontransversal and depends on the object distance in addition to the radial coordinate.

camera *away* from the aquarium. However, the center of the field of view (FOV) is enlarged in water, as if the camera is *closer*. This demonstrates that the refraction causes severe 3D distortions. They confuse human observers, who are used to a perspective world, as well as computer vision systems that use a wrong imaging model. A movie that shows the scene change as the water fills the aquarium is available in [45].

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,

$$r_{\mathrm{g}} = r_{\mathrm{i}} d / f, \tag{9}$$

where r_{i} is the radial coordinate of the image pixel corresponding to the ray. As in Section 2.2, f is the focal length of the camera in air. Note that in this analysis, we assume that the sensor is parallel to the interface. Section 7.6 elaborates on the case when this assumption does not hold. Using (9) in (3) yields

$$\left(r_{\mathrm{w}} - \frac{d}{f} r_{\mathrm{i}}\right)^2 \left[\left(\frac{fn}{r_{\mathrm{i}}}\right)^2 + (n^2 - 1)\right] = z_{\mathrm{w}}^2, \tag{10}$$

which relates the world coordinates $(r_{\mathrm{w}}, z_{\mathrm{w}})$ to the image coordinate r_{i}, as a function of d, f, and n. To solve for r_{i}, (10) results in a fourth degree polynomial. Its four roots are valid mathematically, but only the root that is valid physically should be chosen (i.e., no complex or negative values).

The setup in Fig. 2 is generalized in Fig. 7a: Rather than a thin lens, it represents a camera having a compound lens. The chief ray is defined as a ray that passes through 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 toward 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

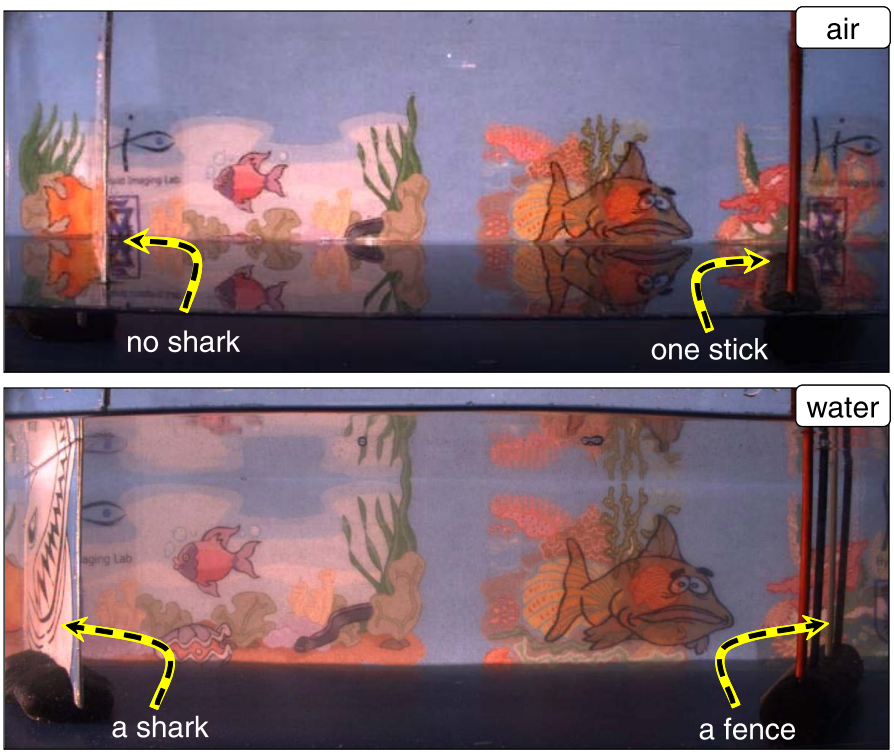


Fig. 6. A scene imaged by a camera looking into an aquarium, empty (top) and full with water (bottom). On the right side, the red stick in the air is seen as a fence in water. On the left side, a shark is suddenly apparent.

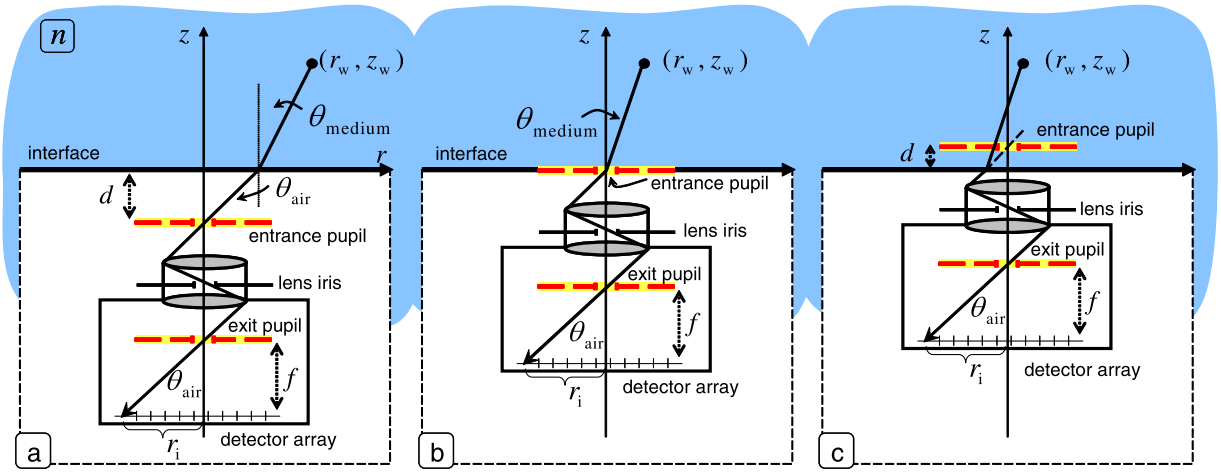


Fig. 7. A lens module looking through a flat interface into a refractive medium. The entrance pupil can lie (a) in the air, (b) on the interface, or (c) in the medium. In all subfigures, the ray is imaged to the *same* pixel r_i, through a fixed focal length f. Moreover, the object distance z_w from 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 $\hat{f}_{\text{effective}}$ in the medium yields values that increase from (a) to (c).

glass interface to the entrance pupil. Thus, (10) applies to compound lenses, with these definitions.

3.2 Entrance Pupil in the Glass

In the setup illustrated in Fig. 7b, 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. 7b.

At small angles, $\theta_{\text{air}} \ll 1$, and thus (1) is linearized to $\theta_{\text{medium}} \approx \theta_{\text{air}}/n$. 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 [20]

$$f_{\text{effective}}|_{\theta_{\text{air}} \ll 1} = nf. \tag{11}$$

The linear approximation above breaks down as θ_{air} increases. The nonlinear relation (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 [9], [32], [39]. 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 (10) directly yields the relation between r_i and r_w:

$$r_i = fn[(z_w/r_w)^2 - (n^2 - 1)]^{-\frac{1}{2}}. \tag{12}$$

Following (5), in radial coordinates:

$$r_w = f_{\text{effective}}|_{\theta_{\text{air}} \ll 1} z_w/r_i. \tag{13}$$

Plugging r_w from (13) in (12), the distortion correction function is

$$r_i^{\text{perspective}} = r_i \left[1 + \left(\frac{r_i}{f_{\text{effective}}|_{\theta_{\text{air}} \ll 1}} \right)^2 (n^2 - 1) \right]^{-\frac{1}{2}}. \tag{14}$$

Suppose that standard calibration yields $f_{\text{effective}}|_{\theta_{\text{air}} \ll 1}$ at small angles around the axis *in situ*. Then, (14) 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, Aggarwal and Ahuja [1] describe a commercial lens module whose entrance pupil is 6 cm 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. 7c. In this case, (10) 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[3] whose world coordinates are (R, Z). This ray is projected to a specific pixel. Thus, each pixel corresponds to a line in (R, Z). A parameter α determines the location along the ray

$$\begin{bmatrix} R(r_i, \alpha) \\ Z(r_i, \alpha) \end{bmatrix} = \mathbf{p}(r_i) + \alpha \mathbf{q}(r_i), \tag{15}$$

where $\mathbf{p}(r_i) = [p_R, p_Z]^T$ are the coordinates of the interface surface and $\mathbf{q}(r_i) = [q_R, q_Z]^T$ is the ray's direction unit vector. From Figs. 2 and 7, note that

3. The azimuthal coordinate is not needed since all the chief rays are meridional rays as defined in [4]. This is true for a camera in air. Since the interface is perpendicular to the optical axis, the chief rays remain meridional.

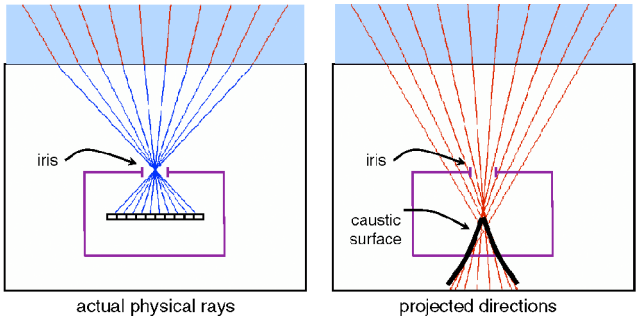


Fig. 8. The creation of the caustic surface. (Left) Actual rays in the imaging system intersect inside the air housing in the center of projection of the perspective camera. (Right) The apparent incident projections do not intersect, but form a bunching which is the caustic surface.

$$\sin\theta_{\text{air}}(r_{\text{i}}) = \frac{r_{\text{i}}}{\sqrt{f^2 + r_{\text{i}}^2}}. \tag{16}$$

Then,

$$p_R = r_{\text{g}}, \qquad p_Z = 0, \tag{17}$$

$$q_R = \frac{\sin\theta_{\text{air}}(r_{\text{i}})}{n} \quad , \quad q_Z = \cos\left\{\sin^{-1}\left[\frac{\sin\theta_{\text{air}}(r_{\text{i}})}{n}\right]\right\}. \tag{18}$$

Thus, the parametric ray representation is

$$R(r_{\text{i}}, \alpha) = \frac{d}{f} r_{\text{i}} + \alpha \frac{1}{n} \sin\theta_{\text{air}}(r_{\text{i}}), \tag{19}$$

$$Z(r_{\text{i}}, \alpha) = \alpha \cos\left\{\sin^{-1}\left[\frac{1}{n}\sin\theta_{\text{air}}(r_{\text{i}})\right]\right\}. \tag{20}$$

The differential change in coordinates from (r_{i}, α) to (R, Z) is expressed by the Jacobian matrix

$$\mathbf{J} = \begin{bmatrix} \left(\frac{\partial p_{\text{R}}}{\partial r_{\text{i}}} + \alpha \frac{\partial q_{\text{R}}}{\partial r_{\text{i}}}\right) & q_{\text{R}} \\ \left(\frac{\partial p_{\text{Z}}}{\partial r_{\text{i}}} + \alpha \frac{\partial q_{\text{Z}}}{\partial r_{\text{i}}}\right) & q_{\text{Z}} \end{bmatrix}. \tag{21}$$

The locus of the singularities in $\mathbf{J}$ represents a surface [6], [12] to which all chief rays are tangent. This is the *caustic* [4]. 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. The creation of the caustic surface is demonstrated in Fig. 8. The physical path of the chief rays (left) intersects at the center of the entrance pupil of the perspective camera, after the refraction. However, there is not a single point in which the original (incident) rays that are imaged intersect (right). Rather, they form a bunching, which is the caustic surface.

To find the caustic surface, we find the points where $|\mathbf{J}| = 0$. Applying this condition to (21) yields

$$\alpha = \frac{\left(q_R \frac{\partial p_Z}{\partial r_{\text{i}}} - q_Z \frac{\partial p_R}{\partial r_{\text{i}}}\right)}{\left(q_Z \frac{\partial q_R}{\partial r_{\text{i}}} - q_R \frac{\partial q_Z}{\partial r_{\text{i}}}\right)}. \tag{22}$$

Using (22) in (15)-(18) yields the caustic coordinates

$$R_{\text{caustic}} = \left(1 - \frac{1}{n^2}\right)\left(\frac{r_{\text{i}}}{f}\right)^3 d, \tag{23}$$

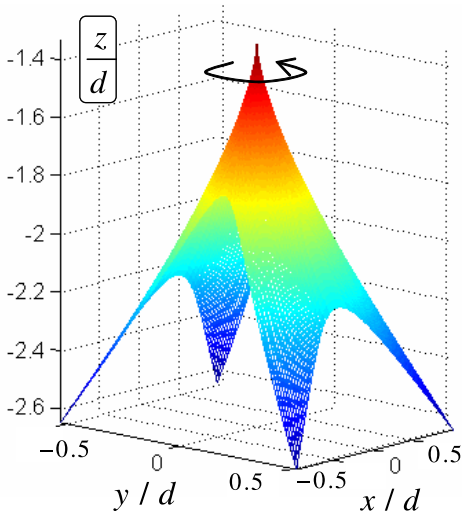


Fig. 9. Caustic of a system having a flat interface with water. The camera has an FOV of $\max(\theta_{\text{air}}) = 50°$. The caustic has radial symmetry which is violated toward the boundaries of the FOV due to the rectangular shape of the sensor. The extent of the caustics is $\mathcal{O}(d)$, and can often reach centimeters or decimeters.

$$Z_{\text{caustic}} = -n\left[1 + \left(1 - \frac{1}{n^2}\right)\left(\frac{r_{\text{i}}}{f}\right)^2\right]^{1.5} d. \tag{24}$$

Obviously, following (23) and (24), the caustic is not a point (unless $d = 0$). Therefore, the system does not have an SVP.

Fig. 9 depicts the caustic in an FOV for which $\max(\theta_{\text{air}}) = 50°$. From (23) and (24), both R_{caustic} and Z_{caustic} depend linearly on d. Therefore, the dimensions in Fig. 9 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 an SVP model more valid.

Equations (23) and (24) depend on the ray slope at the center of projection r_{i}/f rather than the pixel coordinate r_{i} by itself. Therefore, the shape of the caustic does not depend on the location and the position of the sensor. Specifically, even if the camera rotates and the optical axis is not perpendicular to the interface, the caustic shape does not change. This is in compliance with (3), where the ray path from the object to the center of projection depends solely on d and the objection location. Nevertheless, when the optical axis is not perpendicular to the interface, the boundaries of the manifold in Fig. 9 change to fit the projection onto the sensor.

The extent of the caustic in Fig. 9 is in the order of d. For $d = 2$ cm, for example, the viewpoint in the image center is ~ 2 cm apart from the viewpoint at the image corners. Furthermore, when looking at an aquarium, d is on the order of tens of centimeters, leading to a similarly large spreading of the viewpoint locus. Note that reducing $\max(\theta_{\text{air}})$ reduces the extent of the caustic. Suppose that one selects a single point in the caustic to be a representative viewpoint for an SVP model. If the caustic has a large spread, the large spread prevents any such point from yielding a good approximation. Nevertheless, when approximating the system as a perspective system, the effective center of projection lies in the region of the caustic (in [25], this location is termed the *fictive center of projection*). Following (23) and (24), for small incident angles ($r_{\text{i}} \ll f$), $R_{\text{caustic}} = 0$. However, $Z_{\text{caustic}} = -nd$. This means, the

effective center of projection is *never* located in the center of projection of the perspective camera behind the interface.

When the entrance pupil is outside the lens (Section 3.3), the caustic shape is flipped inside out, and the entire wide set of viewpoints is in the medium.

5 Non-SVP Calibration

Calibrating a system which does not have an SVP involves calibrating a *ray map*, i.e., the trajectory of the light ray that each pixel captures. There are nonparametric methods for calibration of such non-SVP systems [12], [34]. However, Ramalingam et al. [34] mention stability difficulties, while the method in [12] may be too complex to implement in a hostile marine environment. Moreover, both methods in [12] and [34] 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.[4] Furthermore, the calibration can be based on two frames.

Based on Section 3, we develop a parametric calibration process for the flat-interface refractive system. The index n is assumed to be known since, according to [26], n varies by less than 3 percent 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 [27]. The same applies to the glass thickness. Usually, it can be measured or known from manufacturer data, but it can be incorporated into the calibration as well.

The embedded camera itself, irrespective of the medium of interest, is an SVP system, modeled as in Section 2.2. The flat refractive system introduces a single additional parameter, d. Overall, the set of parameters of this model is $\Omega = \{d, f, \mathbf{c}, \kappa_1, \kappa_2\}$. A naive calibration scheme involves imaging objects in unknown 3D locations and optimizing the parameters to fit

$$r_{\mathrm{w}} = \frac{d}{f} r_{\mathrm{i}} + \frac{z_{\mathrm{w}}}{\sqrt{\left(fn/r_{\mathrm{i}}\right)^2 + n^2 - 1}}, \tag{25}$$

based on (10). Note that r_{i} is the undistorted image coordinate, obtained from the observed image coordinate using (6) or a similar model (Fig. 3). Therefore, (25) assumes lens distortion has been corrected. This calibration scheme requires additional optimization for extrinsic parameters: the system position and the location of objects. In addition, this calibration scheme is nonlinear. Note that ideally, the calibration approach should minimize the reprojection error, i.e., using (25) in (10) and minimize errors in r_{i}. However, such an approach has to handle the fact that r_{i} in (10) does not have a closed-form solution.

The easiest optimization of (25) would be obtained if several objects exist at known 3D coordinates. However, while the distance z_{w} of the objects can be known, a precise knowledge of their radial coordinate r_{w} is difficult to obtain. Here, we propose a more practical solution that requires little prior information about the objects. Our method works well *in situ*, as we demonstrate in field experiments.

The proposed method requires imaging of a few objects of a known length s_{known} and known range z_{w}. The objects should be positioned frontoparallel to the interface and camera. The range can be obtained by range meters, which are often mounted on underwater vehicles. Alternatively, research divers sometimes achieve constant range by mounting two angled underwater lasers on a camera and fixing them such that they converge to a dot at a specific distance. Acquiring a few objects of a known length is easy by using a few lines on a standard checkerboard calibration target.[5] Identifying the object's extremities, we index these points as $m = 1, 2$. Their corresponding image coordinates are $(r_{\mathrm{i},1}, \phi_1)$ and $(r_{\mathrm{i},2}, \phi_2)$, where ϕ_m denotes the azimuthal coordinate of a point (Fig. 10). The object may appear bent in the frame, due to the distortions. Nevertheless, in the world coordinates, it maintains its straight form.

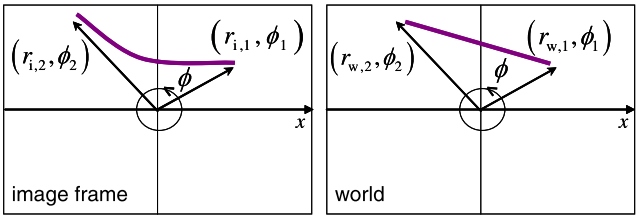


Fig. 10. A straight object is imaged for calibration purposes. The object may appear bent in the raw frame. However, the world coordinates estimated by the model are unaffected.

Based on (25), the corresponding world coordinates $r_{\mathrm{w},1}$ and $r_{\mathrm{w},2}$ should satisfy

$$\hat{r}_{\mathrm{w},m} = \frac{d}{f} r_{\mathrm{i},m} + \frac{z_{\mathrm{w}}}{\sqrt{\left(fn/r_{\mathrm{i},m}\right)^2 + n^2 - 1}}, \tag{26}$$

where $r_{\mathrm{i},m}$ are coordinates after compensating for lens distortion, as in (6). Based on the law of cosines, the object length should satisfy

$$\hat{s} = \sqrt{\left(\hat{r}_{\mathrm{w},1}\right)^2 + \left(\hat{r}_{\mathrm{w},2}\right)^2 - 2\hat{r}_{\mathrm{w},1}\hat{r}_{\mathrm{w},2}\cos\left|\phi_1 - \phi_2\right|}. \tag{27}$$

Following (26)-(27), the calibrated value for $\hat{\Omega}$ is the one that satisfies

$$\hat{s}(\hat{\Omega}) = s_{\mathrm{known}}. \tag{28}$$

Hence, $\hat{\Omega}$ is a set of parameters of the image-formation model that lead to a fit to 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 camera's entrance pupil is in the medium. As (28) is not linear, it is worth starting the optimization from a good initial estimate. An initial estimate for $\hat{f}$ can be obtained by calibrating the system using an SVP model [25], yielding $f_{\mathrm{effective}}$ from (11). Then, $\hat{f} = f_{\mathrm{effective}}/n$ (this is further explained in Section 7.1).

The calibration of this system can be simplified by calibrating the perspective camera in air prior to using it underwater. Then, the inherent properties f, $\mathbf{c}$ and the lens distortion parameters κ_1, κ_2 of the perspective camera are

4. A physics-based parameter calibration was applied to non-SVP catadioptric cameras [42] in air.

5. When having no prior knowledge on the objects, it is common to use a calibration plane having a geometry of the type $\pi_1 x_{\mathrm{w}} + \pi_2 y_{\mathrm{w}} + \pi_3 z_{\mathrm{w}} + \pi_4 = 0$. Then, the calibration process should also solve for the parameters $\Pi = [\pi_1, \pi_2, \pi_3, \pi_4]$, in addition to Ω, using (6) and (25) together with the plane equation.

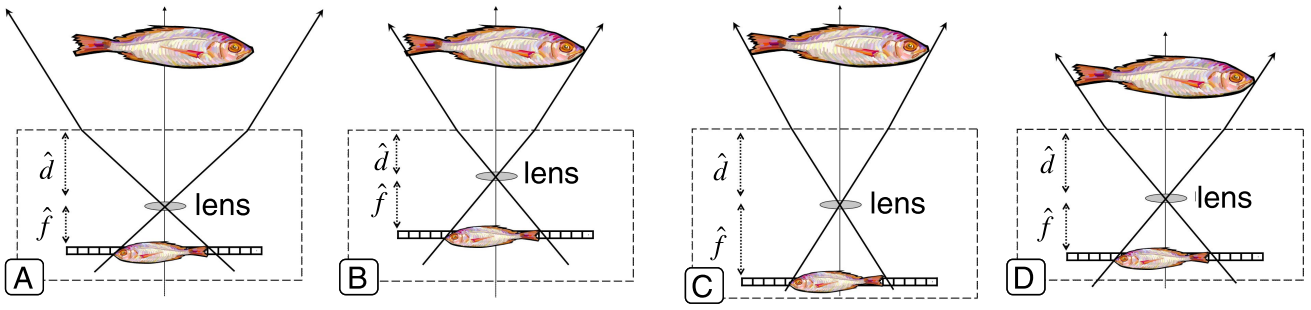


Fig. 11. Calibration using a single object. (A) The fitting problem: an object and its image projection. The initial guess for $\hat{f}$ and $\hat{d}$ does not fit the physical mode. (B), (C) Two possible solution pairs $(\hat{d}, \hat{f})$ for the same object and its corresponding projection. (D) A solution can also be obtained when using a wrong value of z_{w}.

measured a priori, as in [14], [15], and [50]. The only parameter that is left to calibrate in water is d, as done in [22] and [44]. This is convenient, but can be applied only if the zoom or focus settings are *not* changed in situ. Therefore, we do not elaborate on this option here.

5.1 Well Posedness and Stability

How many objects are needed to perform the calibration using (28)? Is their position significant? For the discussion here, let us assume that $\mathbf{c}$ and κ are known and that we only seek $\hat{f}$ and $\hat{d}$. Obviously, using just a single object to assess two continuous parameters is expected to be ill-posed.

Fig. 11A shows a fitting problem: an object and its image projection. The initial guess for $\hat{f}$ and $\hat{d}$ does not fit the physical model. Figs. 11B and 11C show two possible solution pairs $(\hat{d}, \hat{f})$ for the same object and its corresponding projection. Fig. 11D shows that a solution can be obtained also for a wrong value of z_{w}, with different values for $\hat{f}$ and $\hat{d}$. This ambiguity is also seen in (25). There, the estimation of $\hat{d}$ depends on $\hat{f}$; for every value of $\hat{f}$, there is a suitable estimated $\hat{d}$.

Fig. 12A shows projection of three different objects onto the image plane through water. The objects are located at $z_{\mathrm{w}} = 50[\mathrm{cm}]$. Define Ψ_j as the infinite set of possible solution pairs $(\hat{d}, \hat{f})$, derived based on object j. For every object, Ψ_j is a different curve in the $(\hat{d}, \hat{f})$ plane, as shown in Fig. 12B. The curves[6] resemble a straight line but a close look shows that they are not exactly straight, matching (26). As Ψ_j is different for each object j, the solution we seek is $\hat{\Omega} = \cap_j \Psi_j$. Therefore, more than a single object is needed for calibration. However, as seen in Fig. 12B, two objects (#1 and #3) that lie in the center of the FOV generate curves that are very similar and make the calibration ill-conditioned. This can be solved by using an object (#2 in Fig. 12B) that is imaged toward the boundary of the frame. There, refraction increases, making the calibration more sensitive and well conditioned.

An alternative to using multiple calibration objects is to use a single object in multiple known ranges. A simulation of such a situation is demonstrated in Figs. 12C and 12D. The same object is projected to different coordinates when imaged from different distances. This yields two different solution sets $\Psi_{z_{\mathrm{w}}}$. The calibration result is their intersection.

Our calibration method assumes that the objects are placed in parallel to the interface (and the sensor). In practice, there might be some deviations from this assumption. Fig. 13(left) plots the expected error in the calibrated values $\hat{f}$ and $\hat{d}$ as a function of the relative range difference between the extremities of the object, $\Delta z_{\mathrm{w1,2}} = \|z_{\mathrm{w},1} - z_{\mathrm{w},2}\|$. Deviations from the fronto-parallel assumption affect $\hat{f}$ more than $\hat{d}$. Fig. 13(right) plots the expected error in $\hat{f}$ and $\hat{d}$ as a function of the error in absolute range, assuming the object is parallel Δz_{w}. In this case, the value of $\hat{d}$ is more sensitive to error in range measurement.

6. The curves were obtained by calculating $\hat{s}_{\mathrm{calc}}$ for every set of $\hat{d}, \hat{f}$.

6 Experiments

6.1 Calibration

We used a Nikon D100 camera and a Nikkor 24-85 mm 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. 14a. We used a checkerboard pattern as a calibration target (Fig. 14b). Each square is $27.5\ \mathrm{mm} \times 27.5\ \mathrm{mm}$. The known sizes enabled us to select, after image capture, a few lines of known length. Distances to the interface were measured using a simple measuring tape. Then, we applied (26)-(28) on them. We show results

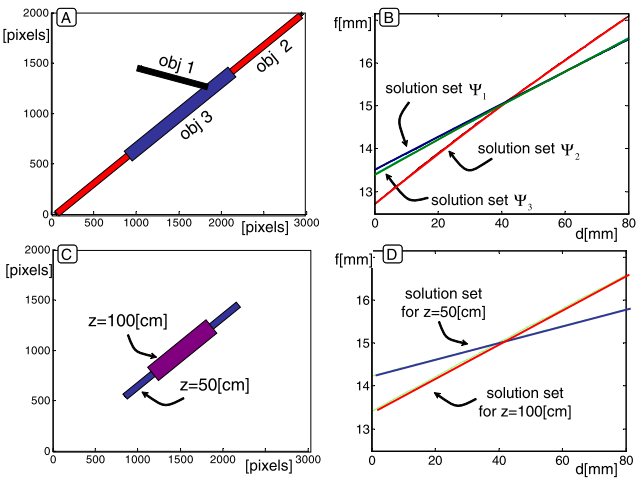


Fig. 12. (A) Projection of three different objects onto the image plane, through water. The objects are located at $z_{\mathrm{w}} = 50[\mathrm{cm}]$. (B) For every object j, the set Ψ_j is a different curve in the $(\hat{d}, \hat{f})$ plane. Thus, using at least two objects, the intersection of the sets provides the calibration result. Two objects (#1 and #3) that lie in the center of the FOV generate curves that are very similar and make the calibration ill-conditioned. This can be solved by using an object (#2) that extends to the extremities of the FOV. An alternative to using two or more calibration objects is using the same object in at least two different known distances. The projected coordinates are shown in (C). This yields (D), two different sets of possible solutions, one per distance. The calibration result lies in their intersection.

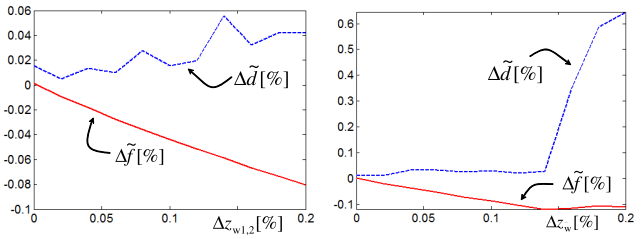


Fig. 13. The expected error in the calibrated values $\hat{f}$ and $\hat{d}$ as a function of the relative range difference between the extremities of the object, $\Delta z_{w1,2}$ (left) and the error in absolute range, assuming the object is parallel (right). The calibration was done by simulating imaging a single object at $z_w = [58\text{ cm}, 88\text{ cm}, 134\text{ cm}]$ with $[d, f] = [45\text{ mm}, 14\text{ mm}]$.

of our method in two sessions. In each session, we used the lens in different zoom and focus settings. In session 1, $z_w = 48$ cm and the calibration yielded $[\hat{d}, \hat{f}] = [7.9\text{ cm}, 24.3\text{ mm}]$. In session 2, $z_w = 78$ cm, and $[\hat{d}, \hat{f}] = [4.2\text{ cm}, 58\text{ mm}]$. We conducted similar experiments in a swimming pool (Fig. 15), in which $[\hat{d}, \hat{f}] = [7.4\text{ cm}, 26\text{ mm}]$. Note that the value of $\hat{d}$ changes significantly when changing the lens setting f.

6.2 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 with known distances that we later measured independently by a ruler. Fig. 14c shows examples of such objects that we measured. On each imaged object, we applied (26) and (27), using the values of $\hat{d}, \hat{f}$, and $\hat{\mathbf{c}}$ calibrated in Section 6.1. To show the generalization of the model, about half of the measured validation objects are at a distance z_w 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.

6.3 Comparison to the Known Art

To demonstrate the significance of our approach, we compare it to the current practice. The current art [11], [29], [31], [32], [39] 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 [5]. This toolbox is based on [15]. It uses calibration images of a checkerboard pattern taken in different positions. We took such underwater images using our system. Then, the standard process [15] obtained the camera parameters of the perspective model, particularly $\hat{f}_{\text{effective}}$ and the presumed radial distortion parameters. In session 2, we got $\hat{f}_{\text{effective}} = 1.3f$. We compensated the images for the estimated radial distortions. Then, we used $\hat{f}_{\text{effective}}$ to find the objects' world coordinates $\hat{r}_w^{\text{perspective}}$ using (30). Using $\hat{r}_w^{\text{perspective}}$ in (27) yielded an estimate for the object length $\hat{s}_{\text{persp}}$, which appears in Table 1 for comparison. Our physics-based model $\hat{s}(\hat{\Omega})$ accounts for the non-SVP nature of the system. It is significantly more accurate: $\hat{s}(\hat{\Omega})$ fluctuates by $\approx$ 1-2 percent around the correct value. In contrast, the error in $\hat{s}_{\text{persp}}$ is $\approx$ 10-15 percent. Moreover, $\hat{s}_{\text{persp}}$ is *biased*. Similar results were obtained in the swimming pool.

7 When does an SVP Approximation Apply?

As discussed in previous sections, the system does not have an SVP and modeling it as a perspective system might yield errors.[7] Thus, Section 5 discussed calibration of the physical model. However, depending on the system and scene parameters, the errors may be small enough and thus negligible in certain applications. Note that although calibration of the accurate model is possible, it might be somewhat more complex than calibrating a perspective model since the physical model has an additional parameter (d). In addition, many existing methods rely on the perspective model. The underlying math, and in particular the perspective epipolar constraint, makes it easy to work with cameras modeled by perspective projection.[8] Thus, there is a trade-off between accuracy and convenience. Suppose we attempt to approximate the system using an SVP model. Such an attempt would yield errors. In this section, we calculate by simulation the expected SVP model errors, for a range of system and scene parameters.

7.1 Aliasing the System as SVP

Sections 3.1 and 3.3 show that when $d \neq 0$, the system does not have an SVP. But, can it be modeled as a perspective projection with radial distortion? Fig. 5 demonstrates why the distortion is not merely radial. Both the square and the round objects are projected into the same coordinate in water. However, in air (undistorted projection), they are projected into different coordinates. Thus, calibrating radial distortion from an object at a distance z_{w1} results in an error when using it to rectify another object, at a different distance z_{w2}. This happens because the distortion depends on the object *distance* in addition to the radial coordinate.

In this paper, we show that a flat refractive system does not have an SVP. Nevertheless, it is possible to treat it as an SVP system and calibrate it using the model in Section 2.2. In fact, this is what has been done in the past. As discussed in Section 4, such an effective center of projection lies in the region of the caustic surface. For every ray in the system

$$f_{\text{effective}}(r_i) = \frac{[z_w - Z_{\text{caustic}}(r_i)][r_i - R_{\text{caustic}}(r_i)]}{[r_w + R_{\text{caustic}}(r_i)]}, \tag{29}$$

where $f_{\text{effective}}$ is different for each value of r_i.

Fig. 16 plots $f_{\text{effective}}$ as a function of r_i/f. As expected, since the system is not really perspective, $f_{\text{effective}}$ is different for different viewpoints. During calibration as an SVP system, we aim to find a single value for the focal length, $\bar{f}_{\text{effective}}$. This focal length would be valid for a certain focal point location $(\bar{R}_{\text{effective}}, \bar{Z}_{\text{effective}})$. As the system has radial symmetry, we expect to get $\bar{R}_{\text{effective}} = 0$, i.e., the optical axis does not move. We denote by $r_i^{\text{perspective}}$ the image coordinate that results from using a perspective projection through $(0, \bar{Z}_{\text{effective}})$ with focal length $\bar{f}_{\text{effective}}$,

$$r_i^{\text{perspective}} = \frac{\bar{f}_{\text{effective}} r_w}{z_w - \bar{Z}_{\text{effective}}}, \tag{30}$$

7. As discussed in Section 4, the error can be reduced by physically minimizing d, i.e., placing the camera close to the interface. However, the location of the actual center of projection along a compound lens is not fixed. Therefore, physical control over d is limited.

8. Recently, Chari and Sturm [7] have developed multiview constraints for the case of two cameras behind a single refractive plane.

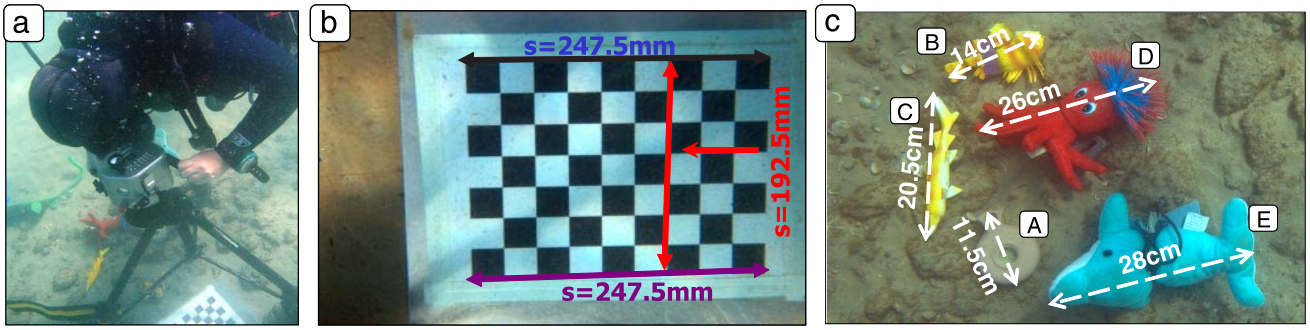


Fig. 14. Calibration and validation experiments done in the ocean while scuba diving. (a) The imaging setup. (b) The checkerboard chart used for *in-situ* calibration. Specific line lengths are marked. (c) Underwater objects are used for validation of the calibration. They were measured independently.

as illustrated in Fig. 3. The resulting focal length of an SVP calibration is a value that minimizes the error in (30) for all points used during calibration. Note that, in water, $n \approx 4/3$; thus, $(n^2-1) \approx 0.77$. At small incident angles, $r_{\mathrm{i}} \ll f$. Consequently,

$$(fn/r_{\mathrm{i}})^2 \gg (n^2-1). \tag{31}$$

Thus, (29) yields $f_{\text{effective}} = f_{\text{effective}}|_{\theta_{\text{air}} \ll 1}$ (11) at the limit of (31).

How does the object distance affect the distortion? To obtain an expression for the distortion, we rewrite (10) to fit the SVP model described in Section 2.2. This means that we assume that the system has an effective focal length $\bar{f}_{\text{effective}}$ (30), and any deviation from the relation in (30) is described by a distortion term.[9]

Recall $r_{\mathrm{i}}^{\text{perspective}}$ from (30). A distortion correction function should relate the refracted image coordinate, r_{i}, to the perspective coordinate, $r_{\mathrm{i}}^{\text{perspective}}$. Plugging r_{w} from (30) into (10) yields the expression for the desired relation, providing the distortion correction function:

$$r_{\mathrm{i}}^{\text{perspective}} = r_{\mathrm{i}}\left[\frac{\bar{f}_{\text{effective}} d}{f(z_{\mathrm{w}} - \bar{Z}_{\text{effective}})} + \frac{\bar{f}_{\text{effective}} z_{\mathrm{w}}}{(z_{\mathrm{w}} - \bar{Z}_{\text{effective}})\sqrt{(fn)^2 + r_{\mathrm{i}}^2(n^2-1)}}\right]. \tag{32}$$

Equation (32) does not depend on the way $\bar{f}_{\text{effective}}$ is assessed. The first term on the left-hand side of (32) shows that the distortion depends on the object distance z_{w}, as demonstrated in Fig. 5. Define the relative distortion as

$$\eta = \frac{r_{\mathrm{i}} - r_{\mathrm{i}}^{\text{perspective}}}{r_{\mathrm{i}}^{\text{perspective}}}. \tag{33}$$

Fig. 17 plots the relative distortion η described by (33) for different values of z_{w}. Here, we used $f = 2,500[\text{pixels}]$, $d = 40[\text{mm}]$, $\bar{f}_{\text{effective}} = fn$, and $\bar{Z}_{\text{effective}} = -nd$ (24). When z_{w} increases, the first term on the left hand-side of (32) decreases until it becomes negligible, practically eliminating the distance dependency. Then, (32) degenerates to (14) and the system becomes practically perspective. This is demonstrated in the following simulations.

9. Recall that the magnification effect caused by the water interface is linear and is expressed by the change in the effective focal length. The distortion we seek to describe here is the deviation from linear magnification.

7.2 Defining the Error

The error in using an SVP model to calibrate a flat refractive system is calculated as follows:

1. Randomly choose 3D coordinates of N_{cal} object points.
2. Project the N_{cal} object points into the image plane using the flat refractive physical model (10).
3. The obtained image coordinates and the known 3D locations are used to perform traditional SVP calibration, by optimization according to the model in Section 2.2.
4. Randomly choose 3D coordinates of a different set of N_{test} object points, each being at a viewing angle θ_{air} (see Fig. 2).

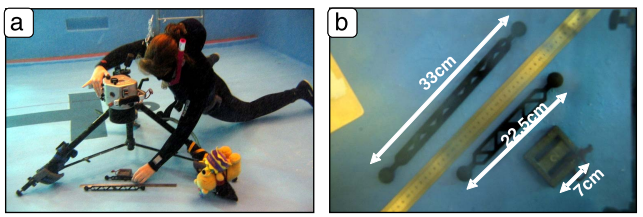


Fig. 15. Calibration experiment done in the pool. (a) The imaging setup. (b) Underwater calibration objects.

TABLE 1

session	object	$z_{\mathrm{w}}^{\text{obj}}$	s_{known}	$\hat{s}(\hat{d},\hat{f},\hat{\mathbf{c}})$	$\hat{s}_{\text{persp}}$
1	A	48	11.5	11.5	8.9
	B	48	14.0	13.3	10.7
	C	48	20.5	20.0	15.7
	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
2	A	153	11.5	11.6	10.4
	B	153	14.0	14.3	12.7
	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

Results of the validation experiments. Units are cm.

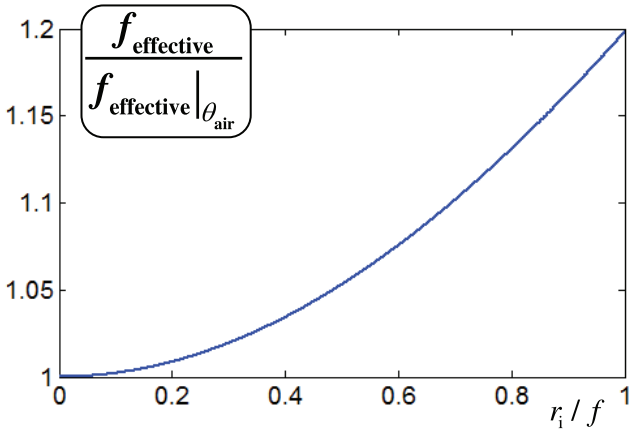


Fig. 16. The ratio of the effective focal length $f_{\text{effective}}$ to $f_{\text{effective}}|_{\theta_{\text{air}}\ll 1}$ as a function of the incidence ray slope in air r_i/f. In practice, the value of $f_{\text{effective}}|_{\theta_{\text{air}}\ll 1}$ does not depend on z_w and has a very small dependency on d. Here, we used $d = 2$ cm.

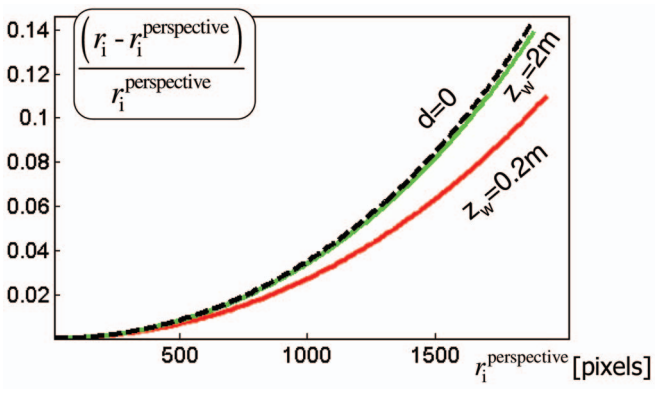


Fig. 17. Apparent (relative) radial distortion η as a function of the object distance and image location. We set $d = 40[\text{mm}], f = 2,500[\text{pixels}]$. There are two interesting things to note here. First, in the case of $d = 0$, i.e., when the system has an SVP, the distortion is more severe than in the case of $d > 0$. Second, as the distance of the object increases, the distortion becomes similar to the SVP case of $d = 0$. Thus, for distant objects, calibration errors stemming from the perspective model are small.

5. Project the new N_{test} 3D locations onto the image plane twice: first, using the SVP calibration results, yielding r_i^{SVP} (see Fig. 3), then, using the correct physical model in (10), yielding r_i^{test}.
6. Calculate the error.

The error is defined as follows:

$$\varepsilon_j(\theta_{\text{air}}) = \left| r_{i,(j)}^{\text{SVP}}(\theta_{\text{air}}) - r_{i,(j)}^{\text{test}}(\theta_{\text{air}}) \right|. \tag{34}$$

Here, j is the experiment index for a certain set of N_{cal} calibration points and N_{test} test points. We repeat each experiment N_{sample} times and then average the error for each value of θ_{air}:

$$\bar{\varepsilon}(\theta_{\text{air}}) = \frac{1}{N_{\text{sample}}} \sum_{j=1}^{N_{\text{sample}}} \varepsilon_j(\theta_{\text{air}}). \tag{35}$$

The maximum error in the entire FOV for a certain setup is

$$\varepsilon_{\max} = \max_{\theta_{\text{air}} \in \text{FOV}} \bar{\varepsilon}(\theta_{\text{air}}). \tag{36}$$

Note that performing step 3 in the real world usually involves estimation of the 3D locations of the objects. In our simulations, we alleviate the need for this and thus assume ideal calibration, in which the 3D locations of the calibration objects are known.

7.3 Simulation Parameters

As seen in (23) and (24), the extent of the caustic increases with the incoming angle. Hence, increasing the in-air FOV makes the SVP approximation less accurate, yielding larger reprojection errors. Therefore, for fair comparison between different focal lengths, we used a varying CCD array size (in pixels) and set a constant maximum FOV $\max(\theta_{\text{air}})$ of the camera *in air*.

To generalize the simulation, we set the parameters relative to the pixel size ρ. As ρ is $\mathcal{O}(1 - 10\ \mu\text{m})$ and f and d are in orders of millimeters, define $\tilde{f} = f/(1000\rho)$ as the relative focal length and $\tilde{d} = d/(1000\rho)$ as the relative distance of the center of projection from the interface. The distance to the objects is in the order of meters; therefore we use normalized units $\tilde{z} = z/(10^5\rho)$. In our simulations, we set the pixel size $\rho = 7.8\,\mu\text{m}$, corresponding to a Nikon D100. Then, the range $\tilde{f}, \tilde{d} \in [1, 8]$ corresponds to $f = 7.8$-62 mm. The range $\tilde{z} \in [0.65, 6.5]$ corresponds to 0.5-5 m. As a result of using normalized units, the results are insensitive to the simulated pixel size. We did not vary the thickness of the interface and set it to 1 cm. The index of refraction of the glass used is $n_{\text{glass}} = 1.46$.

Stable calibration of an SVP model in the presence of distortions as described requires many calibration points. Therefore, in this simulation, we use $N_{\text{cal}} = 50$ points (a number which might be difficult to obtain in situ) for SVP calibration, to achieve low calibration errors. The $N_{\text{test}} = 50$ test points are distributed uniformly across the FOV. We set $N_{\text{sample}} = 100$.

As discussed in Section 7.1, the distortion depends on the object distance. Nevertheless, for each fixed distance, the distortion is transversal. Thus, if the test object distances are kept approximately constant, then the calibration object should be at the same distance, to conform to the transversal distortion model. Such an approach, where all objects share the same distance is feasible if the imaging system has a range meter, such as sonar. For example, Hogue [16] recalibrates the camera whenever the imaging distance changes.

On the other hand, what happens if the distances of the test objects vary? Then, the objects used for SVP calibration should be at a variety of distances as well. Thus, in the simulation, we explore two cases for SVP calibration. First, the distances of the calibration and test objects are relatively uniform. Second, the calibration and object distances vary within each session.

7.4 A Constant Imaging Range

In this section, we explore the errors that result when using a relatively uniform range. We used six distances $\tilde{z} \in [0.65 - 6.5]$. For each value of $\tilde{z}$, the calibration points are distributed around that $\tilde{z}$, with slight perturbations, having a standard deviation (STD) of 5 percent.

Fig. 18 plots $\bar{\varepsilon}(\theta_{\text{air}})$ for the case $\tilde{f} = 6, \tilde{d} = 6$ (corresponding to 39[mm] in a Nikon D100), for different values of $\tilde{z}$ and $\max(\theta_{\text{air}})$. The results are presented for two values of $\tilde{z}$,

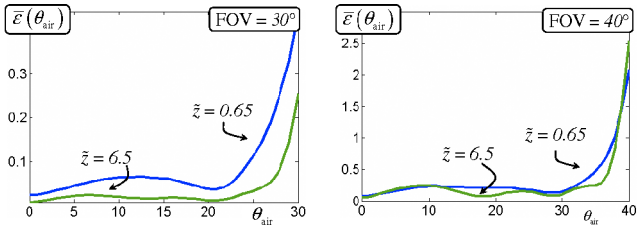


Fig. 18. Reprojection errors $\bar{\varepsilon}(\theta_{\text{air}})$ [pixels] when calibrating and imaging objects at similar distances. We show two cases of two different values of $\tilde{z}$. Here, $\tilde{f} = 6, \tilde{d} = 6$. The error is not very sensitive to the distance. It increases fast at the edge of the FOV.

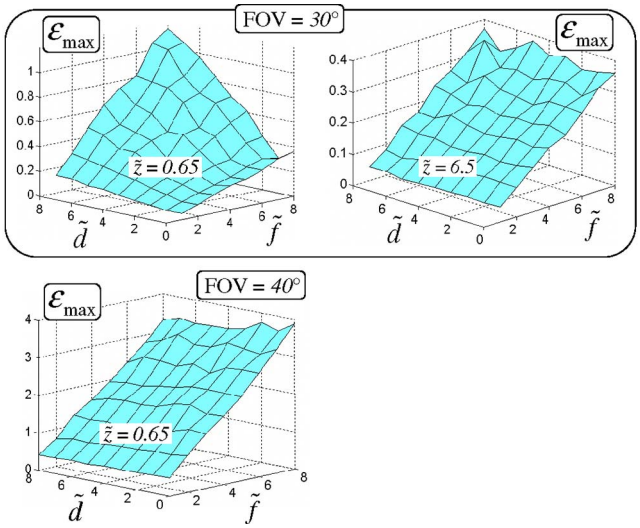


Fig. 19. Reprojection errors $\varepsilon_{\max}$ [pixels], when the calibration objects are in distances similar to those of the imaging objects. At $\max(\theta_{\text{air}}) = 30°$, the errors are relatively small. They decrease with the distance. The error is slightly sensitive to $\tilde{d}$ only at short range. At $\max(\theta_{\text{air}}) = 40°$, the error is insensitive to $\tilde{z}$ and $\tilde{d}$. Decreasing ρ relative to f or increasing $\max(\theta_{\text{air}})$ increases the errors.

since the error is not very sensitive to the distance. The error increases fast at the edge of the FOV.

Fig. 19 plots $\varepsilon_{\max}$ for different values of $\tilde{f}$, $\tilde{d}$, and $\tilde{z}$. At $\max(\theta_{\text{air}}) = 30°$, the errors are relatively small and decrease with the distance (as expected in Section 7.1). At short range, the error increases with $\tilde{d}$. At $\max(\theta_{\text{air}}) = 40°$, the errors are insensitive to $\tilde{z}$ and $\tilde{d}$. Decreasing ρ relative to f or increasing $\max(\theta_{\text{air}})$ increases the errors (in pixels). At $d = 0$, the error is not null. This happens since the perspective distortion function (14) is not polynomial. Thus, the polynomial approximation (6) cannot describe it completely, and this induces errors.

7.5 Unknown Imaging Distances

Section 7.4 showed that calibrating and working at the same distance can reduce the errors significantly. What happens when the SVP calibration distance does not match the distance of the objects? Fig. 20 shows what happens when the SVP calibration is based on objects at a specific distance ($\tilde{z} = 1.3$) but tested on objects at a different distance ($\tilde{z} = 6.5$). The errors in this case are very high. Therefore, if the distance of the objects in the imaging is unknown, SVP calibration should be done on objects at a variety of distances.

As demonstrated in Fig. 20, when the distances of the objects during the imaging session are unknown, calibration

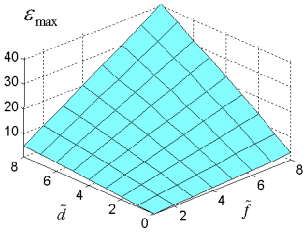


Fig. 20. Reprojection errors $\varepsilon_{\max}$ [pixels] for object distances different than that used during calibration. Calibration was done on objects in $\tilde{z} = 1.3$ and the test objects were placed in $\tilde{z} = 6.5$ (FOV $= 40°$). The errors are large.

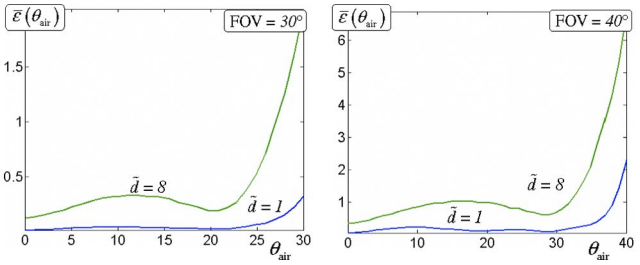


Fig. 21. Reprojection errors $\bar{\varepsilon}(\theta_{\text{air}})$ [pixels] when calibrating and imaging objects at unknown distances. Here, $\tilde{f} = 5$ and $\tilde{d}$ varies. The error increases with $\tilde{d}$.

should use objects that are spread in the entire potential 3D scene domain. In this section, we calculate the errors in this scenario. The calibration points have distances uniformly distributed in the range $\tilde{z} \in [0.65 - 7.8]$. Then, we calculate $\bar{\varepsilon}(\theta_{\text{air}})$ for objects whose distances are also uniformly distributed there. Fig. 21 plots $\bar{\varepsilon}(\theta_{\text{air}})$ for $\tilde{f} = 5$ and different values of $\tilde{d}$ and FOV. Also here, the error increases fast at the edges of the FOV. However, significant errors are also encountered in $\theta_{\text{air}} \in [10 - 20]$. The error increases with $\tilde{d}$. Fig. 22 plots $\bar{\varepsilon}(\theta_{\text{air}})$ for a few setups. The error increases both with $\tilde{d}$, $\tilde{f}$, and the FOV.

7.6 Rotation of the Camera inside the Housing

Until now our analysis focused on a setup in which the camera sensor is parallel to the flat interface. However, slight camera rotation might cause misalignment.[10] Rotation around the optical axis does not effect our analysis as the problem has radial symmetry. However, rotation around a transversal axis means that the sensor plane is not parallel to the plane of the interface. As discussed in Section 4, such a rotation does not change the caustic. However, it effects the projection to image coordinates and thus changes the prior relevant analysis. Now, we try to evaluate the errors that stem from improper alignment of the sensor relative to the interface, as demonstrated in Fig. 23(left). The angle between the sensor and the interface planes is denoted by β. In this case, the errors are not radially symmetric. Fig. 23(right) plots an example of $\bar{\varepsilon}(\theta_{\text{air}})$ for the case of unknown distances. The sensor is tilted around the x-axis. The test points are spread along a line in an angle of $\pi/4$ with the x-axis. Here, $\tilde{f} = 5$, $\tilde{d} = 5$, and $\max(\theta_{\text{air}}) = 40°$. The error relative to $\beta = 0$ increases everywhere, particularly at the edges of the FOV. To decrease these errors, the tilt angle β can be incorporated into the calibration process, as done in [44].

10. Misalignment in omnidirectional cameras was studied in [40].

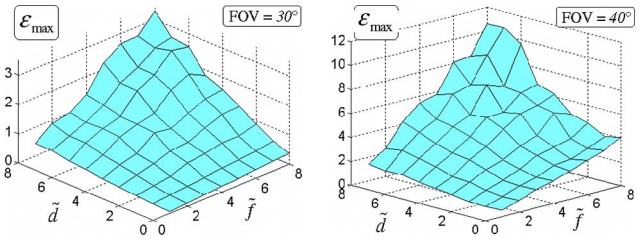


Fig. 22. Reprojection errors $\varepsilon_{\max}$ [pixels] when calibrating and imaging objects at unknown distances. The error increases both with $\tilde{d}, \tilde{f}$, and the FOV.

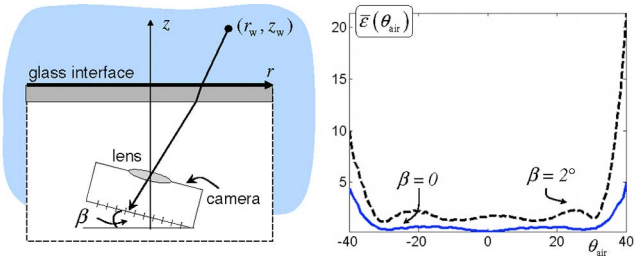


Fig. 23. (Left) The imaging system when the CCD is not parallel to the interface. (Right) Reprojection errors $\bar{\varepsilon}(\theta_{\text{air}})$ in this case. Here, $\tilde{f} = 5$, $\tilde{d} = 5$, FOV $= 40°$, and the distances are unknown. The CCD is tilted around the x-axis. The test points are spread along a line in an angle of $\pi/4$ with the x-axis. Even a small tilt angle $\beta = 2°$ induces large errors. In addition, the errors are asymmetric.

7.7 Conclusions

The errors in the simulation appear small and bearable when the imaging distances are kept constant. Note that Kwon and Casebolt [19] reported appreciable errors when doing geometrical calculations underwater. This might be because the situation in the simulation is simplified: Both the 3D locations and the exact image coordinates of the objects are known. In real life, there is often uncertainty in these attributes. This increases the error. For example, automatic algorithms to identify corners of a checkerboard pattern often fail in the presence of severe distortions. In addition, the simulated camera behind the interface has no lens distortion. The errors stemming from automatic corner detection or lens distortion may accumulate during the process of 3D estimation and reprojection, resulting in errors that are detectable at the end of the process.

Our simulation showed that the errors in the case of a constant imaging range are significantly smaller than in that of unknown distances. Thus, in a constant range, SVP camera models can often be used, as shown in [16]. Increasing $\max(\theta_{\text{air}})$ from 30 to 40 degrees highly increases the errors in both cases. Therefore, the SVP approximation probably should not be used in a wide FOV, e.g., higher than 30 degrees. In addition, as shown in Section 7.6, proper axial alignment of the camera relative to the interface is important. This should be considered versus alignment issues with dome ports when designing a system [18].

8 3D Errors

Until now, we have examined 2D reprojection errors. Nevertheless, using a wrong model has major implications in 3D geometrical applications such as stereo or size measurement. Assume correspondences between pixels are correctly found in a stereo pair [43] taken by a flat refractive system, despite the failure of the perspective epipolar constraint. Then, depth estimates using an SVP approximation are incorrect, as this approximation does not describe properly the pixel-to-ray mapping (15). The reprojection errors in pixels presented in Sections 7.4 and 7.5 can be used to assess 3D errors in other setups, with no need for further simulations.

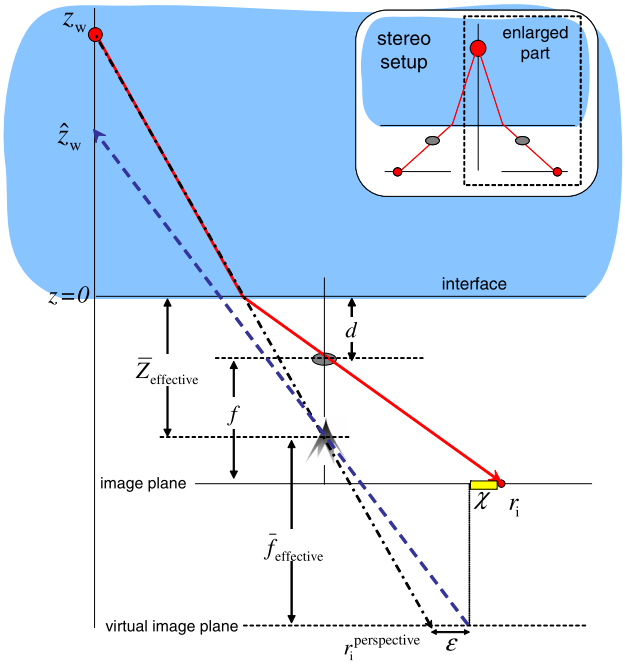


Fig. 24. 3D errors that stem from the errors in the SVP approximation.

Consider Fig. 3. There, a 3D object is imaged using a flat refractive system through a physical pinhole. The SVP approximation projects the same object through the effective pinhole with focal length $\tilde{f}_{\text{effective}}$ to the coordinate $r_i^{\text{perspective}}$. Then, a distortion term χ is added (8), resulting in the SVP image coordinate r_i^{SVP}. The difference between the SVP projection r_i^{SVP} and the actual image coordinate r_i is the error, ε. Fig. 24 demonstrates the inverse process, where image coordinates are projected into 3D space, as done for 3D calculations. The physical image coordinate is compensated for the distortion approximately by χ. Then, the image pixel is backprojected through the effective pinhole. This results in an error since the perspective backprojection was supposed to originate from $r_i^{\text{perspective}}$, approximately ε pixels away. Thus, the error in 3D space stems from the difference between the two dashed rays in Fig. 24.

Let us look, for example, at a stereo setup where both cameras have a symmetric configuration, as shown in Fig. 24. The error in the distance estimation is the difference between the axial coordinates of intersecting corresponding rays. In Fig. 24, the wrongly estimated distance is $\hat{z}_w$, as opposed to true object distance z_w. By simple trigonometry:

$$z_w - \hat{z}_w = \frac{\varepsilon}{\varepsilon + r_i^{\text{perspective}}}(z_w - \bar{Z}_{\text{effective}}). \tag{37}$$

Thus, for example, an error of $\varepsilon = 1$ pixel out of $r = 400$ pixels results in an error of $|\hat{z}_w - z_w| \approx 2.5$ mm at $z_w = 1$ m (for $\bar{Z}_{\text{effective}} \ll 1$ m). This is in agreement with the order of result magnitude in [18].

Stereoscopic distortions caused by refraction are also relevant for biological models. Barta and Horváth [3] and Schuster et al. [38] studied how submerged fish see aerial objects.

9 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 Section 1).

In this paper, we neglected the effect of the lateral shift caused by the glass as it is smaller in magnitude than the shift caused by angular refraction. Nevertheless, as the glass thickness increases, the effect of the shift increases and may become relatively significant. This scenario requires additional analysis.

When attempting 3D stereo reconstruction through a flat interface, 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, maybe using the epipolar constraint for flat-interface systems introduced in [7]. Other possible extensions include self-calibration methods to alleviate the need for a known calibration object. 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.

In addition, Section 7 presents a framework for analyzing the effect of an SVP approximation on a non-SVP system. This framework can be applied on other non-SVP systems. Specifically for underwater applications, it can be used for assessing the errors stemming from a misaligned refractive system having a dome-shaped interface.

Acknowledgments

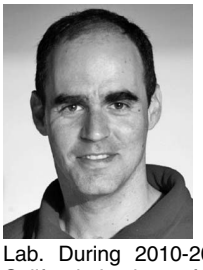
The authors thank the reviewers for their insightful comments and Hank Chezar of the USGS and Boaz Zion of Volcani Center for letting them use their images in Fig. 1. They thank Ben Herzberg and Gal Gur-Arye for help in the experimental dives. Yoav Y. Schechner is a Landau Fellow —supported by the Taub Foundation. This work was supported by the US-Israel Binational Science Foundation (BSF grant 2006384) and the Israeli Ministry of Science, Culture and Sport (Grant 3-3426). This reasearch was supported by the Ollendorff Minerva Center for Vision and Image Science. Minerva is funded through the BMBF. Yoav Y. Schechner was partially supported by US Department of the Navy Grant N62909-10-1-4056 issued by the US Office of Naval Research (ONR) Global and the United States has a royalty-free license throughout the world in all copyrightable materials contained herein. Funding was also provided in part by the CenSSIS ERC of the US National Science Foundation (NSF) under Grant EEC-9986821. Tali Treibitz was partailly supported under grants from NSF ATM-0941760 and ONR N00014-08-1-0638. Tali Treibitz is an Awardee of the Weizmann Institute of Science—National Postdoctoral Award Program for Advancing Women in Science.

References

[1] M. Aggarwal and N. Ahuja, "A Pupil-Centric Model of Image Formation," *Int'l J. Computer Vision,* vol. 48, pp. 195-214, 2002.
[2] S. Baker and S.K. Nayar, "A Theory of Single-Viewpoint Catadioptric Image Formation," *Int'l J. Computer Vision,* vol. 35, pp. 175-196, 1999.
[3] A. Barta and G. Horváth, "Underwater Binocular Imaging of Aerial Objects versus the Position of Eyes Relative to the Flat Water Surface," *J. Optical Soc. Am. A,* vol. 20, pp. 2370-2377, 2003.
[4] M. Born and E. Wolf, *Principles of Optics,* seventh ed., chapters 3 and 4. Cambridge Univ. Press, 1999.
[5] J.Y. Bouguet, "Camera Calibration Toolbox for Matlab," www.vision.caltech.edu/bouguetj/calib_doc, 2011.
[6] D.G. Burkhard and D.L. Shealy, "Flux Density for Ray Propagation in Geometrical Optics," *J. Optical Soc. Am.,* vol. 63, pp. 299-304, 1973.
[7] V. Chari and P. Sturm, "Multiple-View Geometry of the Refractive Plane," *Proc. British Machine Vision Conf.,* 2009.
[8] D.R. Edgington, I. Kerkez, D.E. Cline, J. Mariette, M. Ranzato, and P. Perona, "Detecting Tracking and Classifying Animals in Underwater Video," *Proc. IEEE Computer Vision and Pattern Recognition,* 2006.
[9] J.G. Fryer and C.S. Fraser, "On the Calibration of Underwater Cameras," *The Photogrammetric Record,* vol. 12, no. 67, pp. 73-85, 1986.
[10] G. Glaeser and H.P. Schröcker, "Reflections on Refractions," *J. Geometry and Graphics,* vol. 4, pp. 1-18, 2000.
[11] N. Gracias and J. Santos-Victor, "Underwater Video Mosaics as Visual Navigation Maps," *Computer Vision and Image Understanding,* vol. 79, pp. 66-91, 2000.
[12] M.D. Grossberg and S.K. Nayar, "The Raxel Imaging Model and Ray-Based Calibration," *Int'l J. Computer Vision,* vol. 61, pp. 119-137, 2005.
[13] M. Gupta, S. Narasimhan, and Y.Y. Schechner, "On Controlling Light Transport in Poor Visibility Environments," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision,* second ed., chapter 6. Cambridge Univ. Press, 2003.
[15] J. Heikkila and O. Silven, "A Four-Step Camera Calibration Procedure with Implicit Image Correction," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1106-1112, 1997.
[16] A. Hogue, "SensorSLAM: An Investigation into Sensor Parameter Estimation for SLAM," PhD thesis, York Univ., 2008.
[17] A. Ivanoff and P. Cherney, "Correcting Lenses for Underwater Use," *J. Soc. Motion Picture and Television Eng.,* vol. 69, pp. 264-266, 1960.
[18] C. Kunz and H. Singh, "Hemispherical Refraction and Camera Calibration in Underwater Vision," *Proc. MTS/IEEE Oceans,* 2008.
[19] Y.H. Kwon and J.B. Casebolt, "Effects of Light Refraction on the Accuracy of Camera Calibration and Reconstruction in Underwater Motion Analysis," *Sports Biomechanics,* vol. 5, pp. 315-340, 2006.
[20] J.M. Lavest, G. Rives, and J.T. Lapresté, "Underwater Camera Calibration," *Proc. European Conf. Computer Vision,* pp. 654-668, 2000.
[21] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic Aperture Confocal Imaging," *ACM Trans. Graphics,* vol. 23, pp. 825-834, 2004.
[22] 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.,* vol. 22, no. 2, pp. 364-375, Apr. 1997.
[23] S.M. Luria and J.A. Kinney, "Underwater Vision," *Science,* vol. 167, pp. 1454-1461, 1970.
[24] H.G. Maas, "New Developments in Multimedia Photogrammetry," *Optical 3D Measurement Techniques III,* 1995.

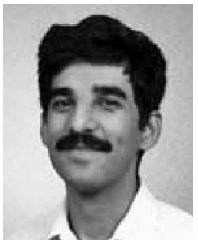
[25] B. Micušík and T. Pajdla, "Structure from Motion with Wide Circular Field of View Cameras," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 7, pp. 1135-1149, July 2006.
[26] C.D. Mobley, *Light and Water: Radiative Transfer in Natural Waters,* chapter 3.6. Academic Press, 1994.
[27] N. Morris and K.N. Kutulakos, "Dynamic Refraction Stereo," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1573-1580, 2005.
[28] 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.,* vol. 31, no. 3, pp. 533-550, July 2006.
[29] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-Acoustic Stereo Imaging System Calibration and 3-D Reconstruction," *Proc. IEEE Beyond Multiview Geometry,* 2007.
[30] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic Stereo Imaging," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 3, pp. 279-290, Mar. 2001.
[31] N. Pessel, J. Opderbecke, and M.J. Aldon, "Camera Self-Calibration in Underwater Environment," *Proc. Int'l Conf. Central Europe Computer Graphics, Visualization and Computer Vision,* 2003.
[32] O. Pizarro, R. Eustice, and H. Singh, "Relative Pose Estimation for Instrumented, Calibrated Imaging Platforms," *Proc. Seventh Digital Imaging Computing, Techniques and Applications Conf.,* pp. 601-612, 2003.
[33] J.P. Queiroz-Neto, R. Carceroni, W. Barros, and M. Campos, "Underwater Stereo," *Proc. Brazilian Symp. Computer Graphics and Image Processing,* pp. 170-177, 2004.
[34] S. Ramalingam, P. Sturm, and S.K. Lodha, "Towards Complete Generic Camera Calibration," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 1093-1098, 2005.
[35] S. Ray, *Applied Photographic Optics,* third ed., chapter 43, pp. 349-352. Focal Press, 2002.
[36] 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,* vol. 202, pp. 402-408, 2007.
[37] Y.Y. Schechner and N. Karpel, "Recovery of Underwater Visibility and Structure by Polarization Analysis," *IEEE J. Oceanic Eng.,* vol. 30, no. 3, pp. 570-587, July 2005.
[38] S. Schuster, S. Rossel, A. Schmidtmann, I. Jäger, and J. Poralla, "Archer Fish Learn to Compensate for Complex Optical Distortions to Determine the Absolute Size of Their Aerial Prey," *Current Biology,* vol. 14, no. 17, pp. 1565-1568, 2004.
[39] M.R. Shortis and E.S. Harvey, "Design and Calibration of an Underwater Stereo-Video System for the Monitoring of Marine Fauna Populations," *Int'l Archives Photogrammetry and Remote Sensing,* vol. 32, pp. 792-799, 1998.
[40] D. Strelow, J. Mishler, D. Koes, and S. Singh, "Precise Omnidirectional Camera Calibration," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2005.
[41] P. Sturm, S. Ramalingam, and S. Lodha, "On Calibration, Structure from Motion and Multi-View Geometry for Generic Camera Models," *Imaging Beyond the Pinhole Camera,* pp. 87-105, Springer, 2006.
[42] R. Swaminathan, M.D. Grossberg, and S.K. Nayar, "Non-Single Viewpoint Catadioptric Cameras: Geometry and Analysis," *Int'l J. Computer Vision,* vol. 66, pp. 211-229, 2006.
[43] Y. Swirski, Y.Y. Schechner, B. Herzberg, and S. Negahdaripour, "Stereo from Flickering Caustics," *Proc. IEEE Int'l Conf. Computer Vision,* 2009.
[44] G. Telem and S. Filin, "Photogrammetric Modeling of Underwater Environments," *ISPRS J. Photogrammetry and Remote Sensing,* vol. 65, pp. 433-444, 2010.
[45] T. Treibitz and Y.Y. Schechner, "Movie," webee.technion.ac.il/people/ttali/flatRefractive.wmv, 2011.
[46] T. Treibitz and Y.Y. Schechner, "Instant 3Descatter," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1861-1868, 2006.
[47] T. Treibitz and Y.Y. Schechner, "Active Polarization Descattering," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 31, no. 3, pp. 385-399, Mar. 2009.
[48] T. Treibitz, Y.Y. Schechner, and H. Singh, "Flat Refractive Geometry," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[49] K. Wolff and W. Förstner, "Exploiting the Multi View Geometry for Automatic Surfaces Reconstruction Using Feature Based Matching in Multimedia Photogrammetry," *Int'l Archives Photogrammetry and Remote Sensing,* vol. 33, pp. 900-907, 2000.
[50] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 11, pp. 1330-1334, Nov. 2000.
[51] B. Zion, V. Alchanatis, V. Ostrovsky, A. Barki, and I. Karplus, "Real-Time Underwater Sorting of Edible Fish Species," *Computers and Electronics in Agriculture,* vol. 56, pp. 34-45, 2007.

Tali Treibitz received the BA degree in computer science and the PhD degree in electrical engineering from the Technion-Israel Institute of Technology in 2001 and 2010, respectively. She is currently a postdoctoral researcher in the Department of Computer Science and Engineering at the University of California, San Diego. She was the recipient of the Google Anita Borg Scholarship in 2009 and the Weizmann Institute of Science National Postdoctoral Award for Advancing Women in Science in 2010. Her research involves physics-based computer vision. She is also an active PADI open water scuba instructor. She is a member of the IEEE.



Yoav Y. Schechner received the BA and MSc degrees in physics and the PhD degree in electrical engineering from the Technion-Israel Institute of Technology in 1990,1996, and 1999, respectively. During 2000-2002, he was a research scientist in the Computer Science Department at Columbia University. Since 2002, he has been a faculty member in the Department of Electrical Engineering at the Technion, where he heads the Hybrid Imaging Lab. During 2010-2011, he has been a visiting associate at the California Institute of Technology. His research is focused on computer vision, the use of optics and physics in imaging and computer vision, and on multimodal sensing. He was the recipient of the Wolf Foundation Award for Graduate Students in 1994, the Gutwirth Special Distinction Fellowship in 1995, the Israeli Ministry of Science (Eshkol) Distinction Fellowship and the Ollendorff Award in 1998, the Schwartz Foundation Award in 1999, and the Morin Fellowship in 2000-2002. He is now a Landau Fellow—supported by the Taub Foundation. He received the Klein Research Award in 2006, the Henry Taub Prize for Academic Excellence in 2008, and Outstanding Reviewer Awards from IEEE CVPR 2007, IEEE ICCV 2007, ECCV 2008, and IEEE ICCV 2009. He is a member of the IEEE.

Clayton Kunz is working toward the PhD degree in the MIT/WHOI joint program in oceanography and applied ocean science and engineering, where he develops autonomous underwater vehicles for scientific applications. Before graduate school, he worked at NASA Ames Research Center on planetary exploration robotics, including the Mars Exploration Rover project. His interests include robot navigation, autonomy, and mapping in support of scientific discovery.



Hanumant Singh received the PhD degree from the MIT-WHOI Joint Program in 1995. He is currently an associate scientist at the Woods Hole Oceanographic Institution. His research interests lie in the area of underwater imaging and underwater robotics. The Seabed class of Autonomous Underwater Vehicles developed under his direction is in use in each of the world's oceans for a variety of oceanographic applications. He is a senior member of the IEEE.