



# Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the SVP camera model

Lai Kang,<sup>1,3,\*</sup> Lingda Wu,<sup>1,2</sup> and Yee-Hong Yang<sup>3</sup>

<sup>1</sup>College of Information System and Management, National University of Defense Technology, Changsha 410073, China

<sup>2</sup>The Key Laboratory, the Academy of Equipment Command and Technology, Beijing 101400, China

<sup>3</sup>Department of Computing Science, University of Alberta, Edmonton, Alberta T6G 2E8, Canada

\*Corresponding author: lkang.vr@gmail.com

Received 6 July 2012; revised 26 September 2012; accepted 26 September 2012;  
posted 27 September 2012 (Doc. ID 172006); published 26 October 2012

In an underwater imaging system, a perspective camera is often placed outside a tank or in waterproof housing with a flat glass window. The refraction of light occurs when a light ray passes through the water-glass and air-glass interface, rendering the conventional multiple view geometry based on the **single viewpoint (SVP) camera model** invalid. While most recent underwater vision studies mainly focus on the challenging topic of calibrating such systems, no previous work has systematically studied the influence of refraction on underwater three-dimensional (3D) reconstruction. This paper demonstrates the possibility of using the SVP camera model in underwater 3D reconstruction through theoretical analysis of refractive distortion and simulations. Then, the performance of the SVP camera model in multi-view underwater 3D reconstruction is quantitatively evaluated. The experimental results reveal a rather surprising and useful yet overlooked fact that **the SVP camera model with radial distortion correction and focal length adjustment can compensate for refraction and achieve high accuracy in multiview underwater 3D reconstruction (within 0.7 mm for an object of dimension 200 mm) compared with the results of land-based systems**. Such an observation justifies the use of the SVP camera model in underwater application for reconstructing reliable 3D scenes. Our results can be used to guide the selection of system parameters in the design of an underwater 3D imaging setup. © 2012 Optical Society of America

OCIS codes: 330.1400, 010.7340.

## 1. Introduction

Image-based three-dimensional (3D) reconstruction has been widely studied during the last two decades [1]. For land-based systems, remarkable success has been achieved [2]. Accurate 3D reconstruction from images captured by underwater cameras, however, has not attracted much attention from the computer vision community until recently [3,4]. **Image-based underwater 3D reconstruction faces challenges** posed by underwater imaging and image distortion caused

by the refraction of light. For underwater imaging, the quality of image degrades due to effects of light attenuation and scattering. In order to improve the quality of underwater images, several algorithms have been proposed in the literature [5,6]. The other major challenge in underwater vision is caused by the refraction of light. In particular, a perspective camera is often placed outside a water-filled tank or in a waterproof housing with a flat glass window. In such an imaging system, the refraction of light occurs when a light ray passes through the water-glass and air-glass interface, rendering the conventional pin-hole camera model invalid in theory. The refractive distortion is known to be highly nonlinear and

depends not only on the distance from the optical axis but also on the depth of a scene point [7].

Recently, there has been a growing interest in using a physically correct refractive camera model in underwater vision. Treibitz *et al.* [7] study the flat refractive geometry and show that **a camera with a flat refractive interface inherently has a non-single viewpoint (SVP) property**. Chari and Sturm [3] analyze theoretically the underlying multiview relationships between cameras of an underwater vision system. The authors show linear relations of such a system with the use of lifted coordinates [8] and demonstrate the existence of geometric entities such as the  $12 \times 12$  refractive fundamental matrix. Unfortunately, no practical applications of these theoretical results are given in [3]. Chang and Chen [4] study a similar configuration involving multiple images of an underwater scene through a single interface. Refractive distortion is explicitly modeled as a function of depth. In their work, an additional piece of hardware called the inertial measurement unit [9] is required to provide the roll and pitch angles of the camera. Also, the normal of the refractive interface is assumed to be known. Based on this additional information, the authors derive a linear solution to the relative pose problem and a closed-form solution to the absolute pose problem. Sedlazeck and Koch [10] study the calibration of housing parameters for underwater stereo camera rigs. Rather than minimizing the reprojection error in image space, the reprojection error on the outer interface plane is minimized by deriving the virtual perspective projection for each 3D point [11]. One issue of this method, as reported in [10], is that the optimization process is time consuming (on the order of 3 h). More recently, Agrawal and Ramalingam [12] show that the underlying geometry of rays in the refractive camera model corresponds to an axial camera [13], based on which a general theory for calibrating such systems is proposed.

In order to analyze the distortions in non-SVP imaging systems, Swaminathan *et al.* [14] introduce a taxonomy of distortions based on the geometry of imaging systems and derive a metric to quantify caustic distortions. The authors also present an algorithm to compute minimally distorted images using simple priors of scene structure. Treibitz *et al.* [7] analyze the refractive distortion and show the error in SVP approximation of a refractive camera using simulations. The analysis of refractive distortion in this paper is similar to that of [7] but more extensive. Specifically, in addition to the simulations performed in [7], we also analyze the SVP approximation when a similarity transformation is applied to the 3D scene, which is more reasonable in the context of metric 3D reconstruction. Another interesting related work shows that spatiotemporal caustic patterns can be exploited for establishing correspondences between two underwater images [15]. It is noteworthy that the influence of refractive distortion on the accuracy of 3D reconstruction is not studied in

[7,14,15]. Besides image-based methods, an active scanning system has also been developed for underwater 3D reconstruction [16].

## 2. Motivations and Contributions

This paper is motivated by the following three observations. First, although it is well known that refractive distortion cannot be modeled by the image space distortion model (e.g., lens radial distortion) in theory [7], the influence of refraction on the accuracy of 3D reconstruction has not been quantitatively evaluated in any previous work. Second, although flat refractive geometry has drawn increasing attention from the computer-vision community in recent years, algorithms explicitly modeling the refraction effect [4,7,10,12] still suffer issues such as efficiency and applicability, which limit their applications in practice. Specifically, existing algorithms are either too time-consuming [10] or require an object with known dimensions [7,11,12,17] or *a priori* information of camera motion [4] for calibration. In comparison, the SVP camera model has been well studied [1,2] and does not suffer any of the above issues. Thus underwater 3D reconstruction based on the SVP camera model would be more practical if the refraction effect could be compensated. Third, essential algorithms, such as fundamental matrix estimation, triangulation [1], multiview stereo [18], etc., involved in underwater 3D reconstruction, have not been well established in the literature. The task of reconstructing an accurate dense 3D model from underwater images based on the refractive camera model remains a challenging one.

Aiming at studying the performance of the SVP camera model in underwater 3D reconstruction, this paper has the following original contributions:

- We provide theoretical analysis of refractive distortion in images and make an important observation that refractive distortion can be well approximated using the SVP camera model after applying a similarity transformation to the 3D scene. This observation justifies the use of the SVP camera model to obtain accurate metric underwater 3D reconstruction up to an arbitrary similarity transformation.
- We quantitatively evaluate the performance of the SVP camera model in underwater 3D reconstruction with respect to several important parameters in an underwater imaging system using both synthetically generated images and real underwater images. The results of underwater 3D reconstruction are compared with the ground-truth 3D model obtained by a land-based system, using a new sparse-to-dense (S2D) point clouds alignment approach particularly suitable for image-based 3D reconstruction.
- The results reveal a rather surprising and useful yet overlooked fact that the SVP camera model with radial distortion correction and focal length adjustment can compensate for refraction and achieve high accuracy in underwater 3D reconstruction (within 0.7 mm for an object of dimension 200 mm) compared with the results of land-based systems.

Such an observation justifies the use of the SVP camera model in underwater applications for reconstructing a reliable 3D scene and also can be used to guide the selection of system parameters in the design of underwater 3D imaging systems.

The remainder of this paper is organized as follows. In Section 3, we analyze the refractive distortion and its SVP approximation in underwater imaging. In Section 4, we present details on the design of experiments. The qualitative and quantitative experimental results are presented in Section 5. Finally, we conclude in Section 6.

### 3. Refractive Distortion and its Approximation

#### A. Characterizing Refractive Distortion

In the presence of a refractive interface, a light ray no longer travels in a straight line, resulting in distortion in the observed images captured using a typical camera [an example is shown in Fig. 1(a)]. As pointed out by Treibitz *et al.* [7], the distortion induced by refraction depends on the position of the object and cannot be modeled as lens radial distortion, which is an image space model. In this subsection, we characterize the refractive distortion with respect to the location of the observed image point and parameters of an underwater imaging system. Without loss of generality, the two-dimensional (2D) case is considered. A 2D illustration of a general underwater imaging system and the description of the corresponding notations are shown in Fig. 1(b) and Table 1, respectively. In this example, the scene is chosen as a straight line for simplicity. Since the parameter vector  $\Theta_n = (n_a, n_g, n_w)^T$  consists of known constant refractive indices of air, glass, and water, the underwater imaging configuration is fully specified by parameters  $\Theta_c = (f, d_a, d_g, \theta_c)^T$  and  $\Theta_o = (d_w, \theta_o)^T$ . In particular, the scene structure of line  $\mathcal{L}$  is determined by the distance  $d_a + d_g + d_w$  and the angle  $\theta_o$  [see Fig. 1(b)]. In order to characterize the refractive distortion in image space, a light ray goes through the camera center and the observed

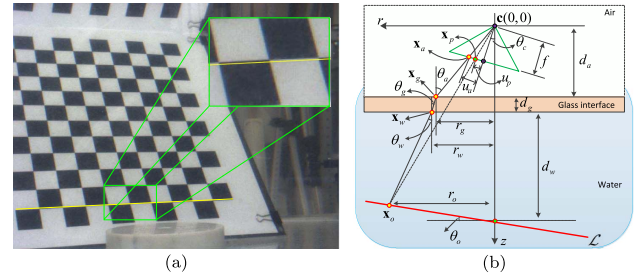


Fig. 1. (Color online) (a) An example showing the refraction effect: an image of a checkerboard submerged in a water-filled tank captured by a camera placed in front of a flat glass interface of the tank. Straight lines are bent in the image due to the refraction of light. (b) 2D illustration of a general underwater imaging system: a camera placed in a waterproof housing with a flat glass interface observing a straight line. See Table 1 for detailed description of notations.

image point is back projected. The light ray is refracted twice (one at the air–glass interface and the other at the water–glass interface) before intersecting with the observed scene at a scene point. Therefore, the derivation of the refractive distortion at an observed image point  $\mathbf{x}_a$  with image coordinate  $u_a$  involves determining three intersection points  $\mathbf{x}_g = (r_g, z_g)^T$ ,  $\mathbf{x}_w = (r_w, z_w)^T$ , and  $\mathbf{x}_o = (r_o, z_o)^T$ , and the projection of  $\mathbf{x}_o$  onto the image plane.

From Fig. 1(b), it is easy to see that the light ray intersects with the air–glass interface at  $\mathbf{x}_g$  with

$$\begin{cases} r_g = d_a \tan \left( \theta_c + \arctan \frac{u_a}{f} \right) \\ z_g = d_a \end{cases} \quad (1)$$

Given  $\mathbf{x}_g$ , the intersection  $\mathbf{x}_w$  between the light ray and the water–glass interface can be determined according to Snell's law [19]:

$$n_a \sin \theta_a = n_g \sin \theta_g, \quad (2)$$

where the trigonometric functions can be calculated as

Table 1. Notations Used in the Underwater Imaging System Shown in Fig. 1(b)

Notation	Description
$\mathbf{c}$	Coordinates of the center of perspective camera: $\mathbf{c} = (0, 0)^T$
$f$	Focal length of the perspective camera and the $z$ -axis
$\mathcal{L}$	Scene structure of a straight line
$d_a$	Distance between the camera and the inner glass interface
$d_g$	Thickness of the glass interface
$d_w$	Distance between the outer glass interface and the intersection between the $z$ -axis and $\mathcal{L}$
$\theta_c$	Angle between the optical axis of the perspective camera and the normal of the glass interface
$\theta_o$	Angle between the glass interface and $\mathcal{L}$
$\theta_w$	Angle between the $z$ -axis and the observed light ray in water
$\theta_a$	Angle between the $z$ -axis and the observed light ray in air
$\mathbf{x}_o$	A scene point lies on $\mathcal{L}$ : $\mathbf{x}_o = (r_o, z_o)^T$
$\mathbf{x}_w$	Intersection between the observed light ray and the outer glass interface: $\mathbf{x}_w = (r_w, z_w)^T$
$\mathbf{x}_g$	Intersection between the observed light ray and the inner glass interface: $\mathbf{x}_g = (r_g, z_g)^T$
$\mathbf{x}_a$	Projection of the scene point $\mathbf{x}_o$ (with image coordinate $u_a$ )
$\mathbf{x}_p$	Projection of the scene point $\mathbf{x}_o$ (with image coordinate $u_p$ ) under the pinhole camera model

$$\begin{cases} \sin \theta_a = \frac{r_g}{\sqrt{r_g^2 + d_a^2}} \\ \sin \theta_g = \frac{r_w - r_g}{\sqrt{(r_w - r_g)^2 + d_g^2}} \end{cases} \quad (3)$$

Combining Eqs. (2) and (3), we get the coordinates of  $\mathbf{x}_w$

$$\begin{cases} r_w = r_g + \frac{n_a r_g d_g}{\sqrt{d_a^2 n_g^2 - n_a^2 r_g^2 + n_g^2 r_g^2}} \\ z_w = d_a + d_g \end{cases} \quad (4)$$

Similar to the way of calculating  $r_w$ ,  $r_o$  can be represented by a function of  $r_w$  by solving the following equations:

$$\begin{cases} n_g \sin \theta_g = n_w \sin \theta_w \\ \sin \theta_w = \frac{r_o - r_w}{\sqrt{(r_o - r_w)^2 + (d_w - r_o \sin \theta_o)^2}} \end{cases} \quad (5)$$

where  $\sin \theta_g$  is the same as in Eq. (3). Then, we get the coordinates of scene point  $\mathbf{x}_o$ :

$$\begin{cases} r_o = A_o \\ z_o = d_a + d_g + d_w - A_o \sin \theta_o \end{cases} \quad (6)$$

where  $A_o$  is the solution of  $r_o$  from Eq. (5). Based on the solution found by the Matlab symbolic toolbox, we have

$$r_o = \frac{r_w B^2 + n_g (d_w - r_w \sin \theta_o) (r_w - r_g) B - d_w n_g^2 (r_w - r_g)^2 \sin \theta_o}{B^2 - n_g^2 (r_w - r_g)^2 \sin^2 \theta_o}, \quad (7)$$

where

$$B = \sqrt{d_g^2 n_w^2 + (n_w^2 - n_g^2) (r_g - r_w)^2}. \quad (8)$$

Under the **pinhole camera model**, the image point of  $\mathbf{x}_o$  is given by

$$u_p = p(\mathbf{x}_o) = f \tan \left( \arctan \left( \frac{r_o}{z_o} \right) - \theta_c \right), \quad (9)$$

where  $p(\cdot)$  represents the function explicitly relating  $\mathbf{x}_o$  and its image point  $u_p$ .

**Definition 1.** In the refractive camera model, the refractive distortion at a given image point is the difference between the projection of the corresponding scene point under the pinhole camera model and the given image point.

According to **Definition 1**, the difference between  $u_a$  and  $u_p$  results in refractive distortion at  $u_a$ , which is

$$E_1(u_a, \Theta_c, \Theta_o) = |u_a - u_p|. \quad (10)$$

Given an image point with coordinate  $u_a$  and parameter vectors  $\Theta_c$  and  $\Theta_o$ ,  $E_1(u_a, \Theta_c, \Theta_o)$  can be computed by sequentially calculating  $r_g$  according to Eq. (1),  $r_w$  according to Eq. (4), and scene point  $\mathbf{x}_o$  according to Eq. (6), and then  $\mathbf{x}_o$  can be projected onto the image plane according to Eq. (9) to obtain  $u_p$ .

Figure 2(a) shows the refractive distortion in an image under five different system configurations. From this figure, we can see that with  $\theta_c = 0$ , different system parameters  $d_a, d_g, d_w$  and  $\theta_o$  preserve the monotonicity of distortion along the radial direction. The refractive distortion under a more general configuration, in which  $\theta_c$  is not equal to zero, is shown in Fig. 2(b) (denoted as original  $E_1$ ). Under this configuration, the distortion becomes much more complex than the previous one.

**Remark 1.** Under general configuration, the refractive distortion in an image is neither symmetric with respect to the image center nor monotonic along the radial direction.

#### B. Approximation Based on the SVP Camera Model

Compared with refractive distortion, lens radial distortion in the SVP camera model is much simpler because it is an image space model. The observed image

point  $u_a$  and its undistorted image point  $u'_p$  are related by [1]:

$$u_a = p'(u'_p) = u'_p \left( 1 + k_1 \left( \frac{u'_p}{f} \right)^2 + k_2 \left( \frac{u'_p}{f} \right)^4 \right), \quad (11)$$

where  $k_1$  and  $k_2$  are two distortion coefficients and  $p'(\cdot)$  represents the function that explicitly relates  $u'_p$  and the observed image point  $u_a$ . Given  $u_a$  and the intrinsic camera parameter vector  $\Theta_s = (f, k_1, k_2)^\top$ ,  $u'_p$  is obtained by solving the following equation:

$$k_2 u_p'^5 + k_1 f^2 u_p'^3 + f^4 u_p' - f^4 u_a = 0. \quad (12)$$

Thus, the lens radial distortion at  $u_a$  can be characterized as

$$E'(u_a, \Theta_s) = |u_a - u'_p|. \quad (13)$$



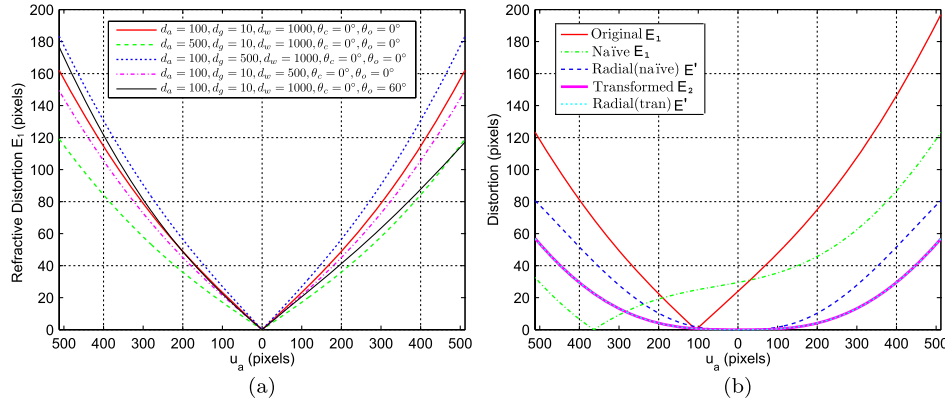


Fig. 2. (Color online) (a) Refractive distortion versus image coordinate under different system configurations. (b) Refractive distortion (denoted as original  $E_1$ ) versus image coordinate under a general configuration ( $d_a = 100$ ,  $d_g = 10$ ,  $d_w = 1000$ ,  $\theta_c = 10^\circ$ ,  $\theta_o = 30^\circ$ ) and approximation to the refractive refraction using two methods based on the SVP camera model. For all imaging setups used in this figure, the focal length  $f$  is 800 pixels and the unit of  $d_a$ ,  $d_g$ , and  $d_w$  is pixels. See text for details.

To this end, both the refractive distortion and the lens radial distortion are characterized with respect to the location of the observed image point and some system parameters. The problem of modeling refraction based on the SVP model corresponds to approximating  $E_1(u_a, \Theta_c, \Theta_o)$  defined in Eq. (10) by  $E'(u_a, \Theta_s)$  defined by Eq. (13). In the following, we analyze three strategies to find such an approximation.  $u_a^i$  ( $1 \leq i \leq M$ ) denotes the coordinates of the  $i$ th image point,  $\mathbf{x}_o^i$  the  $i$ th scene point.

(1) A straightforward way of finding the approximation is to directly search for the best intrinsic camera parameter  $\Theta_s$  by minimizing the following objective function:

$$\arg \min_{\{\Theta_s\}} \sum_{1 \leq i \leq M} |E_1(u_a^i, \Theta_c, \Theta_o) - E'(u_a^i, \Theta_s)|^2. \quad (14)$$

In this paper, we refer to this strategy as the naïve method, in which the scene structure is not allowed to change and only the intrinsic camera parameter  $\Theta_s$  is optimized. The result of this strategy under the general configuration is shown in Fig. 2(b), where naïve  $E_1$  corresponds to the refractive distortion and Radial (naïve)  $E'$  the radial distortion using the optimized parameter  $\Theta_s$ . Obviously, the SVP approximation obtained by this strategy is poor due to the large difference between naïve  $E_1$  and radial (naïve)  $E'$ .

(2) Recall that metric 3D reconstruction is up to an arbitrary similarity transformation [1], so we can apply a similarity transformation to the scene structure before calculating the SVP approximation. This strategy calculates the smallest error in SVP approximation under the assumption of perfect metric 3D reconstruction. It leads to an alternative approximation by minimizing

$$\arg \min_{\{\Theta_s, T_2\}} \sum_{1 \leq i \leq M} |E_2(u_a^i, \Theta_c, \Theta_o) - E'(u_a^i, \Theta_s)|^2, \quad (15)$$

where

$$E_2(u_a^i, \Theta_c, \Theta_o) = |u_a^i - p(T_2(\mathbf{x}_o^i))| \quad (16)$$

is a transformed version of Eq. (10). Here, the function  $p(\cdot)$  is defined by Eq. (9) and  $T_2$  is a similarity transformation, which converts a scene point  $\mathbf{x}_o$  to

$$T_2(\mathbf{x}_o) = s\mathbf{R}_2\mathbf{x}_o + \mathbf{t}_2, \quad (17)$$

where  $s$  is a scale factor,  $\mathbf{R}_2$  is a  $2 \times 2$  rotation matrix, and  $\mathbf{t}_2$  is a  $2 \times 1$  translation vector. Since there exists an ambiguity of an arbitrary similarity transformation in metric 3D reconstruction, the result of this strategy shows the smallest error of SVP approximation, assuming that the 3D reconstruction is accurate. Also, as an example, the result of this strategy under the general configuration is shown in Fig. 2(b), where transformed  $E_2$  corresponds to the optimized refractive distortion and radial (tran)  $E'$  the optimized radial distortion. It is easy to see that the difference between the curves of transformed  $E_2$  and radial (tran)  $E'$  is very small, which means that this strategy could be a good approximation without introducing distortion in the reconstructed 3D scene.

(3) In practice, since both the system parameters and the ground-truth scene structure are not available, we cannot calculate the refractive distortion based on Eq. (10) or apply a similarity transformation to the scene. The success in land-based 3D reconstruction can be attributed to the deployment of bundle adjustment [2,20], in which the scene points are adjusted separately rather than as a group under a similarity transformation. The intrinsic camera parameters and scene structure are simultaneously optimized by minimizing the following objective function:

$$\arg \min_{\{\Theta_s, \mathbf{y}_o\}} \sum_{1 \leq i \leq M} |E_3(u_a^i, \mathbf{x}_o^i) - E'(u_a^i, \Theta_s)|^2, \quad (18)$$

where  $\mathbf{y}_o = \{\mathbf{x}_o^1, \mathbf{x}_o^2, \dots, \mathbf{x}_o^M\}$  is the set of  $M$  scene points and

$$E_3(u_a^i, \mathbf{x}_o^i) = |u_a^i - p(\mathbf{x}_o^i)|. \quad (19)$$

Since the scene is not constrained by a similarity transformation as in the second strategy, even better approximation in terms of the objective function value can be obtained using this strategy. The result of this strategy is not shown in Fig. 2 because the curves almost overlap the results of the second approximation strategy.

**Remark 2.** Refractive distortion can be well approximated by lens radial distortion correction and focal length adjustment under the SVP camera model if a proper similarity transformation is applied to the scene structure. This observation justifies the use of the SVP camera model in metric 3D reconstruction. In practice, the approximation can be even better because the transformation is not constrained to similarity transformation. Although distortion would be introduced in the 3D reconstruction if there is no constraint of the similarity transformation, such distortion is negligible when multiple images are used, as will be demonstrated in our experiments.

#### 4. Experimental Design

In this paper, both synthetically rendered images and real images are used for experiments, where rendered images are captured by virtual imaging setups and real images are captured by real imaging setups in our laboratory. For the virtual imaging setups, the system parameters  $d_a$ ,  $d_g$ ,  $d_w$ , and  $f$  are all measured in pixels. For the real imaging setup,  $d_a$ ,  $d_g$ , and  $d_w$  are measured in millimeters and the focal length  $f$  is measured in pixels. For non-SVP systems, if the size of the object relative to its distance from the camera is too small or if the dimensions in the setup to the focal length are small, then it can be approximated by SVP, which basically means that the locus of caustics is small [14]. Also, if the field of view (FOV) of a non-SVP system is too small, it again can be approximated using SVP. In order to evaluate the performance of SVP-based underwater 3D reconstruction under general imaging setups, we have avoided the above cases in our experimental design. Specifically, both the size of objects and the camera parameters are selected to ensure that the refractive effect is significant and the objects occupy a large portion of FOV. More details regarding this issue are given in the following subsections.

##### A. Imaging Setups

In order to generate realistic underwater images, we use POV-Ray [21] (a publicly available ray tracer) to create a synthetic camera placed in waterproof housing with a flat glass window. The synthetic 3D scene is a texture-mapped bunny [22] sitting on a square base with rounded corners. Two images of the synthetic imaging setup when the pool is empty and when it is filled with water are shown in Figs. 3(a) and 3(b), respectively. The setup in Fig. 3(a) is used to capture the ground-truth images (32 images

sampled on a ring) and the setup in Fig. 3(b) is used to capture underwater images. A detailed structure of each camera in the setup shown in Fig. 3(b) is presented in Fig. 3(c). The focal length of the synthetic camera is 600 pixels and the FOV is about  $56^\circ$ . The resolution of the synthetic image is  $640 \times 480$  pixels. In the simulation, the refractive indices of air, water, and glass are assumed to be  $\Theta_n = (1, 1.3333, 1.5333)^\top$ . The ratio of the object size to its distance from the camera is about 0.56, and the object occupies a large portion of the FOV of the setup (see captured underwater images in Fig. 5).

For capturing real ground-truth images and underwater images, we use two different setups. The first setup consists of a camera and a turntable placed in front of the camera [see Fig. 3(d)]. This setup is used to capture the ground-truth images of 3D objects. In the second setup, we employ a large water-filled tank with multiple flat glass interfaces in our lab. Two imaging methods under this setup are considered: first, a single camera is placed in front of the largest interface of the water tank [see Fig. 3(e)], and multiple views of an object are captured by turning the object; second, an eight-camera array is deployed around the water tank as shown in Fig. 3(f) with each camera placed in front of a separate flat glass interface. All cameras employed in the real imaging setups are the commonly used Point Grey Research Flea2 cameras with a focal length of 1800 pixels and a FOV of  $32^\circ$ . The resolution of the real image is  $1032 \times 776$  pixels. The physical dimension of a pixel is  $4.65 \mu\text{m}$ . Note that while the setup for capturing real underwater images is slightly different from the synthetic one, the underlying refractive geometry is exactly the same except that the Grey Research Flea2 camera has a smaller FOV than the virtual cameras in the synthetic imaging setups. For real experiments, the ratio of the object size to its distance from the camera ranges from 0.23 to 0.41 and the objects occupy nearly the full FOV of the setup (see captured underwater images in Fig. 9). Although the FOV of real cameras is smaller than that used in synthetic experiments, the distortion in 3D reconstruction caused by the refractive effect is significant (see the last row of Fig. 9).

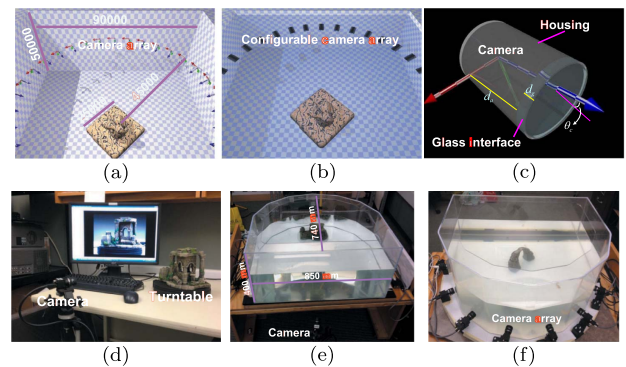


Fig. 3. (Color online) Image acquisition setups used in our study. See text for details.

Table 2. Synthetic Datasets and System Configurations ( $d_a$ ,  $d_g$  are measured in pixels)

Datasets	$N_v$	$d_a$	$d_g$	$\theta_c$
SC <sub>1</sub>	$3 \leq N_v \leq 32$	1000	100	0°
SC <sub>2</sub>	32	$1000 \leq d_a \leq 10,000$	100	0°
SC <sub>3</sub>	32	1000	$100 \leq d_g \leq 5000$	0°
SC <sub>4</sub>	32	1000	1	$0^\circ \leq \theta_c \leq 30^\circ$

## B. Datasets

Using the synthetic datasets allows us to systematically study the influence of system parameters of underwater imaging systems, including the number of images  $N_v$ , the distance between the camera and its refractive interface  $d_a$ , the thickness of refractive interface  $d_g$ , and the angle between the optical axis of the camera and the normal of the glass interface  $\theta_c$ . Note that changing  $d_a$  and  $d_g$  also causes  $d_w$  to change because  $d_a + d_g + d_w$  is fixed at 45,000 pixels. For all synthetic images, the angle between the normal of two adjacent housing interfaces is fixed at 11.25°. We use the notation  $cf(N_v/d_a/d_g/\theta_c)$  to specify the dataset captured under a specific synthetic system configuration. For example,  $cf(16/1000/10/10^\circ)$  refers to the set of 16 images captured with  $d_a = 1000$ ,  $d_g = 10$  and  $\theta_c = 10^\circ$  (length is measured in pixels). The four sets of synthetic datasets are summarized in Table 2.

For the experiments with real images, three aquarium decorations are used, namely Shell, Well, and Arch (see Fig. 4). The four real datasets used in our experiments are listed in Table 3. Specifically, the datasets shell\_Ring, well\_Ring and arch\_Ring are captured by the underwater imaging setup shown in Fig. 3(e) and the images are sampled on a circular ring. The dataset arch\_Sparse consists of eight images captured by the underwater imaging setup shown in Fig. 3(f). In order to evaluate the quality of 3D reconstruction under varying system configurations, different values of system parameters  $d_a$ ,  $d_w$ , and  $\theta_c$  are selected for each dataset. Note that in practice the distance  $d_a$  is normally small and  $\theta_c$  is within a few degrees, so the system configuration for the dataset arch\_Ring (with large  $d_a$  and  $\theta_c$ ) is a more challenging one than for most practical systems. The dataset arch\_Sparse poses challenges as well because the number of images is small. The thickness of the glass interface is around 6 mm for all real datasets.

## C. Metric 3D Reconstruction and Evaluation Methodology

In this paper, we assume that the intrinsic parameters of the cameras are known and underwater 3D reconstruction based on the SVP camera model is performed in two steps. In the first step, the structure and motion (SaM) algorithm [2] is employed to obtain the camera calibration data and sparse 3D point cloud of a scene. Specifically, the above SaM uses bundle adjustment [20] to refine both the camera parameters and scene structure by minimizing reprojection errors like in Eq. (18). For comparison purposes, we investigate the performance of four

different reconstruction strategies, namely NRF, RDist, FAdj, and RDist + FAdj. NRF refers to camera calibration without radial distortion correction or focal length adjustment, RDist refers to camera calibration with radial distortion but without focal length adjustment, FAdj refers to camera calibration with focal length adjustment but without radial distortion correction, and RDist + FAdj refers to using a combination of radial distortion and focal length adjustment in camera calibration. The objective functions used in the above four methods are variants of Eq. (18) by adjusting different numbers of elements of  $\Theta_s$ . Given the calibration data, the multiple-view stereo algorithm proposed by Furukawa and Ponce [23] is performed to reconstruct dense 3D point cloud in the second step. For each underwater dataset, we evaluate the resulting dense 3D point cloud by comparing it with the ground-truth point cloud reconstructed from images captured in the air. For the reconstruction of the ground-truth model, the strategy RDist + FAdj is used.

Unlike existing evaluation benchmarks for 3D reconstruction in land-based systems [18,24], where the cameras are assumed to be calibrated, and thus the reconstructed model and the ground-truth model are already located in the same coordinate system, comparing 3D models obtained in our experiments is more difficult due to the ambiguity of metric 3D reconstruction for uncalibrated image sequences. Denote by  $\mathcal{M} = \{\mathbf{X}_k | 1 \leq k \leq M\}$  the ground-truth model consisting of  $M$  3D points and  $\mathcal{D} = \{\mathbf{Y}_k | 1 \leq k \leq N\}$  the reconstructed model consisting of  $N$  3D points;  $\mathcal{M}$  and  $\mathcal{D}$  need to be aligned for quantitative evaluation. Unfortunately, the most popular point cloud alignment algorithm iterative closest point (ICP) [25] cannot be applied to our data due to the following three reasons: (1)  $\mathcal{M}$  and  $\mathcal{D}$  are both the results of metric 3D reconstruction, each of which is defined up to an arbitrary similarity transformation. Hence we need to estimate not only the rigid transformation but also the scale factor between  $\mathcal{M}$  and  $\mathcal{D}$ . (2) ICP requires a good initialization of alignment that is not available



Fig. 4. (Color online) Sample images of real objects captured in the air. The size ( $L \times W \times H$  mm) of (a) shell is  $165 \times 108 \times 95$ , (b) well is  $229 \times 152 \times 165$ , and (c) arch is  $178 \times 76 \times 165$ .



**Table 3. Real Datasets and the System Configurations**  
( $d_a$  and  $d_w$  are measured in millimeters)

Dataset	Object	$N_v$	$d_a$	$d_w$	$\theta_c$
shell_Ring	Shell	32	10	400	$\approx 0^\circ$
well_Ring	Well	32	10	600	$\approx 0^\circ$
arch_Ring	Arch	32	200	600	$\approx 0^\circ$
arch_Sparse	Arch	8	5	500	$\approx 0^\circ$

in our case. (3) Since  $\mathcal{M}$  and  $\mathcal{D}$  could have only small partial overlay, a large portion of data will be regarded as outliers.

To address the evaluation issue, we develop a new S2D point cloud alignment algorithm particularly suitable for registering point clouds obtained from image-based 3D reconstruction algorithms. The S2D algorithm consists of coarse alignment of the two models using sparse data and alignment refinement using dense 3D point clouds. Here, the sparse data  $\mathcal{M}' = \{\mathbf{X}'_k | 1 \leq k \leq M'\}$  and  $\mathcal{D}' = \{\mathbf{Y}'_k | 1 \leq k \leq N'\}$  refer to the ground truth sparse 3D point cloud and the reconstructed sparse 3D point, respectively. For each 3D point  $\mathbf{X}'$  in the sparse 3D point cloud  $\mathcal{M}'$ , we calculate its feature descriptor as

$$\text{Desc} \cdot (\mathbf{X}') = \frac{1}{N_t} \sum_{i=1}^{N_t} \text{Desc} \cdot (\mathbf{x}_i), \quad (20)$$

where  $\text{Desc} \cdot (\mathbf{x}_i) (1 \leq i \leq N_t)$  is the 128-dimension scale invariant feature transform (SIFT) feature descriptor [26] of the  $i$ th image point used for triangulating  $\mathbf{X}'$  in the SaM algorithm. The feature descriptor for each point in  $\mathcal{D}'$  can be calculated in a similar way. Given the feature descriptors, putative correspondences between  $\mathcal{M}'$  and  $\mathcal{D}'$  are established by performing fast approximate nearest-neighbor search in a high-dimensional feature space [27]. Then, we run three-point absolute orientation with scale estimation [28] using RANSAC to get a robust estimate of the similarity transformation between  $\mathcal{M}'$  and  $\mathcal{D}'$ , which serves as an initial coarse alignment between  $\mathcal{M}$  and  $\mathcal{D}$ . The initial transformation is refined by running a robust ICP alignment algorithm [29] on the dense point clouds. In both the coarse alignment and the alignment refinement, outliers are identified as correspondences whose distances are larger than  $\delta$  ( $\delta = 1\%$  in this paper) times the maximal size of the bounding box  $\text{BB}(\mathcal{M})$  of the ground-truth model. Denoted by  $T_3$  the final similarity transformation, a 3D point  $\mathbf{Y}$  in  $\mathcal{D}$  is transformed as

$$T_3(\mathbf{Y}) = s\mathbf{R}_3\mathbf{Y} + \mathbf{t}_3, \quad (21)$$

where  $s$  ( $s > 0$ ) is the scale factor,  $\mathbf{R}_3$  is a  $3 \times 3$  rotation matrix, and  $\mathbf{t}_3$  is a  $3 \times 1$  translation vector.

Based on the result of accurate alignment between  $\mathcal{M}$  and  $\mathcal{D}$ , the underwater 3D reconstruction is

evaluated by examining the following three quantities.

- The effectiveness of reconstruction is defined as the ratio of the number of 3D points in  $\mathcal{D}$  that have correspondences in the ground-truth model  $\mathcal{M}$  to the total number of 3D points in  $\mathcal{D}$ , i.e.,

$$\text{Eff}(\mathcal{M}, \mathcal{D}) = \frac{|\mathcal{D}''|}{|\mathcal{D}|}, \quad (22)$$

where  $\mathcal{D}'' = \{\mathbf{Y}_k | 1 \leq k \leq N, d(\mathbf{Y}_k, \mathcal{M}) < \delta \cdot \text{BB}(\mathcal{M})\}$ ,  $|\cdot|$  represents the cardinality of a set, and

$$d(\mathbf{Y}_i, \mathcal{M}) = \min_{1 \leq j \leq M} \|\mathbf{T}(\mathbf{Y}_i) - \mathbf{X}_j\|_2 \quad (23)$$

is the minimal distance between the 3D point  $\mathbf{Y}_i$  and each 3D point in  $\mathcal{M}$ .

- The completeness of reconstruction is defined as the ratio of the number of 3D points in  $\mathcal{M}$  that have correspondences in the reconstructed model  $\mathcal{D}$  to the total number of 3D points in  $\mathcal{M}$ , i.e.,

$$\text{Comp}(\mathcal{M}, \mathcal{D}) = \frac{|\mathcal{M}''|}{|\mathcal{M}|}, \quad (24)$$

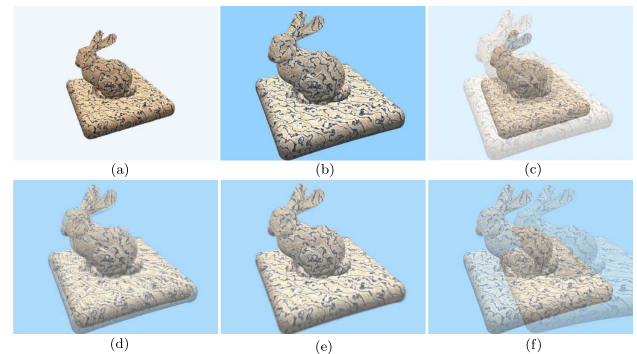
where

$$\mathcal{M}'' = \{\mathbf{X}_k | 1 \leq k \leq M, d(\mathbf{X}_k, \mathcal{D}) < \delta \cdot \text{BB}(\mathcal{M})\}. \quad (25)$$

- The accuracy of reconstruction is defined as the root-mean-square distance of correspondences between  $\mathcal{D}$  and  $\mathcal{M}$ , i.e.,

$$\text{Acc}(\mathcal{M}, \mathcal{D}) = \sqrt{\frac{1}{|\mathcal{D}''|} \sum_{i=1}^{|\mathcal{D}''|} d((\mathcal{D}'')_i, \mathcal{M})^2}, \quad (26)$$

where  $(\mathcal{D}'')_i$  represents the  $i$ th element of  $\mathcal{D}''$ .



**Fig. 5. (Color online) Visual comparison of images captured under different system configurations.** The image in (a) is the image captured in the air. The image in (b) is the underwater image with ( $d_a = 1000$ ,  $d_g = 100$ ,  $\theta_c = 0^\circ$ ). The image in (c) is an overlay of the two images in (a) and (b). An overlay of the image in (b) and the underwater image with ( $d_a = 10,000$ ,  $d_g = 100$ ,  $\theta_c = 0^\circ$ , ( $d_a = 1000$ ,  $d_g = 5000$ ,  $\theta_c = 0^\circ$ ) and  $d_a = 1000$ ,  $d_g = 100$ ,  $\theta_c = 30^\circ$  are shown in (d), (e), and (f), respectively.



## 5. Results

In this section, we present qualitative and quantitative evaluation of the performance of the SVP camera

model in underwater 3D reconstruction. The results on synthetically rendered images and real images are presented in Subsections 5.A and 5.B, respectively.

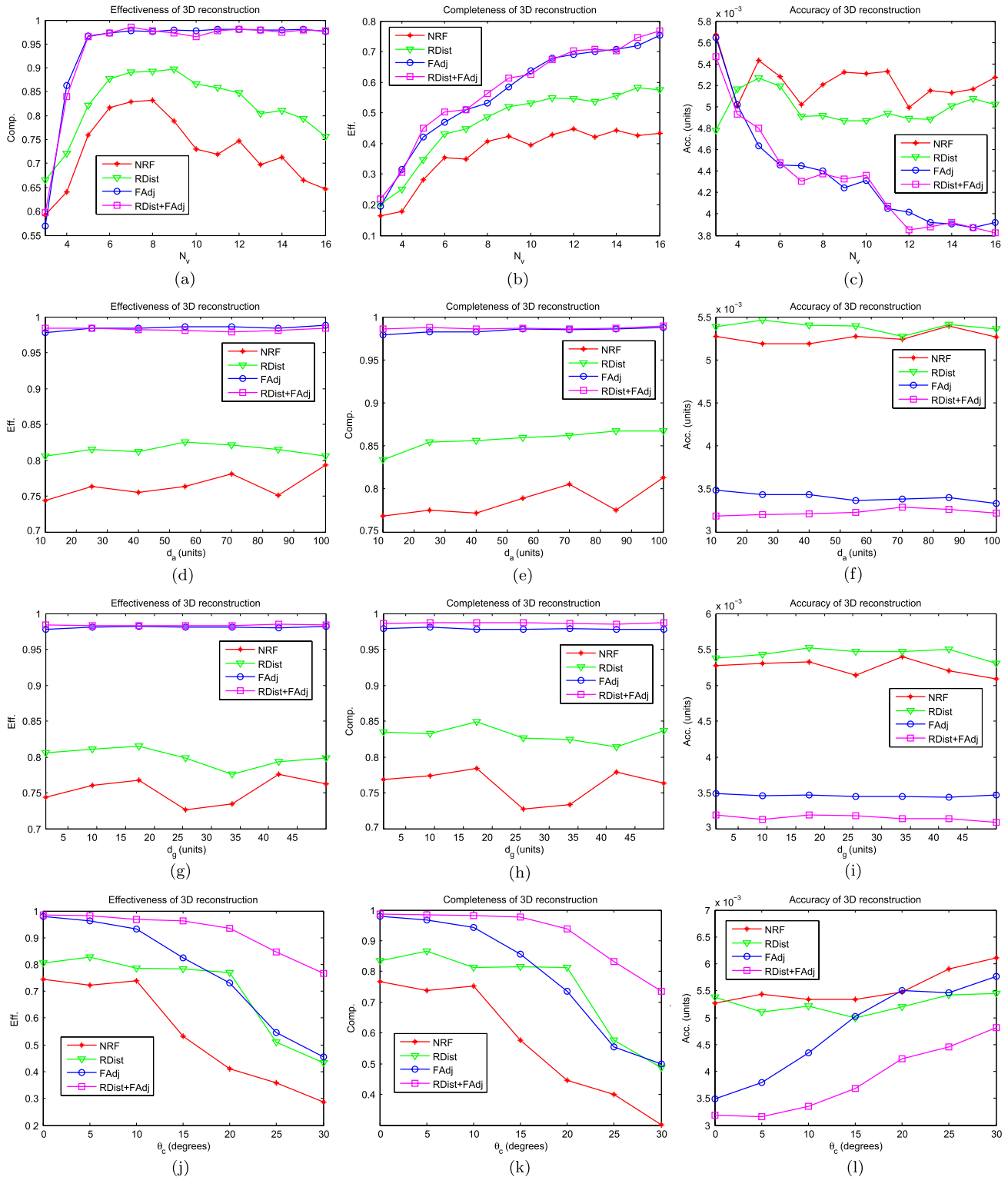


Fig. 6. (Color online) Quantitative evaluation of underwater 3D reconstruction using synthetic datasets listed in Table 2. Shown in (a), (b), and (c) are the results on dataset SC<sub>1</sub>. Shown in (d), (e), and (f) are the results on dataset SC<sub>2</sub>. Shown in (g), (h), and (i) are the results on dataset SC<sub>3</sub>. Shown in (j), (k), and (l) are the results on dataset SC<sub>4</sub>. See text for details.

### A. Synthetic Data

Before presenting the results of 3D reconstruction, we first show a comparison of visual difference between images captured under different system configurations. All the images shown in Fig. 5 are captured by cameras placed at the same position and with the same intrinsic camera parameters. The comparison in Fig. 5(c) shows that an object appears to be closer to the camera and the scene structure is distorted (the edge of the base is bent) when submerged in water due to refraction of light. It also shows that changing the distance between the camera and its refractive interface [see Fig. 5(d)] or changing the thickness of the refractive interface [see Fig. 5(e)] results in similar shift along the radial direction. However, increasing the angle  $\theta_c$  between the optical axis of a camera and the normal of interface causes large distortion in the image [see Fig. 5(f)]. Actually, as will be seen in the following subsections, a large  $\theta_c$  is also the main reason leading to poor underwater 3D reconstruction using the SVP camera model.

Next, we present quantitative results of systematic experiments using the datasets listed in Table 2. In order to facilitate the evaluation of the accuracy of 3D reconstruction, the ground-truth 3D model is scaled such that the maximal size of its bounding box is equal to 1 unit. The first set of experiments studies the influence of  $N_v$  using dataset SC<sub>1</sub> [see Figs. 6(a)–6(c)]. The results show that RDist + FAdj performs significantly better than NRF and RDist in terms of effectiveness, completeness, and also the accuracy of 3D reconstruction. The performance of FAdj on this dataset is close to that of RDist + FAdj. The results also suggest that better results can be obtained by using more images and the SVP camera model can achieve high accuracy of 3D reconstruction (e.g., the 3D error is below 0.004 units for RDist + FAdj using 16 images). The second set of experiments investigates the influence of  $d_a$  using dataset SC<sub>2</sub>. In all comparison, superior performance is achieved by using RDist + FAdj and FAdj [see Figs. 6(d)–6(f)]. Note that although the accuracy of 3D reconstruction using NRF and RDist is

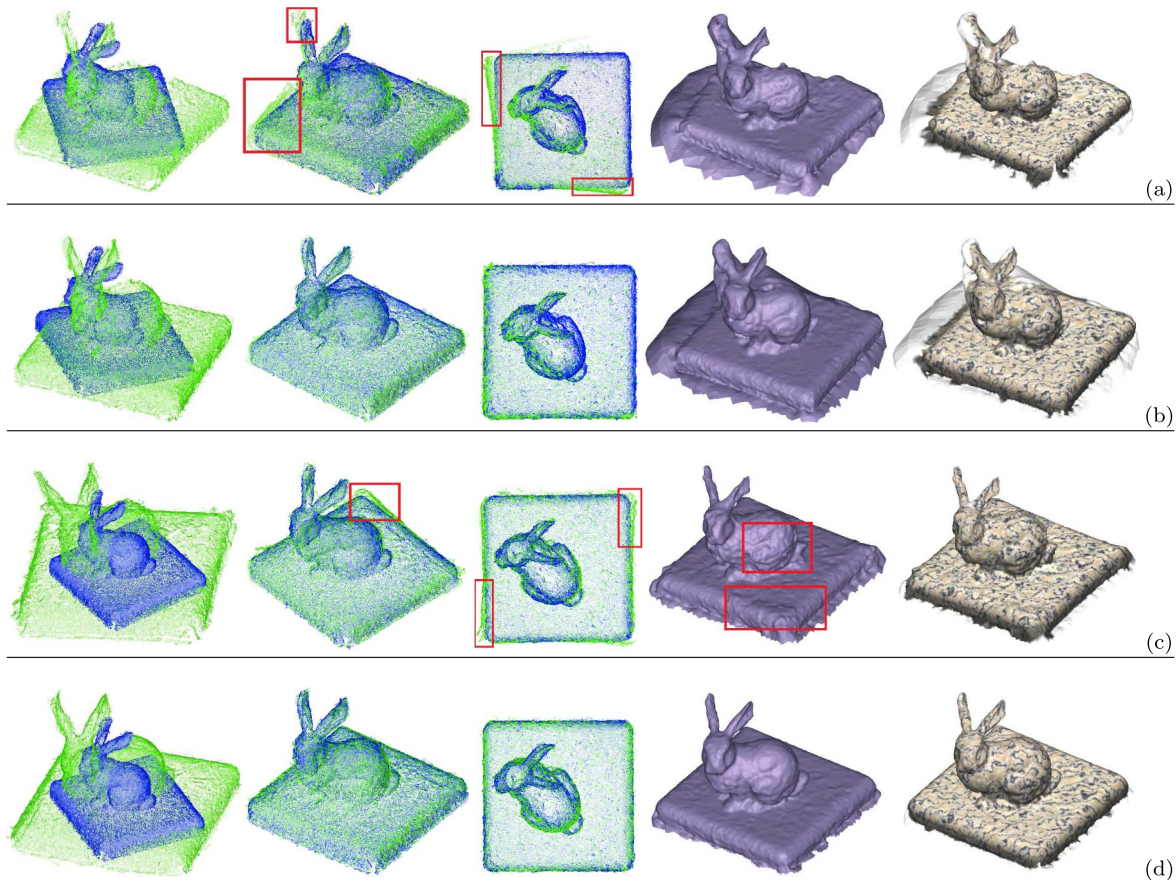


Fig. 7. (Color online) Comparison of 3D reconstruction using different methods. The results of NRF and RDist + FAdj on dataset  $cf(16/1000/100/0^\circ)$  are shown in the first and second rows, respectively. The results of FAdj and RDist + FAdj on dataset  $cf(32/1000/100/15^\circ)$  are shown in the third and last rows, respectively. From left to right: the initial pose of reconstructed point clouds (green) with respect to the ground-truth model (blue); a side view of reconstructed point cloud and the ground-truth model after applying the S2D point clouds alignment algorithm; a top view of reconstructed point cloud and the ground truth model after applying the S2D point clouds alignment algorithm; mesh generated from underwater 3D reconstruction without texture mapping; texture-mapped mesh generated from underwater 3D reconstruction. (Regions of interest are highlighted.)

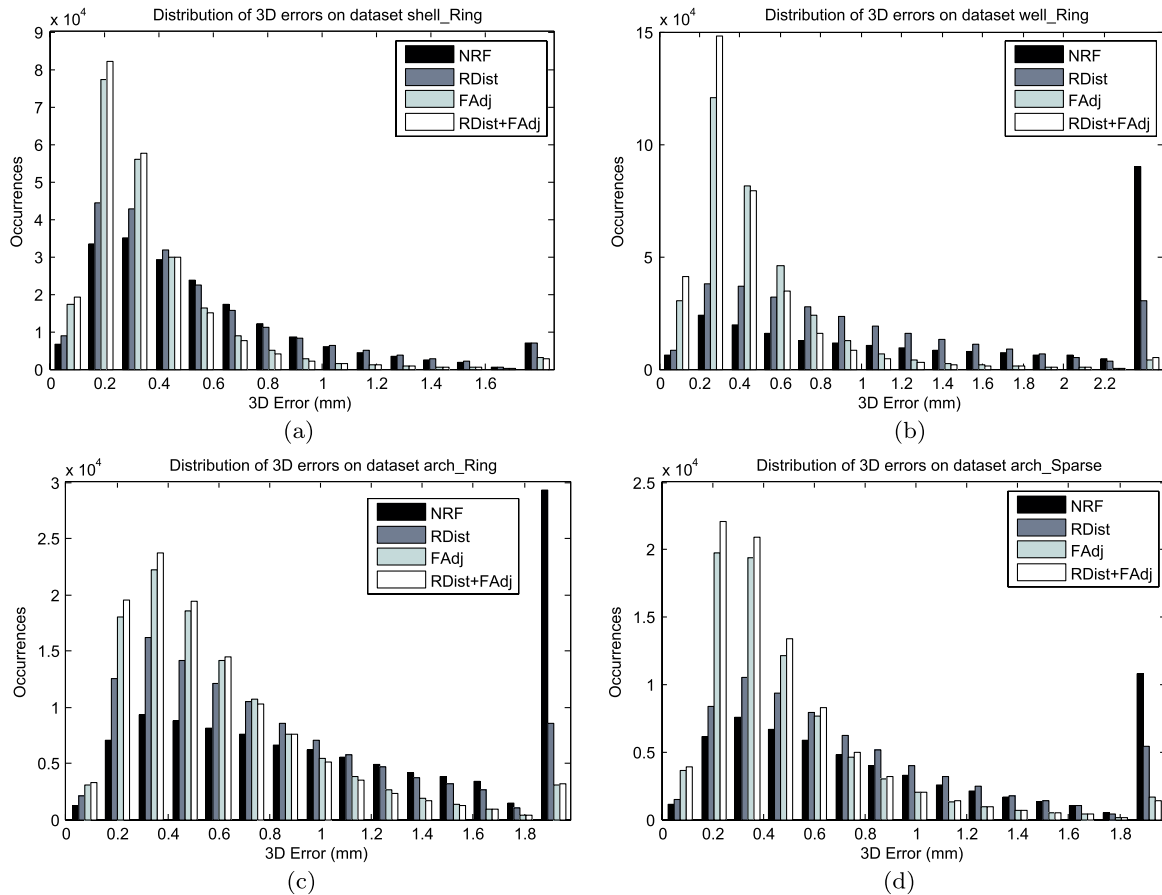
**Table 4. Performance of the SVP Camera Model on Real Images**

Dataset	Method	Eff (%)	Comp (%)	Acc (mm)
shell_Ring	NRF	96.3767	94.5636	0.6409
	RDist	96.6886	96.4833	0.6018
	FAdj	98.6018	98.4560	0.4159
	RDist + FAdj	98.8064	98.6558	0.3875
well_Ring	NRF	62.3939	56.7704	1.1511
	RDist	89.0766	84.1615	1.0287
	FAdj	98.7560	96.7316	0.5721
	RDist + FAdj	98.5369	96.3586	0.5088
arch_Ring	NRF	72.7131	74.3449	0.9570
	RDist	92.3916	93.6748	0.8415
	FAdj	97.3044	96.4522	0.7043
	RDist + FAdj	97.2812	97.4996	0.6839
arch_Sparse	NRF	81.8516	22.5566	0.8822
	RDist	92.1610	29.5537	0.8463
	FAdj	97.9026	32.9819	0.7171
	RDist + FAdj	98.3631	34.0221	0.6915

high (less than 0.006 units), the quality of the resulting 3D model is poor due to its low effectiveness and completeness. Compared with RDist + FAdj, FAdj performs equally well in terms of effectiveness and completeness of 3D reconstruction, while RDist + FAdj gives slightly more accurate results [see Fig. 6(f)]. No apparent influence of  $d_a$  on 3D reconstruction is observed from the results. The third

set of experiments examines the influence of  $d_g$  using dataset SC<sub>3</sub>. The results are similar to those of the previous set of experiments, except that RDist + FAdj consistently achieves about 1% better in completeness of 3D reconstruction than FAdj [see Fig. 6(h)]. Again, no apparent influence of  $d_g$  on 3D reconstruction is observed from the results. The last set of experiments focuses on the influence of  $\theta_c$  using dataset SC<sub>4</sub>. This dataset poses more challenges because severe distortion is caused by increasing  $\theta_c$ . Unlike the results from the previous experiments, the performances of the four reconstruction methods vary significantly from each other [see Figs. 6(j)–6(l)]. Overall, the performance of the four methods all decreases (at different rate) with increasing  $\theta_c$ . However, RDist + FAdj works fairly well when  $\theta_c$  is below 15°, making it a practical one because  $\theta_c$  is within a few degrees if not zero in most underwater imaging systems.

To facilitate qualitative evaluation, some results of underwater 3D reconstruction on two datasets are shown in Fig. 7. For the dataset *cf*(16/1000/100/0°), the point cloud obtained using NRF is very noisy. This can be seen from the highlighted regions in the alignment between the reconstructed point cloud and the ground-truth model [see Fig. 7(a)]. For the dataset *cf*(16/1000/100/0°), FAdj



**Fig. 8.** (Color online) Distribution of 3D errors for reconstruction on real underwater images. The 3D error of a reconstructed 3D point is defined by Eq. (23). Errors larger than 1% of the size of the ground-truth model are collected in the last bin.



also gives poor results as can be seen from the highlighted regions both in the alignment of point clouds and in the reconstructed mesh [see Fig. 7(c)]. In comparison, the results of RDist + FAdj on both datasets are of high visual quality, as shown in Figs. 7(b) and 7(d).

## B. Real Data

With the success of the SVP camera model in underwater 3D reconstruction using synthetic data, next we present results on real datasets listed in Table 3. Since the physical dimensions of the objects used in real datasets are known, we are able to measure the accuracy of 3D reconstruction in millimeters. The results are summarized in Table 4, from which we see that the performance of each reconstruction method varies on each real dataset because refractive distortion depends on the scene structure. Nevertheless, the 3D reconstruction of RDist + FAdj is of high accuracy (ranging from 0.3875 mm for dataset

shell\_Ring to 0.6915 mm for dataset arch\_Sparse). Note that the completeness of 3D reconstruction for dataset arch\_Sparse is low for all four methods. This is because this dataset consists of only eight images and each camera can only see a part of the object in the setup shown in Fig. 3(f). The overall performance of RDist + FAdj is the best among the four methods and the performance of FAdj is close to that of RDist + FAdj, both of which coincide with the results on synthetic datasets. A more detailed analysis of the distribution of errors in 3D reconstruction is shown in Fig. 8. The superior performance of RDist + FAdj is clearly demonstrated, since the 3D errors concentrate in the lower value range for all four real datasets. We also present visualization of 3D reconstruction on real datasets in Fig. 9 for comparison. In all cases, the strategy RDist + FAdj successfully compensates for refraction and reconstructs high-quality 3D models [see Figs. 9(a)–9(d)]. Without radial distortion correction and focal length

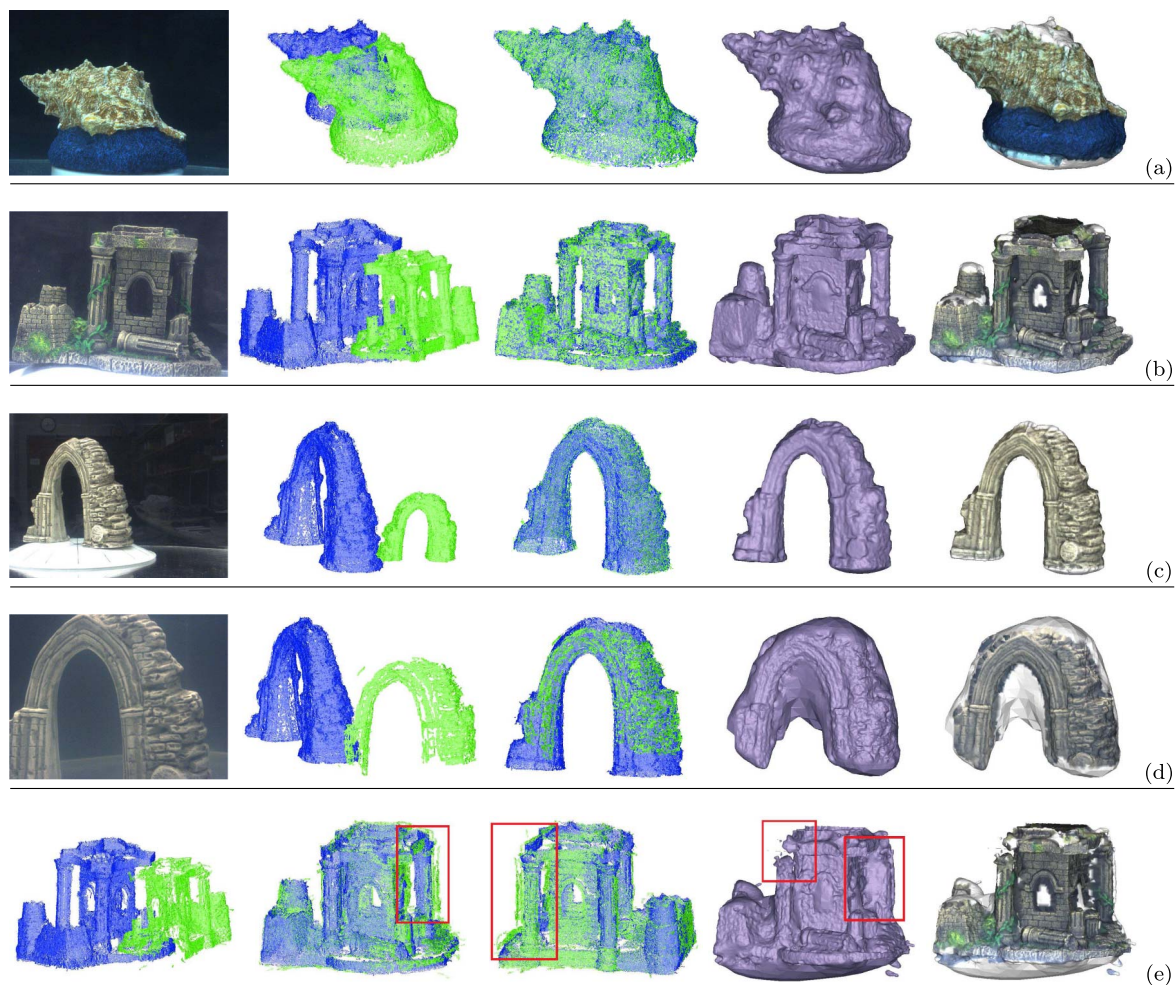


Fig. 9. (Color online) Comparison of 3D reconstruction using different methods. Shown in the first four rows are the results of RDist + FAdj on the dataset shell\_Ring, well\_Ring, arch\_Ring, and arch\_Sparse, respectively. From left to right: a sample underwater image from the dataset; the initial pose of reconstructed point clouds (green) with respect to the ground-truth model (blue); a side view of the reconstructed point cloud and the ground-truth model after applying the new S2D point clouds alignment algorithm; mesh generated from underwater 3D reconstruction without texture mapping; texture-mapped mesh generated from underwater 3D reconstruction. The last row shows the results of NRF on dataset well\_Ring. From left to right: the initial pose of reconstructed point cloud; two images of point cloud alignment; two images of mesh generated from underwater 3D reconstruction. (Regions of interest are highlighted.)



adjustment, refraction causes serve distortion 3D reconstruction as shown in Fig. 9(e), where both the alignment of point clouds and the reconstructed mesh are noisy.

## 6. Conclusions

Refraction of light has been recognized as one of the major challenges in underwater vision. While most existing work on underwater vision focus on calibration, no previous work has systematically studied the influence of refraction on underwater 3D reconstruction. In this paper we provide theoretical analysis of refractive distortion in images and make an important observation that refractive distortion can be well approximated using the SVP model after applying similarity transformation to the 3D scene. This observation indicates the possibility of using the SVP model for accurate underwater metric 3D reconstruction, which is also up to an arbitrary similarity transformation. We then systematically evaluate the performance of the SVP model in underwater 3D reconstruction using both synthetically rendered images and real images. The results reveals a rather surprising and useful yet overlooked fact that the SVP camera model with radial lens distortion correction and focal length adjustment can compensate for refraction and achieve high accuracy in underwater 3D reconstruction. Such an observation justifies the use of the SVP model in underwater application for reconstructing a reliable 3D scene and can also be used to guide the selection of system parameters in the design of underwater 3D imaging systems.

This work was partially supported by the Chinese Scholarship Council (grant no. 2010611068), the Hunan Provincial Innovation Foundation for Postgraduate (grant no. CX2010B025), the Excellent Graduate Student Innovative Project of the National University of Defense Technology, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the University of Alberta. The thank the anonymous reviewers for their constructive comments.

## References

1. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. (Cambridge University, 2004).
2. N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *Int. J. Comput. Vis.* **80**, 189–210 (2008).
3. V. Chari and P. Sturm, "Multiple-view geometry of the refractive plane," in *Proceedings of British Machine Vision Conference* (BMVA, 2009).
4. Y. Chang and T. Chen, "Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction," in *Proceedings of International Conference on Computer Vision* (IEEE, 2011).
5. P. C. Y. Chang, J. C. Flitton, K. I. Hopcraft, E. Jakeman, D. L. Jordan, and J. G. Walker, "Improving visibility depth in passive underwater imaging by use of polarization," *Appl. Opt.* **42**, 2794–2803 (2003).
6. W. Hou, S. Woods, E. Jarosz, W. Goode, and A. Weidemann, "Optical turbulence on underwater image degradation in natural environments," *Appl. Opt.* **51**, 2678–2686 (2012).
7. T. Treibitz, Y. Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 51–65 (2012).
8. P. Sturm and J. P. Barreto, "General imaging geometry for central catadioptric cameras," in *Proceedings of European Conference on Computer Vision* (Springer, 2008), pp. 609–622.
9. Z. Kukelova, M. Bujnak, and T. Pajdla, "Closed-form solutions to minimal absolute pose problems with known vertical direction," in *Proceedings of Asian Conference on Computer Vision* (Springer, 2010), pp. 216–229.
10. A. Sedlazeck and R. Koch, "Calibration of housing parameters for underwater stereo-camera rigs," in *Proceedings of British Machine Vision Conference* (BMVA, 2011), paper 118.
11. G. Telem and S. Filin, "Photogrammetric modeling of underwater environments," *ISPRS J. Photogramm. Remote Sens.* **65**, 433–444 (2010).
12. A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012), pp. 3346–3353.
13. S. Ramalingam, P. Sturm, and S. K. Lodha, "Theory and calibration algorithms for axial cameras," in *Proceedings of Asian Conference on Computer Vision* (Springer, 2006), pp. 704–713.
14. R. Swaminathan, M. D. Grossberg, and S. K. Nayar, "A perspective on distortions," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2003), pp. 594–601.
15. Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "CauStereo: range from light in nature," *Appl. Opt.* **50**, F89–F101 (2011).
16. L. Bartolini, L. De Dominicis, M. Ferri de Collibus, G. Fornetti, M. Guarneri, E. Paglia, C. Poggi, and R. Ricci, "Underwater three-dimensional imaging with an amplitude-modulated laser radar at a 405 nm wavelength," *Appl. Opt.* **44**, 7130–7135 (2005).
17. C. Kunz and H. Singh, "Hemispherical refraction and camera calibration in underwater vision," in *Proceedings of MTS/IEEE Oceans* (IEEE, 2008), pp. 1–7.
18. S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2006), pp. 519–528.
19. G. Glaeser and H.-P. Schrocker, "Reflections on refractions," *J. Geom. Graph.* **4**, 1–18 (2000).
20. B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice* (Springer, 1999), pp. 153–177.
21. <http://www.povray.org>.
22. <http://graphics.stanford.edu/data>.
23. Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 1362–1376 (2010).
24. C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2008), pp. 1–8.
25. P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256 (1992).
26. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**, 91–110 (2004).
27. M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proceedings of International Conference on Computer Vision Theory and Applications* (INSTICC, 2009).
28. B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Am. A* **4**, 629–642 (1987).
29. D. S. D. Chetverikov, D. Svirkov, and P. Krsek, "The trimmed iterative closest point algorithm," in *Proceedings of International Conference on Pattern Recognition* (IEEE, 2002), Vol. 3, pp. 545–548.