

Underwater 3D Surface Capture Using Multi-view Projectors and Cameras with Flat Housings

RYO KAWAHARA^{1,a)} SHOHEI NOBUHARA^{1,b)} TAKASHI MATSUYAMA^{1,c)}

1. Introduction

This paper is aimed at realizing a practical image-based 3D surface capture system of underwater objects. Image-based 3D shape acquisition of objects in water has a wide variety of academic and industrial applications because of its non-contact and non-invasive sensing properties. For example, 3D shape capture of fertilized eggs and young fish can provide a quantitative evaluation method for life-science and aquaculture.

On realizing such a system, we utilize fully-calibrated multi-view projectors and cameras in water (**Fig. 1**). Underwater projectors serve as reverse cameras while providing additional textures on poorly-textured targets. To this end, this paper focuses on the refraction caused by flat housings, while underwater photography involves other complex light events such as scattering [3, 16, 17], specularity [4], and transparency [13]. This is because one of the main difficulties in image-based 3D surface estimation in water is to account for refractions caused by flat housings, since flat housings cause epipolar lines to be curved and hence the local support window for texture matching to be inconstant.

To cope with this issue, we can project 3D candidate points in water to 2D image planes taking the refraction into account explicitly. However, projecting a 3D point in water to a camera via a flat housing is known to be a time-consuming process which requires solving a 12th degree equation for each projection [1]. This fact indicates that 3D shape estimation in water cannot be practical as long as it is done by using the analytical projection computation.

To solve this problem, we model both the projectors and cameras with flat housings based on the pixel-wise varifocal model [9]. Since this virtual camera model provides an efficient forward (3D-to-2D) projection, we can make the 3D shape estimation process feasible.

The key contribution of this paper is twofold. Firstly we propose a practical method to calibrate underwater projectors with flat housings based on the pixel-wise varifocal model. Secondly we show a system for underwater 3D surface capture based on space carving principle [12] using multiple projectors and cameras in water.

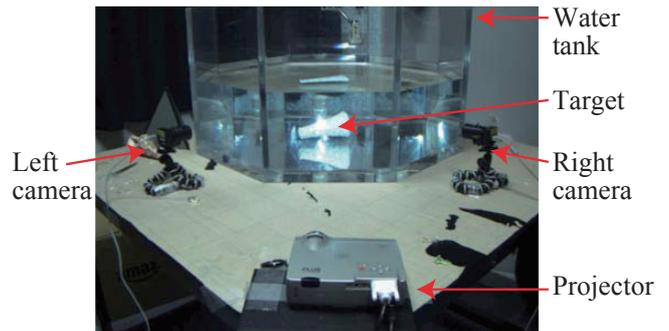


Fig. 1 A projector-camera system for underwater 3D capture. An object in an octagonal water tank is lit by a projector (center) and captured by two cameras (left and right). This is optically-equivalent to capturing the object in water by the projector and cameras with flat housings.

2. Related Work

Image-based 3D shape acquisition has been widely studied in computer vision for objects in the air, such as humans [5] and buildings [7, 20], and in particular algorithms utilizing active illuminations are known to be a practical solution for real environments [6, 10]. However, such methods cannot handle underwater objects since they do not manage refractions caused by flat housings.

Towards realizing 3D shape estimation in water, Sedlazeck and Koch have proposed an underwater structure-from-motion (SfM) technique which accounts for the refraction caused by a flat housing in computing the reprojection error [8]. However SfM requires rich texture on the target in general while underwater objects are not always well-textured. On the other hand, our projector-camera solution can handle such a poorly-textured environment in theory.

In the context of projector-camera calibrations [2, 11, 15, 18, 21], Narasimhan *et al.* have proposed a method for underwater environments which utilizes a pair of flat calibration targets [17], based on an assumption that both of their 3D positions in water are known a priori. On the other hand, our approach does not require providing the pose of the calibration target in water. Hence we can conclude that our method is more simple and practical.

3. Calibration of Underwater Projector-Camera System

3.1 Measurement Model

As shown in **Fig. 2**, suppose we have an underwater projector C and an underwater camera C' . We also have a flat calibration

¹ Graduate School of Informatics, Kyoto University

a) kawahara@vision.kuee.kyoto-u.ac.jp

b) nob@i.kyoto-u.ac.jp

c) tm@i.kyoto-u.ac.jp

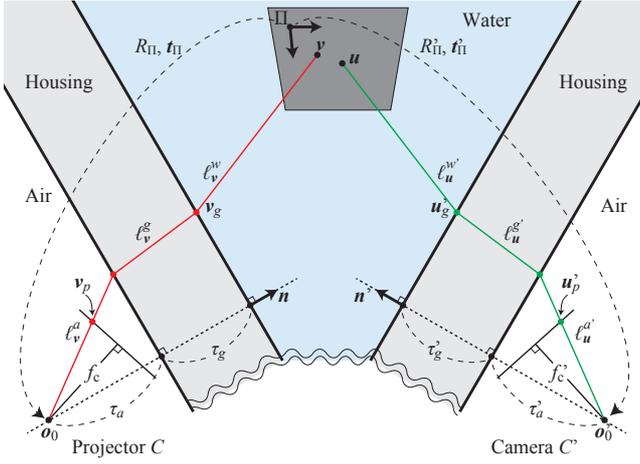


Fig. 2 Measurement model. A projector C and a camera C' observe a point u on a plane Π in water via flat housings.

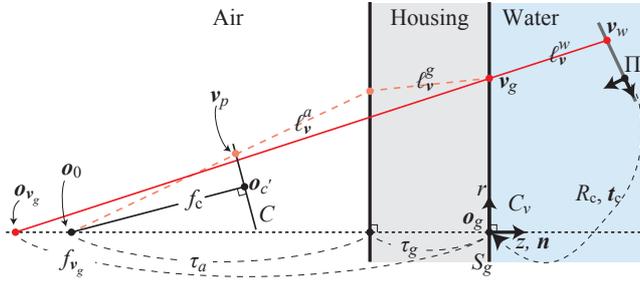


Fig. 3 Pixel-wise varifocal camera / projector model

panel Π in water. Here we assume that a 2D pattern U of known geometry is printed on Π , and the camera C' is calibrated by the pixel-wise varifocal camera model as C_v' beforehand [9] (**Fig. 3**).

Based on the principle of reversibility of light, an underwater projector is equivalent to an underwater camera, and hence we can model the rays emitted by the projector using pixel-wise varifocal lengths (**Fig. 3**). That is, each ray in water is modeled by a virtual pixel v_g on S_g and an associated virtual focal length f_g on the surface normal. We denote this as pixel-wise varifocal projector model (PVPM) hereafter. Notice that we use X' to denote a parameter of PVCM which corresponds to X of PVPM, and x_X , y_X , and z_X to denote the x , y , and z element of a vector X respectively, and d_X^Y to denote the direction of the line ℓ_X^Y .

The goal of this section is,

- calibrating projector C by PVPM as C_v ,
- estimating the relative pose R and t between C_v and C_v' .

For the calibration as done in the air [15], we use a 2D pattern U printed on Π as well as a 2D pattern V projected by C onto Π , and set a model coordinate system $\{K\}$ for using two or more poses of Π . That is, we first estimate the pose R_c', t_c' of C_v' w.r.t. Π by capturing U . Then we project a pattern V on Π by C and estimate the geometry of V on Π in $\{K\}$ by capturing it by C_v' . Lastly we estimate the PVPM parameters of C_v and its pose R_c, t_c w.r.t. Π using V with its estimated geometry on Π . As a result, we obtain the relative pose R, t in $\{K\}$ between C_v and C_v' via Π .

3.2 Pose Estimation of PVCM using a Planner Pattern in Water

Estimation of R_c' and t_c' can be done by using the flat refraction

constraint [1]. That is, the direction $d_u^{w'}$ of the ray $\ell_u^{w'}$ to a known point u of U is identical to the vector from the incident point u_g' to u_w' , where the known point $u = (x_u, y_u, 0)^T$ is described as u_w' in the coordinate system of C_v' .

$$\begin{aligned} d_u^{w'} \times (u_w' - u_g') &= d_u^{w'} \times ((R_c' u + t_c') - u_g') = \mathbf{0}, \\ \Leftrightarrow (x_u [d_u^{w'}]_{\times} \quad y_u [d_u^{w'}]_{\times} \quad [d_u^{w'}]_{\times}) \begin{pmatrix} r_{c,1}' \\ r_{c,2}' \\ t_c' \end{pmatrix} &= [d_u^{w'}]_{\times} u_g', \end{aligned} \quad (1)$$

where $r_{X,i}$ denotes the i th column vector of R_X , and $[X]_{\times}$ denotes the 3×3 skew-symmetric matrix defined by a 3D vector X . Since this equation provides three constraints for 9 unknowns $r_{c,1}'$, $r_{c,2}'$, and t_c' , we can solve this system of equations linearly by using at least three points. Once $r_{c,1}'$ and $r_{c,2}'$ are obtained, $r_{c,3}'$ is given by their cross product.

3.3 3D Geometry Estimation of a Projected Pattern

The next step is to estimate the 3D positions of a projected pattern V on Π in the model coordinate system $\{K\}$, in order to establish 2D-to-3D correspondences between projector pixels and 3D points on Π .

Suppose the projector C casts a known pattern onto Π which consists of feature points such that their projections can be identified in the captured image of C_v' even under refractions. Let v denote a 3D feature point on Π projected from a pixel v_p of C . The goal here is to estimate $v = (x_v, y_v, 0)^T$ from its projection v_g in the camera C_v' image.

Since v is on $\ell_{v_p}^{w'}$, we can represent its 3D position to C_v' with a scale parameter $\lambda_{v'}$ as

$$v = R_c'^T (v_w' - t_c') = R_c'^T (\lambda_{v'} d_{v_p}^{w'} + o_{v_g}' - t_c'). \quad (2)$$

Here we know that $z_v = 0$ because of the fact that V is on Π , and it is trivial to determine the other unknown parameters $\lambda_{v'}$, x_v and y_v . As a result, we obtain 3D points \hat{v} on Π in $\{K\}$ from v .

3.4 PVPM Calibration using 2D-to-3D Correspondences

Up to this point, we obtained a set of correspondences between 2D projector pixels v_p and 3D points \hat{v} on Π in $\{K\}$. In order to estimate the PVPM parameters, we first estimate the pose of the projector R_{Π} and t_{Π} w.r.t. Π in $\{K\}$ using the plane-of-refraction constraint [1]:

$$\begin{aligned} v_p^T (\mathbf{n} \times (R_{\Pi} \hat{v} + t_{\Pi})) &= 0, \\ \Leftrightarrow \begin{pmatrix} x_{v_p} \\ y_{v_p} \\ f_c \end{pmatrix}^T \left(E \begin{pmatrix} x_{\hat{v}} \\ y_{\hat{v}} \\ z_{\hat{v}} \end{pmatrix} + s \right) &= 0, \end{aligned} \quad (3)$$

where \mathbf{n} denotes the housing surface normal described in the local coordinate system of C , $E = \mathbf{n} \times R_{\Pi}$ and $s = \mathbf{n} \times t_{\Pi}$. As proven in [1], E and s are given linearly by using 11 or more points, and we can compute R_{Π} , t_{Π} , \mathbf{n} from E and s . Then τ_a , and τ_g can be computed based on Snell's law.

Once the above parameters are obtained, the pose of PVPM can be given as follows:

$$\begin{aligned} R_c &= (R_h)^T R_{\Pi}, \\ t_c &= (R_h)^T t_{\Pi} - (0, 0, \tau_a + \tau_g)^T, \end{aligned} \quad (4)$$

where R_h is an arbitrary rotation matrix which transforms the z -axis of PVPM to be identical to \mathbf{n} and hence has a 1 degree of

freedom about the rotation around n .

Notice that the 3D points v fed to Equation (3) are not necessarily from a single Π . In fact by capturing the panel Π with different poses, they can cover a larger area of the scene and contribute to improve the accuracy and robustness of the parameter estimation as pointed out in [1].

3.5 Extrinsic Calibration of PVPM and PVCMM

The extrinsic parameter R, t between the PVPM C_v and the PVCMM C_v' is given by

$$R = R_c'(R_c)^\top, \quad t = t_c' - R_c'(R_c)^\top t_c. \quad (5)$$

In addition, we can apply a bundle-adjustment which nonlinearly minimizes the reprojection error using the value by Equation (5) as the initial values.

4. Underwater Space Carving

As described in Section 1, underwater photography causes epipolar lines to be curved and hence simple image-based corresponding point search strategies based on sliding a static template window in the captured images cannot be valid. Besides, the structured light methods utilizing multiple images such as gray-coded patterns cannot handle dynamic targets.

To solve this problem, we employ the space-carving approach [12] which examines the photo-consistency of each voxel by projecting it to the images. While this approach requires computing the forward (3D-to-2D) projections repeatedly in order to account for the refraction explicitly, the PVPM and PVCMM representations solve this difficulty as it realizes an efficient forward computation [9].

In practice, our underwater space carving is conducted as follows. **1.** Capture the background images without the object. **2.** Capture the foreground images with the object while projecting a fixed pseudo-random-dot pattern on it. **3.** Estimate the object silhouettes from each foreground and background images. **4.** Apply the shape-from-silhouette [14] to estimate the rough object volume by the visual hull. **5.** Carve photo-inconsistent voxels from the visual hull.

Notice that we also use the projector as a reverse underwater camera in **5.**, and examines the photo-consistency utilizing the knowledge about the projected pattern.

5. Evaluation

5.1 PVPM Calibration using Synthesized Data

Fig. 4 shows calibration errors before (blue) and after (red) the non-linear optimization of the reprojection error, under different noise levels using a synthesized dataset (**Fig. 5**). Given randomly-synthesized three chessboard patterns V in water, we virtually project the chesscorners to C' and add Gaussian noise with zero-mean and standard deviation $\sigma = 0.0, \dots, 1.0$ pixels to the projections. Then we estimate R_c', t_c', V, R_c and t_c with such noisy 2D points.

The three plots report the average errors of 100 trials at each noise level, where E_{R_c} is the Riemannian distance between R_c and the ground truth, E_{t_c} is the RMS error of t_c normalized by the norm of the ground truth, and E_p is the average reprojection

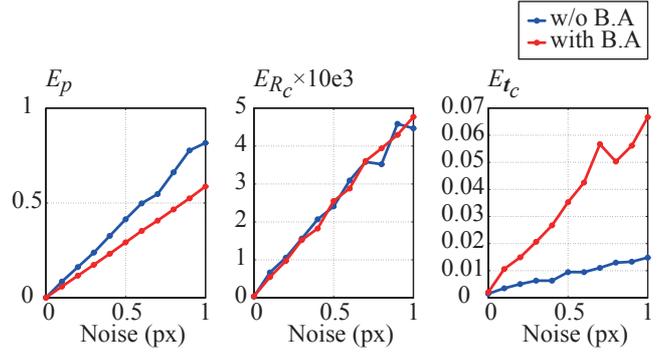


Fig. 4 Calibration errors

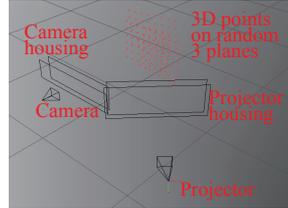


Fig. 5 Synthesized dataset

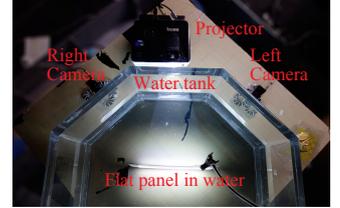


Fig. 6 Top-view of plane object capture setup

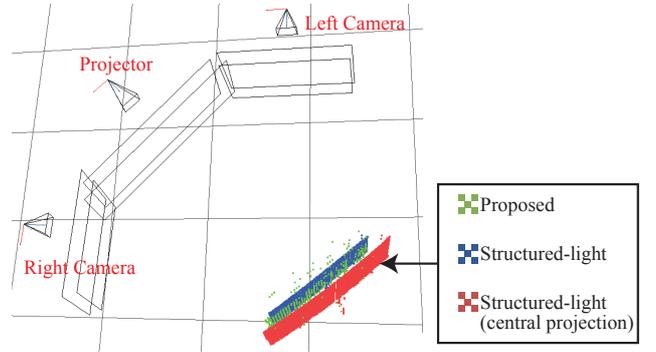


Fig. 7 Reconstructed 3D shapes of plane by a graycode method with pixel-wise varifocal model (blue), the same graycode method with central projection model (red), and our underwater space-carving method (green).

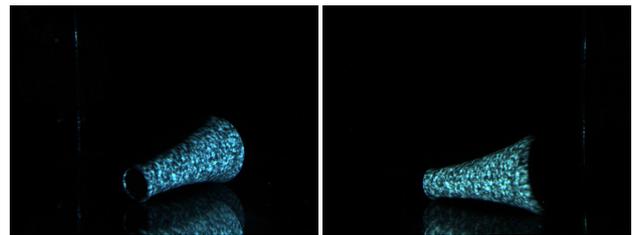


Fig. 8 Captured images of vessel by the left and right cameras

tion error per pixel in the original image plane of the projector C . These results quantitatively demonstrate that our calibration method performs robustly against the observation noise.

Notice that t_c after the optimization is worse than before in the plot (**Fig. 4** right). A possible reason is because both t_c and τ_a can move the camera along the housing surface normal and hence the model is over-parameterized. This point must be verified in future work.

5.2 3D Shape Estimation with Real Data

Fig. 6 shows our underwater projector-camera system for 3D shape capture in water. This system consists of a single projector

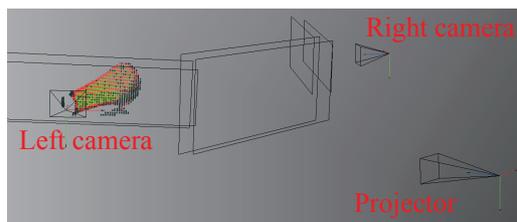


Fig. 9 Reconstructed 3D shape of vessel. The red, green and black points represent outmost, interior, and carved voxels.

and two cameras, and performs as a trinocular active stereo system. Notice that the object is submerged in an octagonal water tank of approximately 90cm width, and the projector and cameras observe the object via the flat acrylic tank surface of about 3.0cm thick. This is optically-equivalent to having an underwater projector and cameras with flat housings of the same thickness.

Fig. 7 shows estimated 3D shapes of the white flat textureless panel in Fig. 6 by three different methods. The blue dots represent a 3D shape (voxels) \hat{S} estimated by a structured light method using graycode images taking account of the refraction into by the housings using the pixel-wise varifocal length model. The red dots represent a 3D shape \tilde{S} by the same structured light method but triangulating the geometry by the central projection model without account for the refraction. The green dots represent the 3D shape S by our underwater space-carving using a single image pair. Notice that S contains only the frontmost voxels of the result of the space carving here.

In this experiment we assume \hat{S} can serve as the ground truth since it has no mismatches in the corresponding point estimation process and models the refraction exactly, and \tilde{S} can be seen as a best estimation except the refraction modeling. By comparing \tilde{S} and S with \hat{S} , we can observe that \tilde{S} is located beyond \hat{S} due to the refraction, while S almost coincide with \hat{S} . This fact can be verified quantitatively by comparing the 90%-accuracy and 1cm-completeness [19] which were 3.88cm and 0.06% for \tilde{S} , and 1.96cm and 92.28% for S respectively. Hence we can conclude that (1) even though the correspondence estimation is correct, the refraction introduces a significant error in the 3D shape, and (2) our method can estimate a shape much closer to the ground truth quantitatively.

In addition **Fig. 9** shows the estimated 3D shape of a white textureless vessel in water (Fig. 1) using an image pair shown in **Fig. 8**. This result also demonstrates qualitatively that our system successfully estimates the 3D shape of a textureless target.

6. Conclusion

Towards realizing a 3D shape capture of underwater objects, this paper proposed an underwater projector-camera system which explicitly handles the refraction caused by the flat housings, and utilizes the projector as a reverse underwater camera. The evaluations quantitatively demonstrated the robustness of our underwater projector-camera system calibration and improvements from the baseline method which does not account for the refraction.

We believe this work is to be one step closer to realizing a practical 3D sensing of underwater objects. Our future work includes full 3D capture, 3D motion estimation, and extension to

semi-transparent targets.

References

- [1] Agrawal, A., Ramalingam, S., Taguchi, Y. and Chari, V.: A theory of multi-layer flat refractive geometry, *Proc. CVPR*, pp. 3346–3353 (2012).
- [2] Audet, S. and Okutomi, M.: A user-friendly method to geometrically calibrate projector-camera systems, *Proc. PROCAMS*, pp. 47–54 (2009).
- [3] Drews Jr., P., do Nascimento, E., Moraes, F., Botelho, S. and Campos, M.: Transmission estimation in underwater single images, *In Proc. of ICCV 2013 Underwater Vision Workshop*, pp. 825–830 (2013).
- [4] Evrenci, D., Iiyama, M., Funatomi, T. and Minoh, M.: Shape and reflectance from scattering in participating media, *Proc. 3DV*, pp. 183–190 (2013).
- [5] Furukawa, Y. and Ponce, J.: Accurate, dense, and robust multi-view stereopsis, *Proc. CVPR*, pp. 1–8 (2007).
- [6] Furuse, T., Hiura, S. and Sato, K.: 3-D shape measurement method with modulated slit light robust for interreflection and subsurface scattering, *Proc. PROCAMS*, pp. 1–2 (2009).
- [7] Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M. and Okamoto, Y.: The great buddha project: digitally archiving, restoring, and analyzing cultural heritage objects, *IJCV*, Vol. 75, pp. 189–208 (2007).
- [8] Jordt-Sedlazeck, A. and Koch, R.: Refractive structure-from-motion on underwater images, *Proc. ICCV*, pp. 57–64 (2013).
- [9] Kawahara, R., Nobuhara, S. and Matsuyama, T.: A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings, *In Proc. of ICCV 2013 Underwater Vision Workshop*, pp. 819–824 (2013).
- [10] Kawasaki, H., Furukawa, R., Sagawa, R. and Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern, *Proc. CVPR*, pp. 1–8 (2008).
- [11] Kimura, M., Mochimaru, M. and Kanade, T.: Projector calibration using arbitrary planes and calibrated camera, *Proc. CVPR*, pp. 1–2 (2007).
- [12] Kutulakos, K. N. and Seitz, S. M.: A theory of shape by space carving, *Proc. ICCV*, pp. 307–314 (1999).
- [13] Kutulakos, K. N. and Steger, E.: A theory of refractive and specular 3D shape by light-path triangulation, *IJCV*, Vol. 76, No. 1, pp. 13–29 (2008).
- [14] Laurentini, A.: How far 3D shapes can be understood from 2D silhouettes, *TPAMI*, Vol. 17, No. 2, pp. 188–195 (1995).
- [15] Moreno, D. and Taubin, G.: Simple, accurate, and robust projector-camera calibration, *Proc. 3DIMPVT*, pp. 464–471 (2012).
- [16] Mukaigawa, Y., Yagi, Y. and Raskar, R.: Analysis of light transport in scattering media, *Proc. CVPR*, pp. 153–160 (2010).
- [17] Narasimhan, S., Nayar, S., Sun, B. and Koppal, S.: Structured light in scattering media, *Proc. ICCV*, Vol. I, pp. 420–427 (2005).
- [18] Sakaue, F. and Sato, J.: Calibration of projector-camera systems from virtual mutual projection, *Proc. ICPR*, pp. 1–4 (2008).
- [19] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms, *Proc. CVPR*, pp. 519–528 (2006).
- [20] Thomas, D. and Sugimoto, A.: A flexible scene representation for 3d reconstruction using an RGB-D camera, *Proc. ICCV*, Vol. 1, No. 2, p. 3 (2013).
- [21] Yamauchi, K., Saito, H. and Sato, Y.: Calibration of a structured light system by observing planar object from unknown viewpoints, *Proc. ICPR*, pp. 1–4 (2008).