Reconstructionwith Underwater Stereo Vision

Lin Gui
Ocean University of China
Qingdao, China
gui_azure@163.com

Bo Yin
Ocean University of China
Qingdao, China
ybfirst@ouc.edu.cn

Zhiqiang Wei
Ocean University of China
Qingdao, China
weizhiqiang@ouc.edu.cn

Lei Huang
Ocean University of China
Qingdao, China
jasonhuangsc@163.com

Abstract—Conventional multi-view theory fails to explain underwater stereo vision because of the refraction bends light rays. This paper focuses on one of the most common scenarios of underwater vision, a stereo vision system of two cameras with a flat housing, in which the light ray refracts twice, once at the air-housing interface, and once at the housing-water interface.The papermodelsthe geometry of the light propagation explicitly. Under the principle of light ray's reversibility,discussion on one camera's view in 2D follows the backward projection by casting a light ray from the camera center to the object point.After the entire light ray's coplanarity is proved, the 2D case extends to 3D space easily. In 3D case, computation of object point from the casted ray is derived from the stereo view model. An 3D reconstruction algorithm is also proposed. The algorithm is concise and feasible because no particular configuration or additional device is required. It is proved accurate and low error rate in the experiment.

Keywords-underwater; stereo vision;reconstruction; flat housing; refraction

I. INTRODUCTION

Stereo vision is well studied and related technologies have been applied in industries, such as 3d reconstruction, measurement, navigation. Nevertheless there is still much work to do on stereo vision underwater, since the refractive mediums introduce challenges to stereo vision.

Conventional multiple-view geometry,based on theprecondition that light travels along straight lines fails in underwater vision.[1]View through a flat housing is one of the most common underwater scenarios. In such scene light rays bends into poly-lines when travelling through water, housing, air and then into a camera. Moreover, the degree of bending is connected with the incidence angle when the light ray enters one material from the other. Consequently, the objects look not only closer but also slightly distorted in the images. Therefore, depth can't be accomplished easily with traditional methods based on the light propagation in air.

Many studies on underwater vision have been proposed in recent years, while most discussion on reconstruction requires cameras to be posed at particular orientation, for example, approximately orthogonal to the air-glass or glass-water interface,which may not be suitable for most real occasions.[2]Some other studies focus on image formation model and light transmission to improve the visibility underwater. A depth map is also obtained, usually with the help of additional means, such as structured light or polarized filter.[3, 4] Models of view through single or multiple refractive layers are also discussed, while these studies focus on single view.[5,6,7]In addition, many researches focus on reconstruction on large scale such as subsea terrain.[8]R. Kawahara's[9] work is more likely to ours, but he has focused on encoded the refraction into the camera model, which is also more suitable for axial cameras.

To reconstruct the scene underwater, this paper focused on building a model of underwater vision with a glass flat housing,and proposed an easy and feasible solution for computing objects' position in water with two views. The model clarifies the geometry withtwice reflection on light ray's path, once at the water-glass interface, once at the glass-air interface. By tacking the light projection, the relationship between image points and corresponding object points is articulated, and a concise 3D reconstruction function is also proposed.

The contribution of this paper is as follows. Firstly, our model describes a general scenario of underwater vision, without requirement for the orientation of the camera, which is more applicable to real use than an axial camera model. Secondly, we propose a easy and feasible algorithm to reconstruct objects in water. We believe such function will build a foundation for underwater 3D reconstruction, measurement, and expand the application of stereo vision.

The rest of this paper is divided into 3sections. Section 2definesour general geometry model of underwater vision, and introduces the computation of depth. Section 3 describes our experiment. Section 4 makes the conclusion and discusses our future work.

II. UNDERWATER VISION MODEL

A. Basic principles

Light propagation is the basic theory of stereo and is also the fundament of the discussion. Two essential principles govern the propagation of light in a refractive situation.

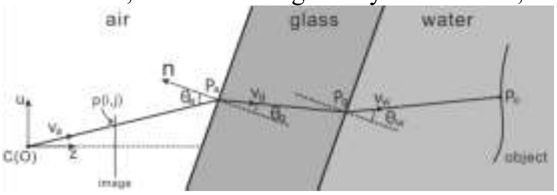

Figure 1. Underwater vision model

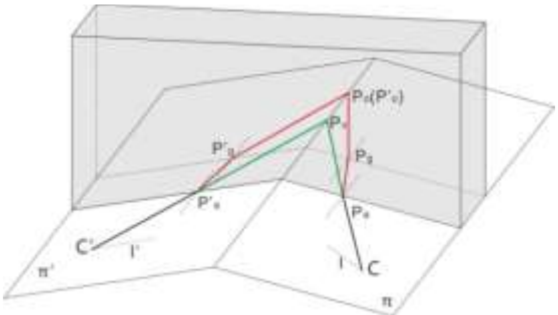

Figure 2.Stereo vision with two cameras

As we know, refraction of light obeys Snell's law, which defines the relationship between the incidence light and the refracted light as the follow equation:

$$\sin\theta_1/\sin\theta_2 = \mu_1/\mu_2 \tag{1}$$

θ_1 and θ_2 are angles of incidence light and the refracted light with the respect of the normal of the interface. $\mu_1$andμ_2 are the refractive indexes of the two materials.

The other principle is the reversibility of light ray, which deducted from Huygens–Fresnel principle.

Since both cameras share the similar situation, we focus on one camera's scenario in Fig. 1 first. The 2D case expands to 3D spaceeasily, which will be proved later. In the underwater vision, a light ray from objects refracts twice before reaching the camera, first at water-housing interface and then at housing-air interface. This forward projection is natural, but the forward derivation is highly nonlinear and difficult to compute because the position of object point is unknown.[2] Therefore, most of our discussion is on backward projection to simplify the derivation. Since the propagation of light is reversible, the former and the latter models are equivalent.

B. 2D model of single view

The two cameras in a stereo vision system share the similar situation of light casting, so our discussion starts with one camera's model in 2D case. Derivation in 2D easily expands to 3D scene because of the coplanarity of light's refraction[5], which will be proved later.

The backward projection assumes light raysare casted from the camera center, and travels through three mediumsonto object. Fig. 1 describes the scene. The coordinate system is defined as the projection of the camera coordinate system in the figure's plane. The origin O is the center of camera, also marked as C. The horizontal axis z is the projectionof the axis of camera, and the vertical axis u is defined orthogonal to z accordingly. For a pixel $p(x, y)$ in the image, the corresponding light ray intersects the air-housing interface at P_a and intersects the housing-water interface at P_g. The intersection angles with the surface normal $\boldsymbol{n}$ are $\theta_a$and θ_g. Then we will have the equations below by Snell's law:

$$\frac{\sin\theta_a}{\sin\theta_g} = \frac{\mu_a}{\mu_g} \tag{2}$$

$$\frac{\sin\theta_g}{\sin\theta_w} = \frac{\mu_g}{\mu_w} \tag{3}$$

$\theta_w$denotes the refractive angle of light ray in water. θ_a,θ_g and θ_w are all acute angles. μ_a, μ_g and μ_w denote the refractive indexes of the three materials respectively. For convenience, we represent the direction of the light ray with vectors in the following discussion. The direction vectors of the light ray in air, housing and water are marked as $\boldsymbol{v_a}$,$\boldsymbol{v_g}$and$\boldsymbol{v_w}$. Obviously $\boldsymbol{v_a}(f_c, u_a)$ is a cameras-related and pixel-wise vector withf_c as the focal length and$u_a = \sqrt{x^2 + y^2}$.Since μ_a, μ_g and μ_w are constants, $\boldsymbol{v_g}$can be derived from taken as a function of $\boldsymbol{v_a}$and $\boldsymbol{n}$.

$$\boldsymbol{v_g}(z_g, r_g) = \boldsymbol{f_g}(\boldsymbol{v_a}, \boldsymbol{n}) \tag{4}$$

$$\text{with}\frac{(v_g \cdot n)^2}{|v_g|^2 + |n|^2} = k_1 + k_1\frac{(v_a \cdot n)^2}{|v_a|^2 + |n|^2} + 1 \tag{5}$$

$$\text{and}(k_1 = \frac{{\mu_g}^2}{{\mu_a}^2}) \tag{6}$$

Eq. (5) is a deducted from Eq. (2). Analogously, $\boldsymbol{v_w}$ can also be denoted as a function of $\boldsymbol{v_g}$, and finally a function of $\boldsymbol{v_a}$.

$$\boldsymbol{v_w} = f_w(\boldsymbol{v_g}, \boldsymbol{n}) = f_w(f_g(\boldsymbol{v_a}, \boldsymbol{n}), \boldsymbol{n}) = g_w(\boldsymbol{v_a}, \boldsymbol{n}) \tag{7}$$

Since $\boldsymbol{v_g}$ and $\boldsymbol{v_w}$ is defined as a unit vector, its elements (z_g, r_g) will be easily calculated withEq. (5) or Eq. (7) and quadratic sum. Replacing the line from camera center to object point with vectors, the transitive relationship is acquired.

$$P_a = C + t_a\boldsymbol{v_a} \tag{8}$$

$$P_g = P_a + t_g\boldsymbol{v_g} \tag{9}$$

$$P_o = P_g + t_w\boldsymbol{v_w} \tag{10}$$

In the equations abovet_a, t_g and t_w are constants denoting thickness of air, housing and water the light ray passes. C can be omitted since it's the origin. By combining equations Eq. (4) to Eq. (10), we finally have the coordinate of point P_o on object.

$$P_o = t_a\boldsymbol{v_a} + t_g f_g(\boldsymbol{v_a}, \boldsymbol{n}) + t_w g_w(\boldsymbol{v_a}, \boldsymbol{n}) \tag{11}$$

C. 3Dmodel withstereo views

Actually the discussion on 2D is based on an implicit precondition that the entire light ray path lies on the same 3D plane. The precondition can be easily proved from Snell's law, and furthermore the 2D case extends naturally to 3D case under the condition. In the backward projection the casted ray travels through camera center C and a image point p by definition. When the light ray intersects the air-housing interface, a plane $\pi$is defined by the light ray and the interface normal at the intersection point. According to Snell's law, the incidence ray, the normal of the refractive surface, and the refracted ray lies on the same plane. Then the refractive light ray P_aP_g also lies onπ. The coplanarity is transited to the refraction on glass-water interface, thus the entire path of light-ray lies on the same planeπ. Consequently the equations in 2D discussion extends to 3D simply by extending the vectors from 2-dimension to 3-demension. Noticeably, definition of 3D vector $\boldsymbol{v_a}$ is as the following equations.

$$\boldsymbol{v_a}(\mathrm{x,y,z}) = (\mathrm{q_x}(\mathrm{x}-\mathrm{x_0}-\delta_\mathrm{x}), \mathrm{q_y}(\mathrm{y}-\mathrm{y_0}-\delta_\mathrm{y}), \mathrm{f_c}) \quad (12)$$

In this equation, $\mathrm{f_c}$ is the camera's focal length, which lies on axis z in Fig. 1's 2D scene.x and y still represent the image point's coordinate in the dimension of length. The principle point lies at $(\mathrm{x_0}, \mathrm{y_0})$. $\mathrm{q_x}$and$\mathrm{q_y}$ denote the amount of pixels in unit length on x and y direction with $\mathrm{q_x} = \mathrm{f_c}/\mathrm{a_x}$ and $\mathrm{q_y} = \mathrm{f_c}/\mathrm{a_y}$.$\delta_\mathrm{x}$andδ_y are radial distortion components.$\mathrm{a_x}$, $\mathrm{a_y}$,δ_x, $\delta_\mathrm{y}$and $\mathrm{f_c}$ are pre-calibrated intrinsic parameters of the camera or can be calculated with the parameters. Actual$\boldsymbol{v_a}$is scaled to unit vector by dividing by the square root of the quadratic sum of all three components. Unit vectors $\boldsymbol{v_a}$and$\boldsymbol{v_a}$ is also derivedfromEq. (4),Eq. (7) and the coplanar constrains.

Now we can extend the 2D model into 3D and take the other camera in to consideration also. Fig. 2 describes the casted light rays of both cameras in the stereo vision. Points on the light ray of the camera on right C, $\mathrm{P_a}$ and$\mathrm{P_g}$ and lies on plane π, while points on the light ray from the other camera $\mathrm{C'}$, $\mathrm{P'_a}$ and $\mathrm{P'_g}$ lies on planeπ'. The object point $\mathrm{P_o}$lies on both light rays' path, also both planes. Extending the light rays in air forward like the red lines, they will intersect at thepoint$\mathrm{P_v}$, the virtual image of the real object point$\mathrm{P_o}$. The object point looks locating at $\mathrm{P_v}$ because light rays are thought to be straight. Obviously, both $\mathrm{P_v}$ and $\mathrm{P_o}$ lie on the intersection line ofπ andπ'.

Since the geometry of $\mathrm{P_v}$ and the cameras obeys the principles of conventional multi-view, the following spatial transformation is valid.

$$\mathrm{P_v} = \mathrm{RP'_v} + \mathrm{T} \quad (13)$$

$\mathrm{P_v}$and$\mathrm{P'_v}$ denote the point in the coordinates of the two cameras respectively. RandT are the extrinsic parameters of the stereo vision system. Representing with vectors, we have

$$\mathrm{k}\boldsymbol{v_a} = \mathrm{R} \cdot \mathrm{k'}\boldsymbol{v'_a} + \mathrm{T} \quad (14)$$

k and $\mathrm{k'}$ are the distance between $\mathrm{P_v}$ and the camera centers, which can also be computed in conventional multi-view model. From Eq. (14) we get

$$\boldsymbol{v'_a} = \frac{1}{\mathrm{k'}}(\mathrm{kR^{-1}}\boldsymbol{v_a} - \mathrm{T}) \quad (15)$$

Then $\mathrm{P'_o}$ can also be derived with (11).Point coordinates derived through Eq. (15) will be used for error analysis.

According to Eq. (11), the surface normal $\boldsymbol{n}$and the distance constants $\mathrm{t_a}$, $\mathrm{t_g}$ and $\mathrm{t_w}$ are necessary for computing $\mathrm{P_o}$. $\mathrm{t_g}$can be omitted because it's only related to the direction of $\boldsymbol{v}_\mathrm{g}$ and the depth of housing. However, $\mathrm{t_a t_w}$ are affected by the position and orientation of the flat housing, so markers as Fig. 3 are introduced to calibrate the housing. Six markers are affixed on the outer surface of the housing, the corner's coordinates can be computed with the fundamental matrix.[1] Then normal $\boldsymbol{n}$ can be computed from the coplanarity. Denoting the corners as $\mathrm{Q_i}(\mathrm{x_i}, \mathrm{y_i}, \mathrm{z_i})$, since $\mathrm{P_a} = \mathrm{t_a}\boldsymbol{v_a}$, $\mathrm{t_a}$ can also be computed with the coplanar constrains of $\mathrm{P_a}$ and $\mathrm{Q_i}$.

$$(\mathrm{P_a} - \mathrm{Q_i}) \cdot ((\mathrm{Q_i} - \mathrm{Q_{i+1}}) \times (\mathrm{P_a} - \mathrm{Q_{i+1}})) \quad (16)$$

The values with all combinations of two markers are computed and the average value is used for computation. The similar equation also works for $\mathrm{t'_a}$, and then $\mathrm{t_w}$ and $\mathrm{t'_w}$ can be computed from $\mathrm{P_o} = \mathrm{P'_o}$. Finally coordinate of $\mathrm{P_o}$is calculated with Eq. (11).

D. *Algorithm*

The actual step-by-step algorithm is listed as follows.

- Calibrate the stereo vision system, including the intrinsic and extrinsic parameters of the two cameras, and compute the fundamental matrix
- Detect the corners of markers and compute the normal of the housing surface
- Extract the SIFT features and match the point pairs [11]
- Compute $\boldsymbol{v_a}$,$\boldsymbol{v_g}$,$\boldsymbol{v_w}$, $\mathrm{t_a}$, and$\mathrm{t_g}$for every matched points in left picture.
- Repeat last stepand compute the according vectors and constants for matched points in right picture.
- Compute $\mathrm{t_w}$ and $\mathrm{t'_w}$ with $\mathrm{P_o} = \mathrm{P'_o}$ for all point pairs
- Reconstruct the points' coordinates.

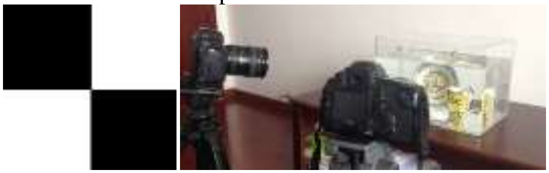
(a) (b)

Figure 3. (a) Checkerboard marker used for interface normal calibration. (b)Experiment configuration

III. EXPERIMENT

A. *Configuration and preparation*

Theactual experiment configuration is as Fig. 3 shows.A 60cm*25cm*40cm glass water tank is used as the flat housing,with the thickness of 0.65cm and the measured refraction index as 1.46. Two Canon 5D cameras are used to simulate a stereo vision system, providing images of the resolution 3168 by 2112. The cameras are located on one side of the same flat surface, about 40cm from the water tank with an approximate 40-degree angle between their orientations.

The calibrationof the stereo vision system is implemented with Zhang's method[10]. 14 image pairs of a checkerboard pattern are taken, and the parameters as follows are derived.

Intrinsic parameters of camera1:

$$\mathrm{A_1} = 1.0e + 03 * \begin{pmatrix} 4.1472 & 0 & 0 \\ 0 & 4.1423 & 0 \\ 2.1350 & 1.4327 & 0.0010 \end{pmatrix}$$

Intrinsic parameters of camera2:

$$\mathrm{A_2} = 1.0e + 03 * \begin{pmatrix} 2.9247 & 0 & 0 \\ 0 & 2.9235 & 0 \\ 1.5843 & 1.0980 & 0.0010 \end{pmatrix}$$

Extrinsic parameters of camera2:

$$\mathrm{R} = \begin{pmatrix} 0.9021 & 0.1302 & -0.5829 \\ -0.1186 & 0.9912 & 0.0582 \\ 0.5853 & 0.0224 & 0.8105 \end{pmatrix}$$

$$\mathrm{T} = (-460.0086 \quad -40.1219 \quad 66.7382)^\mathrm{T}$$

Fundamental matrix:

$$\mathrm{F} = \begin{pmatrix} -0.0000 & 0.0000 & -0.0108 \\ 0.0000 & 0.0000 & 0.1040 \\ 0.0061 & -0.1764 & -2.7499 \end{pmatrix}$$

B. *Result and analysis*

Some objects with different color design are immersed in the water for experiment. Images of the scene taken by the cameras are shown in Fig.4. The refraction through the water tank is quite obvious seeing from the table's edge behind the tank, while the color design of the objects are clear enough to be detected. The SIFT points in both imageare shown in Fig. 5. Most points lie on the standing plate because of the sophisticated and colorful pattern on it. Points on the mug are also detected. After matching the SIFT points, the object points' coordinates are computed, shown in Fig. 7.We can see that many points are lost during matching. Points on the markers are omitted because they're unnecessary for our reconstruction. Points on the reflected image on the housing are removed automatically since there is no pair for them.Only most of points on the mug fails to match. However points on the standing plane are well matched and recovered. From the dimensions in the figure we can see the points lie in the depth 0.6-0.65 with meter as the unit, which is approximate right at the object's location.

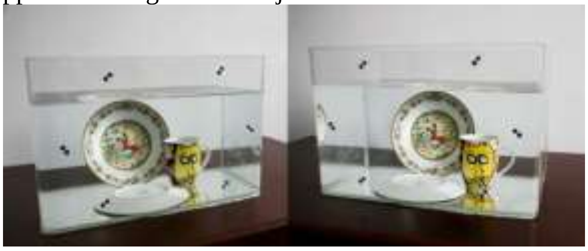

Figure 4.Real experimental image

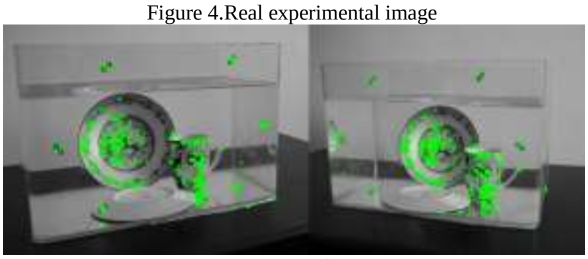

Figure 5.SIFT points in left and right images

Considering the error analysis, to compare the result with reconstruction of the same scene in air is difficult for real operation.Therefore, anreprojection error for analysis$\boldsymbol{v_d}$ is defined as the differencebetween $\mathrm{P_o}$and $\mathrm{P'_o}$. During the computation of $\mathrm{P_o}$, the unit vectors of incidence light ray in air $\boldsymbol{v_a}$ and $\boldsymbol{v'_a}$are derived from their according image points, while in the computation of $\mathrm{P'_o}$, one of vectors is calculated with the other byEq. (15). By plotting all $\boldsymbol{v_d}$ as points in 3D space, we visualize the errors as Fig. 6. Majority of the points lie in the range from -6 to 6 on three dimensions with millimeter as the unit. Then the distances $\mathrm{d_p}$between every pair of $\mathrm{P_o}$ and $\mathrm{P'_o}$ are also calculated, which is square root of the quadratic sum of the three components of $\boldsymbol{v_d}$. Under the dimension of millimeter, the values of $\mathrm{d_p}$fall between to 0.19 to 11.81, of which the average value is 4.1313 and the STD is 2.1994. That is to say the error rate is no more than 2%.

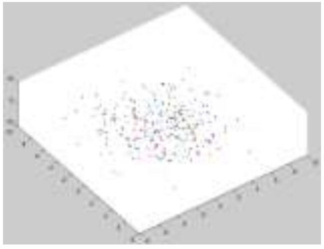

Figure 6.Projection errors

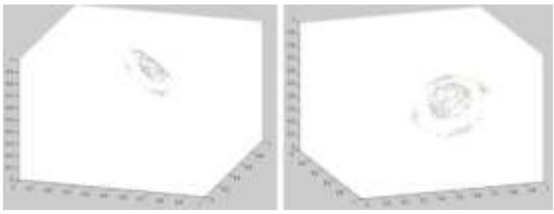

Figure 7.Reconstruction result

IV. Conclusion and future work

We modelthe underwater stereo vision with a flat housing explicitly based on the light propagation in refractive mediums. Backward projection of single vision in 2D case is discussed sufficiently as the foundation. Then we expand the case into 3D on the basis of refractive light ray's coplanarity. Furthermore, case of stereo vision with two cameras is also fully analyzed. Meanwhile, a 3D reconstruction algorithm is derived on the basis. In the algorithm, relationship between the casted light from two cameras is built for computing the object points, involving the extrinsic parameters and fundamental matrix of the stereo vision system. Simple markers on the housing's surface are also introduced to calibrate the orientationof the housing. Experiment results prove that the method is effective and accurate.

Our method shows low error rate on decimeter scale, which is similar to the distance of repairing underwater robots and its objects. The future work will be focused on two respects to make the method more practicable. Attempt with other features will be executed to match more points and new features may be introduced for specific scenario. On the other respect, scattering caused by particles in the water will be taken into consideration.

References

[1] Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

[2] Chang, Y. J., & Chen, T. (2011, November). Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 351-358). IEEE.

[3] Narasimhan, S. G., Nayar, S. K., Sun, B., & Koppal, S. J. (2005, October). Structured light in scattering media. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (Vol. 1, pp. 420-427). IEEE.

[4] Y. Y. Schechner, & Karpel, N. (2004, June). Clear underwater vision. In*Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 1, pp. I-536). IEEE.

[5] Agrawal, A., Ramalingam, S., Taguchi, Y., & Chari, V. (2012, June). A theory of multi-layer flat refractive geometry. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 3346-3353). IEEE.

[6] Chen, Z., Wong, K. Y., Matsushita, Y., Zhu, X., & Liu, M. (2011, November). Self-calibrating depth from refraction. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 635-642). IEEE.

[7] Shimizu, M., & Okutomi, M. (2008, June). Calibration and rectification for reflection stereo. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.

[8] Raimondo, S., & Silvia, C. (2010). Underwater image processing: state of the art of restoration and image enhancement methods. EURASIP Journal on Advances in Signal Processing, 2010.

[9] Kawahara, R., Nobuhara, S., & Matsuyama, T. (2013, December) A Pixel-wise Varifocal Camera Model for Efficient Forward Projection and Linear Extrinsic Calibration of Underwater Cameras with Flat Housings. *2013 IEEE International Conference on Computer Vision Workshop*. IEEE.

[10] Bouguet, J. Y., Camera Calibration Toolbox for Matlab.http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/parameters.html

[11] Vedaldi, A., SIFT for Matlab. http://www.robots.ox.ac.uk/~vedaldi/code/sift.html