



Axial image restoration of a single dome window view for stereo vision in unmanned underwater vehicles (UUV)

Ihab Elaff¹

Received: 15 March 2025 / Revised: 2 April 2025 / Accepted: 6 May 2025
© The Author(s) 2025

Abstract

Unmanned Underwater Vehicles (UUVs) are typically used for monitoring and inspection tasks where it's critical to locate objects and create 3D models of the surroundings. One of the most dependable techniques for accomplishing 3D localization and modeling jobs is stereo vision. This isn't the case in underwater conditions, though, as cameras will record distorted views, which results in inaccurate 3D estimation. In real-world applications, it is more practicable and a more difficult task to place the two cameras of the stereo vision system inside a single dome glass. The goal of this study is to recover the axial image views of the two cameras within a single dome window from the refracted views that were recorded. The cameras' position (along the Z and X axes) and orientation (pan and tilt) within a single dome window—of which there are 448 possible combinations—are taken into account by the kinematics model. The approach has been improved and validated using around 18 million experiments. Lastly, when compared to the original axial images, our approach may report axial images with an average overlapped similarity of 96%.

Keywords Stereo vision · Dome window · Axial image · Refracted image · UUV

Abbreviations

UUV	Unmanned Underwater Vehicles
ROV	Remotely Operated Vehicle
FOV	Field Of View
PTRC	Pan-Tilt Rotation Center
CSC	Common Spherical Center
CC	Camera Center
OP	Observation Plan

1 Introduction

Many scientists are still concerned about the difficult topic of exploring the underwater environment. The ability of Unmanned Underwater Vehicles (UUVs) to locate an object underwater has several vital uses, including navigation, sea floor modeling, modeling objects beneath the water's surface, and more [1, 2]. Numerous studies have addressed the

axial image restoration of a single camera behind a dome-port or flat-port (aligned at the central axis of the dome) [3–7]. While two cameras (or more) can localize the 3D location of the targeted object, a single camera only displays the image in 2D, making it unsuitable for estimating the 3D location of any object [8]. Stereo vision, which usually consists of two cameras, is more useful in practical applications, though some experiments have even been conducted to determine the distance to an object using a laser beam and a single camera [9].

Localization of a particular object in 3D begins with simultaneous image capture by both cameras, followed by the specification of matching primitives of the desired object and, lastly, a guess as to the 3D location of those matching primitives for both the parallel stereo vision system setup and the converging vision system setup [8, 10]. Numerous approaches have been used to solve this problem, and the findings have been shown to be highly accurate [11–14]. Localization of a particular object in 3D begins with simultaneous image capture by both cameras, followed by the specification of matching primitives of the desired object and, lastly, a guess as to the 3D location of those matching primitives for both the parallel stereo vision system setup and the converging vision system setup [8, 10]. Numerous

✉ Ihab Elaff
ihabelaff@qu.edu.qa

¹ Computer Science and Engineering Department, Collage of Engineering, Qatar University, Doha, Qatar

approaches have been used to solve this problem, and the findings have been shown to be highly accurate [11–14].

Li et al. [15] used the normalized position of the underwater target to create a measuring model for refractive stereo vision based on light propagation routes. On the basis of the idea that light travels the shortest distance between two sites, they also developed an underwater simulation imaging model. In order to detect cylindrical pipes, Lodi Rizzini et al. [16] detailed how to integrate a stereo vision system into the MARIS intervention AUV and how to configure it. For each, two different cameras with windows have been employed. Tests conducted in an outdoor swimming pool under various lighting conditions demonstrate that the algorithmic approach used enables target pipe detection and offers an assessment that is sufficiently accurate. The accuracy of the fish 3D location estimation in the aquaculture cage was examined by Ubina et al. [17] using the stereo camera images that were taken. The study is predicated on the rectified camera system's known geometrical arrangement. The resolution of the calculated depth maps of fish is the crucial factor limiting the accuracy of 3D fish metric estimation. Reliable convolutional neural networks (CNNs) are used to address this problem by proposing an object-based matching for depth computing and underwater fish tracking. The implementation and experimental outcomes of underwater StereoFusion, an algorithm for real-time 3D dense reconstruction and camera tracking, are presented by Rossi et al. [18]. With the use of two distinct cutting-edge underwater Remotely Operated Vehicles (ROVs), it has been successfully tested in both an ocean and a lake. For a stereo vision system with each camera positioned in a different dome window, Dunkley et al. [19] created comprehensive open source hardware and software solutions. Images from the left and right cameras are synchronized for calibration using a flashing light. A stereo vision system with two movable cameras that use motors to separately pan, tilt, and slide was proposed by Sagara et al. [20]. This vision system can track a target item in underwater situations since it is installed inside a waterproof cylindrical container.

Because it offers more viewing angles and can withstand a greater depth of water than other port types, a dome-port was selected for this project. Although multiple dome windows could be used for stereo vision (as previously introduced), it is more feasible for real-world applications and a more difficult problem to have two cameras inside a single dome window. This model achieves 97% overlap between the restored image and the axial image. Previous research [7] addressed axial image restoration of dome window view for a single camera that is aligned at different displacements along the center axis of the dome at any combination of pan-tilt-depth. In order to make it suitable for stereo vision from a single dome window, this manuscript discusses an

improvement to the previous work that covers the restoration of axial images when the camera is shifted away from the center axis of the dome. In comparison to the original axial images, it is successful in obtaining axial images with an average overlapping similarity of 96%.

2 Materials and methods

2.1 Stereo vision system

Since a single camera can only depict a view in two dimensions without providing depth information, three-dimensional vision can be accomplished with two or more cameras. Two cameras are employed by the stereo vision system to determine the view's depth. There are two configurations for the stereo vision system: chatting vision and parallel stereo vision [8, 10]. The center axes of the cameras are parallel to one another in the first configuration, while they intersect in the second.

High camera resolution (which will also take more processing time) and a big base distance between the two cameras are some issues with more precise 3D localization of a specific location. The process of localizing a particular object in three dimensions begins with the simultaneous capture of images by both cameras, followed by the specification of matching parameters for the desired object in both views, and the computation of the 3D location of those matching parameters.

2.2 Snell's law

The light-ray will be subjected to refraction when passing through different materials, which can be represented by [7]:

$$\hat{s}_2 = \left(\frac{n_1}{n_2} \right) \left(\hat{s}_1 - \left(\hat{N}_{\text{surf}} \hat{s}_1 \right) \hat{N}_{\text{surf}} \right) + \sqrt{1 - \left(\frac{n_1}{n_2} \right)^2 \left(1 - \left(\hat{N}_{\text{surf}} \hat{s}_1 \right)^2 \right)} \hat{N}_{\text{surf}} \quad (1)$$

Where n_1 and n_2 represent refraction indices for 2 the different materials, \hat{s}_1 and \hat{s}_2 are the incident ray and the refracted ray normalized vectors, and \hat{N}_{surf} is the normal vector on the interface surface between the 2 materials.

2.3 UUV's stereo vision system setup

Using two servo motors, a pan-tilt rotation mechanism has been created to rotate the Spaiser ROV's camera [21] inside an acrylic dome. On the inner ring of that system, two fixed small-thin-lens cameras have been put. They are spaced

30 mm apart along the camera's X-axis from the Pan-Tilt Rotation Center (PTRC) of the system. The term Common Spherical Center (CSC) refers to the center of the inner and outer surfaces of the dome.

3 Methodology

3.1 Kinematic modeling

A kinematics model has been defined to identify the center of the Observation Plane's (OP) pose of one camera of the stereo vision system with respect to CSC by the following transformation sequence (Fig. 1):

- 1) Displacement between CSC and PTRC along camera's Z-axis: $\text{Disp}(0,0,Z_1)$.
- 2) Rotation around camera's Y-axis: $\text{Rot}(Y_{\text{cam}}, \phi_2)$.
- 3) Rotation around camera's X-axis: $\text{Rot}(X_{\text{cam}}, \theta_3)$.
- 4) Displacement between the PTRC and the Camera-Center (CC) along X-axis of the base: $\text{Disp}(X_4,0,0)$.

Then, the CC pose with respect to CSC, can be calculated using:

$$T_{\text{Cam}} = \begin{bmatrix} C_2 & S_2S_3 & S_2C_3 & C_2X_4 \\ 0 & C_3 & -S_3 & 0 \\ -S_2 & C_2S_3 & C_2C_3 & -S_2X_4 + Z_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

The OP of each camera, where the images are received by the camera, is shifted away from the CC along the camera's Z-axis by Z_{OP} . Then each pixel's physical location $G(G_x, G_y, G_z)$ on the OP (camera's sensor) can be calculated with respect to the CRC using the following formulas:

$$T_G = T_{\text{Cam}} \text{Disp}(X_G, Y_G, Z_{\text{OP}}) \quad (3)$$

where

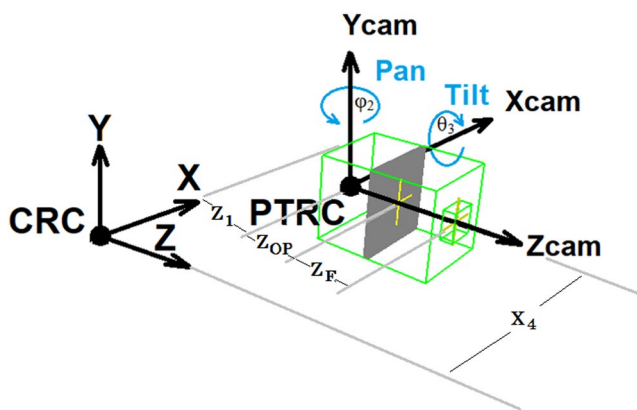


Fig. 1 Kinematics model

$$X_G = \left(\frac{U}{M}\right) m \text{ where } -\frac{M}{2} \leq m \leq \frac{M}{2} \quad (4)$$

$$Y_G = \left(\frac{V}{N}\right) n \text{ where } -\frac{N}{2} \leq n \leq \frac{N}{2} \quad (5)$$

Then

$$G(G_x, G_y, G_z) = (C_2X_G + S_2S_3Y_G + S_2C_3Z_{\text{OP}} + C_2X_4, C_3Y_G - S_3Z_{\text{OP}}, -S_2X_G + C_2S_3Y_G + C_2C_3Z_{\text{OP}} - S_2X_4 + Z_1) \quad (6)$$

Where the captured image of resolution $N \times M$ (in pixels) is the mapping of the rectangular area $U \times V$ (in mm) area on the OP. The camera's lens is located away from the OP along Z-axis by the focal length f . This can be determined by the following transformation:

$$T_L = T_{\text{Cam}} \text{Disp}(0,0, Z_{\text{OP}} + f) \quad (7)$$

$$L(L_x, L_y, L_z) = (S_2C_3[Z_{\text{OP}} + f] + C_2X_4, -S_3[Z_{\text{OP}} + f], C_2C_3[Z_{\text{OP}} + f] - S_2X_4 + Z_1) \quad (8)$$

3.2 Axial view image restoration

As Snell's law is bijective, then the normalized incident vector \hat{S}_{air} from any point G on the camera's OP towards the dome will be refracted on the surface of the dome at point $P(P_x, P_y, P_z)$ with the normalized vector \hat{S}_{dome} which intersects with the dome-water interface point $Q(Q_x, Q_y, Q_z)$ and again it is refracted as the normalized vector \hat{S}_{water} (Fig. 2).

It is not possible to immediately construct the axial image because the entering water rays will not cross at the same location [7]. By projecting the outgoing water rays onto a plan that is far from the camera (the point R), and then rerouting incoming vectors from that plan towards the focal point of the camera's lens and subsequently to the OP (the point V), homographic transformation [22] is used to solve this problem. Although this can be improved, it will fix the regenerated view on the OP with an adequate match to the axial view. Results could be much enhanced by slightly moving the virtual lens point away from the original lens location during homographic translation.

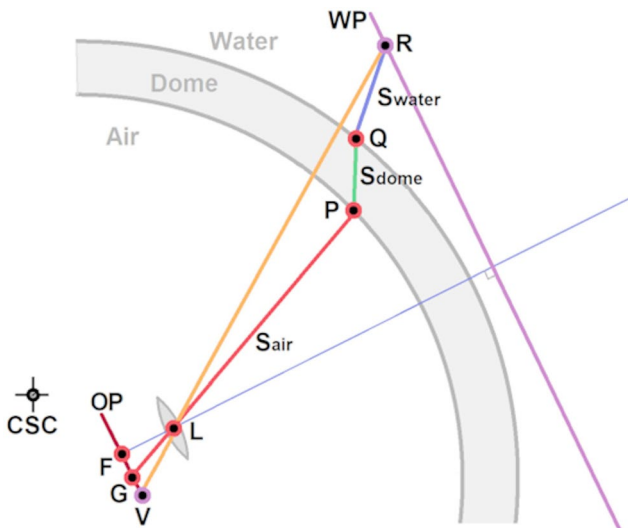


Fig. 2 Axial view image restoration

4 Experimental phase

4.1 Images acquisition

In this study, a verified Blender's mimic [23] for the water pool and dome port of a real-world arrangement was used. For testing and confirmation, axial and refracted images of certain views have been produced using Spaiser's camera settings [21], which are 3.6 mm lens and 4.8×3.6 mm target surface. The axial

and corresponding refracted images for the testing object have been produced (448 different combinations) using a range of camera displacements along the Z-axis (0 to 30 mm step 10), along the X-axis (0 to 30 mm step 10), and Pan-Tilt angles (-30 to 30 step 10 and 0 to 30 step 10 respectively).

4.2 Refraction index

The Euclidian distance between the position vectors \hat{S}_{air} and \hat{S}_{water} is used to compute a refraction index, which shows the impact of the camera's pose inside the dome window:

$$Ref_{Idx} = Sqrt((\hat{S}_{air}.X - \hat{S}_{water}.X)^2 + (\hat{S}_{air}.Y - \hat{S}_{water}.Y)^2 + (\hat{S}_{air}.Z - \hat{S}_{water}.Z)^2) \quad (9)$$

High values signify significant refraction-induced deformation. Figure 3 illustrates the impact of Z-axis displacement and Pan angle for each displacement along the X-axis. In all cases, the tilt angle has no discernible effect on the Ref_{Idx} (the average for all combinations is 0.005 ± 0.005 for the tested tilt angles between 0 and 30 degrees).

It is obvious that the observed image will be significantly affected by more displacements along the X and Z axes as

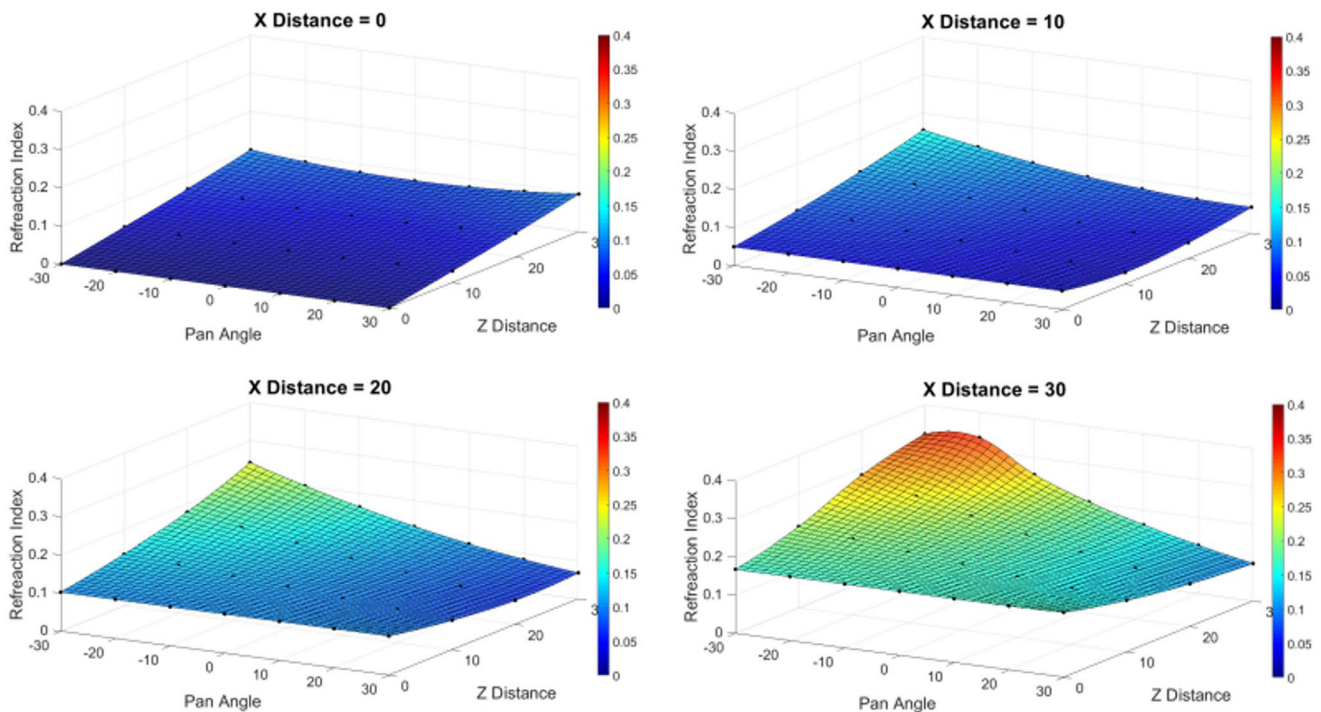


Fig. 3 Refraction index of dome window

well as more angles, which in turn leads to more. However, limiting the pan angle will narrow the field of view, and decreasing the X-axis displacement will impact the 3D vision's precision. In stereo vision applications, rectification of the seen view would be a superior option because reduction is not a feasible solution.

4.3 Shifting coefficients profiles

Overlap similarity coefficients have been produced for each of the generated images both before and after the virtual lens point has been shifted for every possible combination. Values that achieve the largest overlap have been chosen for each combination, and the shifting of dZ and dX with values from -1 mm to $+1$ mm at step 0.01 mm has been used (for each combination

which form in total roughly 18 million testing cases) (Fig. 4).

Figure 4 shows that the values of dZ and dX for the tested combinations have a non-linear relationship with the system settings. Using MATLAB, a feedforward neural network with a single hidden layer of 40 neurons is built and trained to estimate values of dZ and dX . The network's inputs are Pan angle, X distance, and Z distance. To precisely address the goal values, all data is used for training. A smaller Pan angle step can only produce more choices because the cameras' Z and X distances are set for each stereo vision system. The network can then be re-trained to use the new combinations.

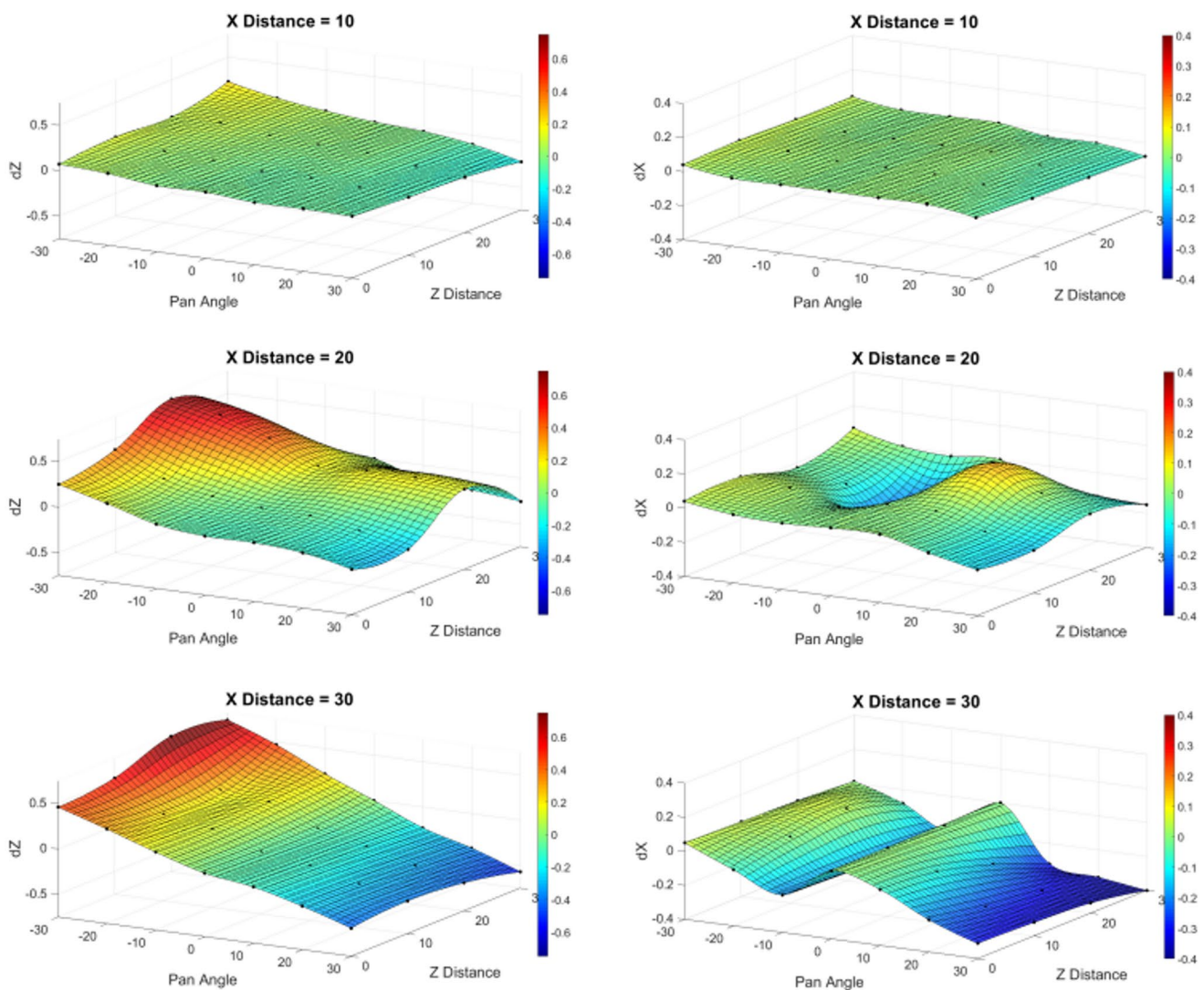


Fig. 4 Shifting coefficient profiles

4.4 Model validation

The average overlapping similarity has improved from 0.86 ± 0.14 (min value is 0.41) to 0.96 ± 0.03 (min value is 0.78), indicating a considerable improvement in overlapping similarity for all cases when the virtual lens is shifted (Fig. 5). Given that shifting along the X-axis significantly

affects systems, Table 1; Fig. 6 demonstrate how applying shifting significantly improves overlapping similarity.

The location and orientation of the recovered view of a sample object almost matches its axial view, even in the worst case (overlapping similarity is 0.78 after shifting) when Z distance, X distance, and Pan angle are 20, 20, and 0 respectively (Fig. 7).

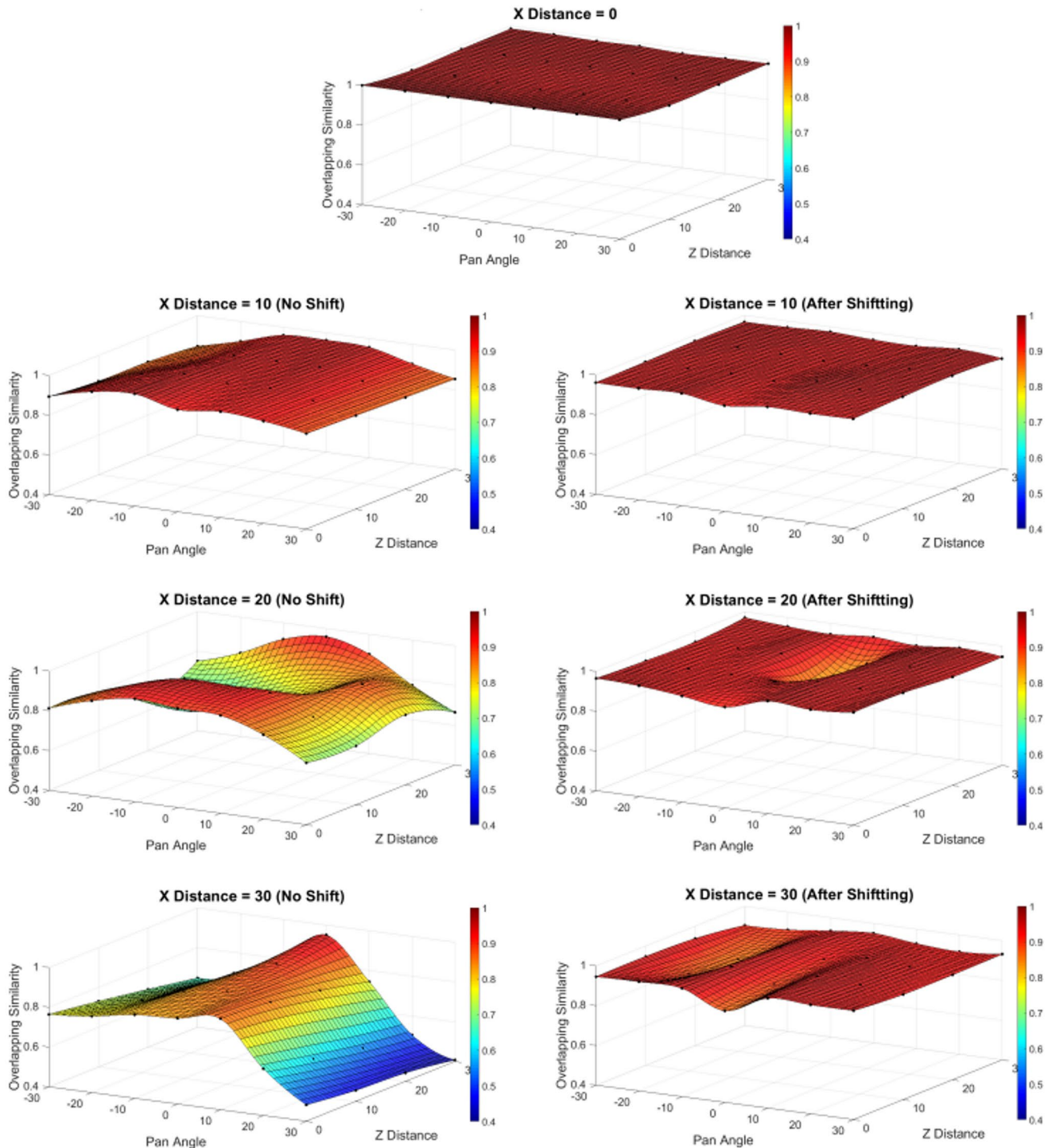
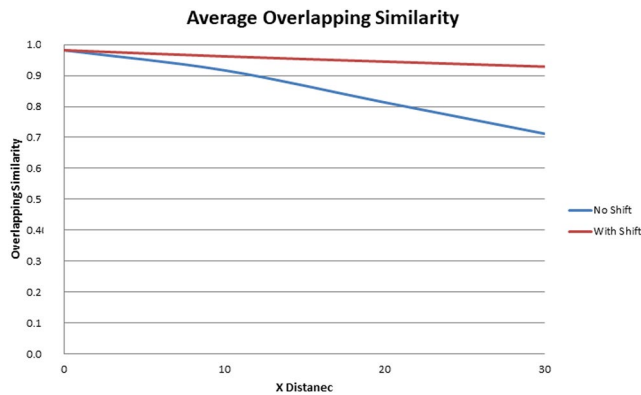


Fig. 5 Overlapping similarity with respect to X distance before and after shifting the virtual lens

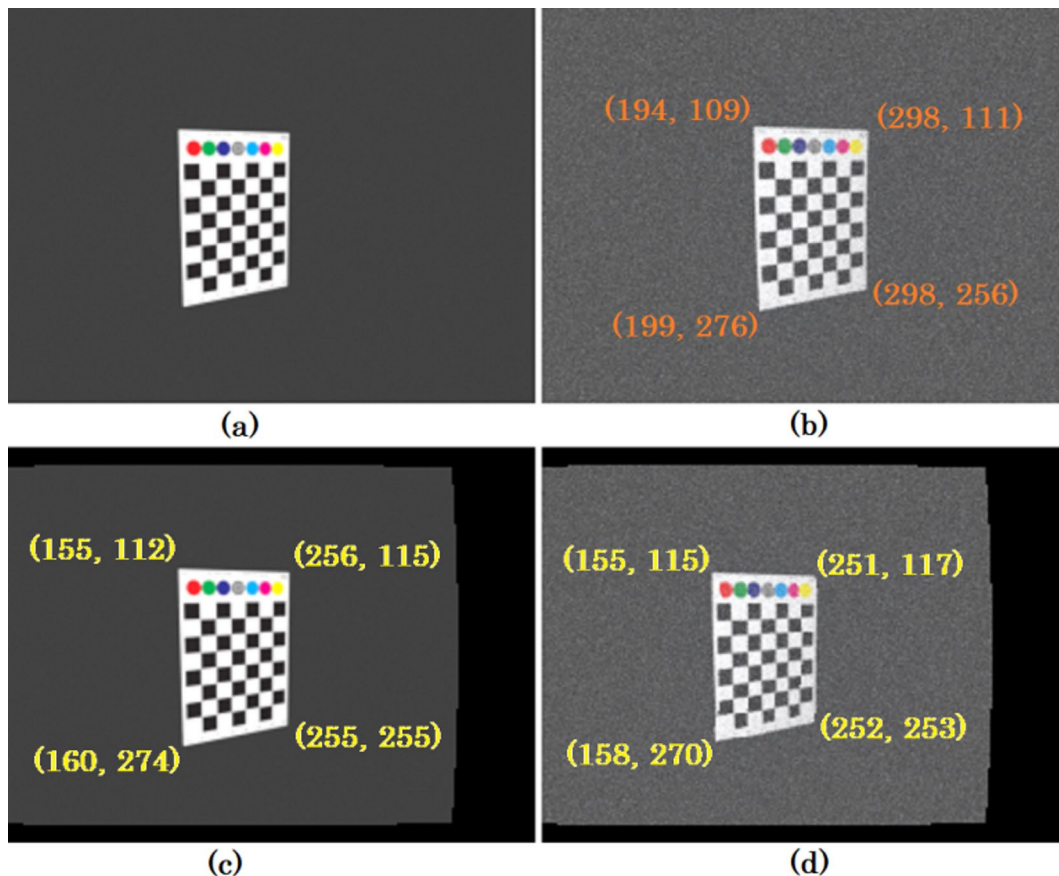
Table 1 Improvement of the average overlapping similarity with respect to X-axis displacement

X Distance	No Shift AVG \pm STDV	After Shifting	AVG Improvement%
0	0.98 \pm 0.012	0.98 \pm 0.012	0%
10	0.92 \pm 0.037	0.96 \pm 0.009	4%
20	0.81 \pm 0.094	0.95 \pm 0.035	17%
30	0.71 \pm 0.152	0.93 \pm 0.027	31%

**Fig. 6** Improvement of the average overlapping similarity with respect to X-axis displacement

5 Discussion

To recover the axial images of a stereo vision system placed behind a UAV system's dome glass, a novel analytical technique has been created. The technique suggests a kinematics model that uses shifting along the Z and X axes of a pan-tilt rotation base to describe a camera's posture. The technique was created using Snell's law, homographic transformation, and 448 distinct ways to rotate and translate a single camera inside a dome window. An average of 96% overlapping similarity between the original axial image and the restored axial image has been attained after about 18 million experiments were used to refine and validate the technique. According to the response index, the observed image will be significantly impacted by higher displacements along the X and Z axes as well as more angles. By changing the virtual lens, the average overlapping similarity has improved from 0.86 ± 0.14 (min value is 0.41) to 0.96 ± 0.03 (min value is 0.78), which is a considerable improvement for all cases. A Feed Forward Neural Network with one hidden layer of 40 neurons is trained to estimate values of dZ and dX based on Z distance, X distance, and Pan angle since ΔZ and ΔX are non-linearly connected to the system's setup. The observed sample object in the reconstructed image nearly

**Fig. 7** Worst case scenario matching. (a) Axial image, (b) Refracted image (c) clipped axial image and (d) Restored image

matches its axial perspective, even in the worst situation, when Z distance, X distance, and Pan angle are 20, 20, and 0 correspondingly.

Acknowledgements Open Access funding provided by the Qatar National Library. David Nakath from GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstrasse, Kiel, Germany for the provision of the mimic Bender project.

Author contributions I am a single author and i contribute to the entire manuscript.

Funding Open Access funding provided by the Qatar National Library.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Wu, H., Chen, Y., Yang, Q., Yan, B., Yang, X.: A review of underwater robot localization in confined spaces. *J. Mar. Sci. Eng.* **12**(3), 428 (2024)
- Petillot, Y.R., Antonelli, G., Casalino, G., Ferreira, F.: Underwater robots: From remotely operated vehicles to Intervention-Autonomous underwater vehicles. *IEEE Robot Autom. Mag.* **26**, 94–101 (2019)
- Fan, W., Zhang, W., Fan, Z.: Research on aberration correction methods of conformal dome based on Haack curve. *Opt. Commun.* 565:130676. (2024)
- Zhou, G., Liu, Y., Dang, B., Yu, C., Li, L., Du, J., Ma, J., Liu, X.: Calculation and analysis of key parameters of underwater optical imaging system. *Sensors*. **24**, 1537 (2024)
- Song, Y., Nakath, D., She, M., Köser, K.: Optical imaging and image restoration techniques for deep ocean mapping: A comprehensive survey. *PFG*. **90**, 243–267 (2022)
- Holak, K., Cieslak, P., Kohut, P., Giergiel, M.: A vision system for pose Estimation of an underwater robot. *J. Mar. Eng. Technol.* **21**(4), 234–248 (2020)
- Elaff, I.: Inverse model for correcting underwater vision by axial image restoration of dome window view, *Measurement* 239:115459. (2025)
- Grewe, L.L., Kak, A.C.: Handbook of Pattern Recognition and Image Processing (Vol. 2): Ch8 Computer Vision. Academic Press, Inc. (1994)
- Muljowidodo, K., Mochammad, A.R., Agus, N.S.A.B.: Vision based distance measurement system using single laser pointer design for underwater vehicle. *Indian J. Mar. Sci.* **38**(3), 324–331 (2009)
- Domnguez-Morales, M., Jimnez-Fernndez, A., Paz-Vicente, R., Linares-Barranco, A., Jimnez-Moreno, G.: Stereo Matching: from the Basis To Neuromorphic Engineering. *Current Advancements in Stereo Vision*. InTech (2012)
- Hamid, M.S., Abd Manap, N.F., Hamzah, R.A.: Kadmin stereo matching algorithm based on deep learning: A survey. *J. King Saud Univ. - Comput. Inform. Sci.* **34**(5), 1663–1673 (2022)
- Bi, Y., Li, C., Tong, X.: An application of stereo matching algorithm based on transfer learning on robots in multiple scenes. *Sci. Rep.* **13**, 12739 (2023)
- Yoshizawa, M., Motegi, K., Shiraishi, Y.: A deep Learning-Enhanced stereo matching method and its application to Bin picking problems involving tiny cubic workpieces. *Electronics*. **12**(18), 3978 (2023)
- Wu, R., Wang, M., Li, Z.: Few-Shot stereo matching with high domain adaptability based on adaptive recursive network. *Int. J. Comput. Vision*. **132**, 1484–1501 (2024)
- Li, G., Huang, S., Yin, Z., Li, J., Zhang, K.: Underwater refractive stereo vision measurement and simulation imaging model based on optical path. *J. Mar. Sci. Eng.* **12**(11), 1955 (2024)
- Lodi Rizzini, D., Kallasi, F., Aleotti, J., Oleari, F., Caselli, S.: Integration of a stereo vision system into an autonomous underwater vehicle for pipe manipulation tasks. *J. Computers Electr. Eng.* **58**, 560–571 (2017)
- Ubina, N.A., Cheng, S.C., Chang, C.C., Cai, S.Y., Lan, H.Y., Lu, H.Y.: Intelligent underwater stereo camera design for fish metric Estimation using reliable object matching. *IEEE Access*. **10**, 74605–74619 (2022)
- Rossi, M., Tršić, P., Sivčev, S., Riordan, J., Toal, D., Dooly, G.: Real-Time Underw. StereoFusion Sens. **18**(11), 3936 (2018)
- Dunkley, K., Dunkley, A., Drewnicki, J., Keith, I., Herbert-Read, J.E.: A low-cost, long-running, open-source stereo camera for tracking aquatic species and their behaviours. *Methods Ecol. Evol.* **14**, 2549–2556 (2023)
- Sagara, S., Ambar, R.B., Takemura, F.: A stereo vision system for underwater Vehicle-Manipulator systems - Proposal of a novel concept using Pan-Tilt-Slide cameras. *J. Rob. Mechatronics*. **25**, 785–794 (2013)
- Elaff, I.: Design and development of spaiser remotely operated vehicle. *J. Eng. Appl. Sci.* **69**, pp14 (2022)
- Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2003)
- Nakath, D., She, M., Song, Y., et al.: An optical digital twin for underwater photogrammetry. *PFG*. **90**, 69–81 (2022)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.