Lanty: A Deep Sea Stereo Vision System

Eric Alfaro-Dufour *, Caterina Muntaner-González *, Antoni Martorell-Torres*, Gabriel Oliver-Codina*
Dept. of Mathematics and Computer Science, University of the Balearic Islands, Palma, Spain
Email: {e.alfaro, c.muntaner, antoni.martorell, goliver}@uib.es
*ORCID: 0009-0004-3555-2582, 0009-0003-9091-7970, 0000-0001-9189-5265, 0000-0001-6910-1940

2024 IEEE International Conference on Industrial Technology (ICIT) | 979-8-3503-4026-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIT58233.2024.10540725

***Abstract*—This work introduces a comprehensive hardware and software solution for an underwater Stereo Vision System. Rigorous robustness and resistance studies were conducted using finite element analysis, accompanied by a meticulous selection of materials and design considerations. Multiple tests were executed to determine optimal camera configuration parameters, ensuring the acquisition of high-quality images and accurate 3D information. A Robot Operating System-based software for fish detection and classification was also integrated into the system, enabling the creation of 3D point clouds. The entire system was successfully tested in a real marine environment, and the results demonstrated the robustness of the hardware design and the potential of the software for fish biomass estimation applications. These findings confirm that the proposed solution is effective and can be practically applied in underwater environments.**

***Index Terms*—stereo vision, underwater, hardware, deep learning.**

I. Introduction

Stereo vision, also known as binocular vision, involves using two cameras to observe a scene from two slightly different viewpoints. As soon as two image points are matched, i.e., identified as corresponding to the same physical point, it is possible to compute the three-dimensional coordinates of this physical point. Stereo vision happens to be very attractive for many applications in robotics, including 3-D object recognition and localization as well as 3-D navigation of mobile robots [1]. Due to significant advancements in camera technology, stereo vision has recently extended to underwater applications, such as deep-sea exploration or navigation control of Remote Operated Vehicles (ROVs) or Autonomous Underwater Vehicles (AUVs).

Underwater applications of Stereo Vision Systems (SVS) present few more challenges than in-air systems. Underwater vision requires a pressure housing with a clear window through which a camera can observe. For this purpose, flat ports are commonly used in underwater housings due to their simplicity and low cost. However, flat ports present some drawbacks, particularly when utilized in deep-sea applications. Firstly, it is important to note that cameras positioned behind a flat port are subject to refraction as a result of the different media. Secondly, very strong, and therefore, expensive materials are needed to resist the demanding high pressure of the underwater environment. Additionally, thicker flat ports can also be used to meet mechanical requirements. Alternatively to flat ports, acrylic or glass dome ports could be used since they are better suited for high-pressure conditions. Besides, the field of view is not limited and cameras centered in the dome will not suffer from refraction. Nevertheless, precise mechanical centering and thorough calibration are essential [2].

In this article, we present a comprehensive hardware and software solution for a deep-sea underwater SVS. The prototype, called Lanty, has been developed within the context of PLOME project, and the main goals are described below:

- Design and test of a high pressure resistance enclosure.
- Assembly of a SVS. This includes cameras and process unit.
- Development of an algorithm for fish detection and creation of 3D point clouds.
- Application of the SVS in real-world conditions to verify the performance of the designed system.

This paper is structured as follows: section II analyses the related work on this field, section III explains the hardware used, section IV describes the design of the SVS, section V covers the software integrated, section VI shows the experiments performed and a real application of the SVS, and finally, section VII illustrates the conclusions and proposes future work.

II. Related Work

When it comes to underwater SVS, few current references have been found in this line due to the challenges that this application presents, specially for deep-sea purposes. Most of the work found refers to monocular vision systems. For example, in [3] a cost-effective video recording system for a rapid appraisal of deep-sea benthic habitats was developed. Authors in [4] presented a compact and robust camera system for deep-sea applications, up to 5500 metres below sea level. Researchers in [5] explained in detail the structural design and scientific payload of a deep-sea camera system. Besides, methodological procedures were described, including camera recovery which becomes a challenging task in deep-sea applications.

Another important challenge that arises in underwater environments is battery duration. In reference [6], a micro-controller with a real-time clock was used to wake up the imaging computer at regular preprogrammed intervals. Once images were captured, the computer was automatically turned off until the next period, leaving the micro-controller as the only component permanently drawing energy from the battery. This kind of functionality offered them the possibility to perform long deployments, up to several weeks, for bio-fouling settlement study.

Focusing on the progress made so far in the field of underwater SVS, in [8] a low-cost SVS prototype was developed for underwater object detection. A trade-off between computational power requirements and electrical power consumption was accomplished in order to perform the expensive computational cost of the vision tasks and still guarantee enough level of autonomy for at least two hours of video acquisition. However, very low-quality 3D point clouds were obtained due to the lack of synchronization between the webcams. In further studies, they showed an improved hardware design of their SVS which could operate up to 50 metres in depth. Besides, a calibration method that exploits the shape of cylindrical submerged objects was also proposed [9].

Authors in [10] used an AUV with an integrated stereo camera vision system to perform real-time pipe mapping. In their study, a semantic segmentation network and two information algorithms were developed in order to detect pipes and valves, extract the information providing pipe vectors or structural elements and then unify data from spatially referenced point clouds, forming an information map of the inspected area. In reference [11], an intelligent SVS for deep-sea operations was presented. Their SVS configuration underwent a thorough component selection process as well as an optimization of the camera housing by using Finite Element Method (FEM). Furthermore, field trials were performed off the shore of Marseille on a commercial ROV.

Regarding fish detection and monitoring through SVS, it has increasingly gained popularity over the last decade due to significant advancements in computer vision and deep learning. Researchers in [12] proposed a low-cost and compact SVS capable of measuring live fish in aquatic conditions, without requiring any special in-water calibration. Even though their system presented a short baseline (about 100 mm), fish measurement accuracy was high. In reference [13], complete hardware and software designs were presented with the aim of recording species and their behaviour in 3D for multiple days. In their camera housing, they utilized dome ports rather than flat ports since the field of view is not restricted and image refraction is reduced. Besides, two different methods were implemented to calculate the temporal offset between cameras as well as spatial calibration throughout an accuracy-improved open-source script.

Lanty has been developed with the aim of gathering images of fishes living between the range of the mesopelagic zone, up to 500 metres deep. Similarly to the work in [11], a detailed description of component selection will be given, both for electronics and hardware. As well, an exhaustive analysis of stresses and deformations in the camera enclosure will be performed. Moreover, software development contribution will be presented as well as an application in real-world conditions.

III. Hardware

Lanty has been designed taking into consideration space limitations and mechanical requirements to accomplish a good performance under extreme conditions of pressure. The aim of the project was to develop a compact system that could be easily transported and also quickly deployed, not only for deep-sea applications but also for shallow water uses.

In this section, we explain the procedure regarding the selection of hardware components. Firstly, the electronics which have been accurately selected to fit in with the requirements of stereo vision, secondly, the integrated power system, and finally, the adequate materials that have been utilized within the enclosure design in order to meet the proposed mechanical performance.

A. Electronics

The SVS is integrated by two USB3 Chameleon3 cameras, which provide high-resolution images, mounted with 5mm focal length and 10 megapixels lenses. These cameras are physically synchronised and powered directly by a high-capacity 14,8V 70 Ah Li-Ion battery. All data received from the cameras are processed by an Intel Nuc Pro12, running on Ubuntu 20.04 and Robot Operating System (ROS) in *Noetic* version. This minicomputer incorporates a 12-core i7 processor, with an operating frequency of up to 4,7 GHz. It also has 16GB of RAM and 2TB of storage capacity.

Regarding the illumination, the system includes two LED lights which are controlled by an Arduino Nano via GPIO pins. Both lights are powered by the battery whereas the microcontroller power supply comes directly from the NUC (see Fig. 1). In table I, there is a brief of the selected electronics.

B. Power system

A compact and long duration power system have been designed in order to ensure sufficient battery for long deployments. The battery was mounted inside an off-the-shelf

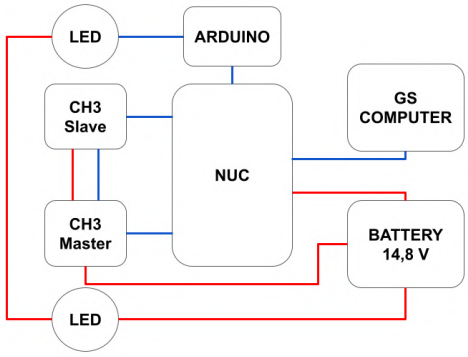


Fig. 1. Electrical scheme. Blue wires represent data connections whilst red cables refer to power wiring.

TABLE I
ELECTRONIC COMPONENTS.

Component description	Units
Board Intel Pro NUC12WSBi7	1
Flir Camera Chameleon3 USB3 31S4C-CS	2
Kowa Lens LM5JC10M	2
Arduino Nano Every	1
Bluerobotics Lumen Subsea Light	2

aluminium enclosure, which has a maximum operating rating of 950 metres in depth.

From the data-sheet provided by the manufacturers, we estimated the theoretical power consumption of the components integrated in the SVS (see table II). Besides, real energy consumption was calculated, considering distinct scenarios (see table III).

Experimental results indicated a significant power demand when the system was fully activated. Therefore, low power strategies were applied to ensure battery life. The operation strategy will be explained in section V-A.

C. Enclosure

The main body of the enclosure (see Fig. 2) has been developed with Acetal Copolymer, which is easy to machine and offers a good resistance to organic and inorganic solvents and acids. The rear cover is made of stainless steel for its excellent mechanical properties and resistance to corrosion in marine environments, as well as to facilitate the heat dissipation of the entire system.

Concerning the protection of the camera lenses, we used acrylic domes with a polycarbonate retaining ring, which guarantee a wide and clear vision and an operation rating of up to 750 metres depth. Regarding communications and power supply, special underwater connectors have been installed on the rear cover, which guarantee a good performance up to a depth of 6000 metres. Table IV details enclosure components.

IV. Design process

3D design and finite element analysis (FEA) was accomplished using *Solidworks 2022* application, whereby it is possible to analyse deformations and stresses when certain loads or forces are applied to a solid body. Through this software, we studied the behaviour of the different components of the enclosure under the conditions of pressure equivalent to 500 metres of water column (MWC).

The first challenge was to design a compact enclosure that would house all the electronic components and still allow easy access to them. The next one, was to verify the strength of the design when external pressure loads were applied to the enclosure walls. It is important to outline that shape is a key factor when designing underwater housings. For example, a spherical shell is widely accepted to be an ideal solution because of the extremely efficient stress and strain distributions in the material. Cylindrical shapes do also perform well under compression forces, therefore they are commonly used for underwater applications [14].

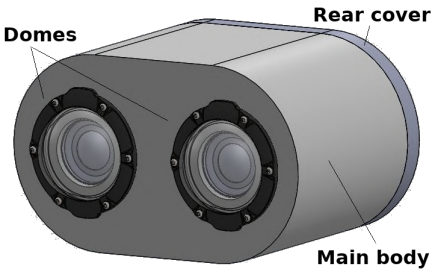


Fig. 2. Stereo camera enclosure.

Considering all factors and requirements exposed above, we defined our enclosure with an oval shape which not only offered more space to store the electronics but also had good enough mechanical properties regarding stress distribution. Thus, the main body of the enclosure measures 246mm in width, has a length of 140mm and a height of 144mm. Inside, two cylindrical rooms of 50mm diameter and 60mm long were designed to store the cameras and their lenses. On the back of the main body, a 40mm depth space has been arranged in order to store the minicomputer, leaving a 20mm wall thickness to ensure sufficient robustness (see Fig. 3).

Another important factor to consider is the O-ring groove design. Special care must be taken when drilling the O-ring groove, as it must fit within the recommended dimensions found in the O-ring groove design guides [15]. In this case, we designed O-ring grooves with the required width and depth for O-rings of 175mm inner diameter and 3mm cross section (rear cover) and of 69.52mm inner diameter and 2.62mm cross section (domes).

A. FEA for pressure load equivalent to 500 MWC

In the following subsection, design parameters of the main body and the rear cover will be presented via interactive FEA studies. It has to be outlined that domes and underwater connectors have not been tested as the manufactured ensures

TABLE II
Theoretical power consumption.

Component	Power (W)	Current (A)
Intel NUC12	120	8.1
Subsea lights	15	1.0
Cameras	6	0.4
TOTAL	141	9.4

TABLE III
Real power consumption.

Component	Current (A)
NUC (idle mode)	0.8
NUC (main node)	3.5
NUC (recording) + cameras + lights	7.0
NUC (recording + inference) + cameras + lights	9.0

TABLE IV
Material list of the enclosure.

Component description	Units
Acetal copolymer body (140x246x144)mm	1
INOX cover (15x246x144)mm	1
Bluerobotics 75mm dome	2
INOX hex button socket screw M5x35mm	22
O-ring (75x2.52)mm	2
O-ring (175x3)mm	1
MacArtney subconn MCBH8M (power)	1
MacArtney subconn DBH8F (data)	1

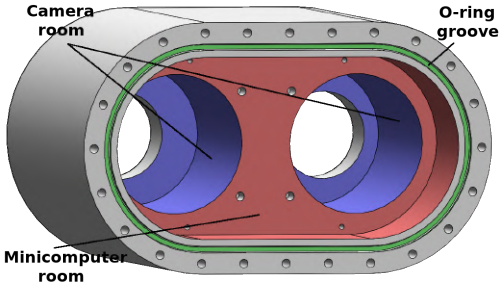


Fig. 3. Enclosure layout.

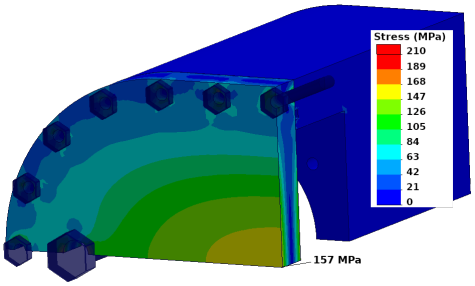


Fig. 4. Results of the simulation process with rear cover width of 15mm.

their resistance up to 750 and 6000 metres below sea level, respectively.

The rear cover of the enclosure was attached to the main body through hexagonal head button sockets M5, which was defined as a threaded bolt in *Solidworks* simulation. Contact definition between the INOX cover and the acetal main body was set as general bonding and the mesh applied was the finest possible in order to get accurate results. Furthermore, aspect ratio and Jacobin values were verified to be lower than 10 so resulting values of stress and deformation would not be distorted.

In order to reduce the computational cost and the elapsed time for the simulation we simplified the model by a fourth of symmetry. Since the pressure is evaluated in MPa we had to apply the conversion between MWC and MPa and then add also the atmospheric pressure (0.1MPa). The total amount of pressure applied to the walls of the enclosure was 5MPa.

The criteria applied for analysing the results obtained from FEA was that the values of stress in the rear cover would not exceed the tensile yield strength (TYS) of the INOX as well as the values of stress in the main body would not pass the tensile ultimate strength (TUS) of the Acetal Copolymer. In table V, the main material properties are listed.

Stress values in the main body of the enclosure were observed to be lower than the threshold TUS of the acetal compolymer, with margin enough to apply a 25% safety factor (see Fig. 4). The most critical part observed was the O-ring groove, in which concentration of stresses occurred as the result of the FEA. This issue could be easily mitigated by applying an increased radii of the fillet as it had been done in [11].

Due to its planar geometry, special attention has been paid when analysing the rear cover. This element became a critical part when high pressure was applied perpendicularly. In order to optimize the design parameters, static analysis were performed for different values of thickness of this material, as shown in table VI.

Results of FEA showed that the thickest rear cover was the best option since stress values were under the threshold of TYS and the deformation percentage remained under the 3%. Besides, 25% margin of operation had to be considered as a safety factor, in order to ensure the enclosure endurance at 500 metres in depth.

V. Software

This sections presents the software workflow of the SVS. Firstly, overall functioning of the SVS is described; and secondly, deep learning object detection model embedded in Lanty is explained.

A. Overall software operation

Lanty was developed mostly in Python programming language and utilises ROS libraries and tools for its operation. The software developed automatically records images every preset time and deactivates cameras and lights when not recording to reduce power consumption.

The workflow of the SVS is shown in Fig. 5. As the first step, the user must preset the periods of time for starting the system and the elapsed time in the recording process. Once the program is running, cameras and lights will be turned on. Next, the recording process will start and images from both cameras will be stored at a rate of 4 frames per second. When the recording process terminates, lights and cameras will be automatically switched off and the object detection model will be launched. The neural network will output fish predictions and automatically discard the pictures if no fishes are found. Finally, the SVS will remain in idle mode until next period is reached.

B. Deep learning object detection model

One of the applications of the developed SVS is fish species detection and classification using a Deep Learning model.

TABLE V
Material properties of the camera enclosure.

Material	TYS (MPa)	TUS (MPa)
Acetal copolymer	-	71,5
Stainless steel 304	210	513

TABLE VI
Stress and deformation values on the rear cover.

INOX width (mm)	Max. stress (MPa)	Max. deformation (mm)
5	844	2.5
10	315	0.8
15	157	0.4

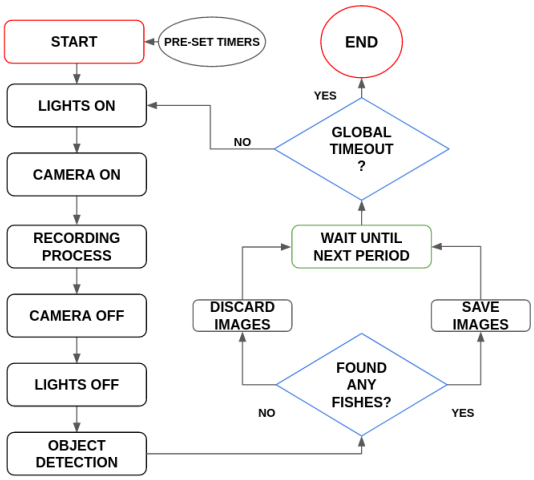


Fig. 5. Software workflow diagram.

The chosen model, YOLOv8-seg, is a state-of-the-art instance segmentation model known for its speed and efficiency. The model generates masks and confidence percentages for detected objects. YOLOv8 incorporates architectural improvements with respect to previous yolo versions, including an anchor-free design and mosaic augmentation, enhancing its generalisation capabilities.

The selection of the YOLOv8 medium version was based on latency-mAP curve analysis, demonstrating the best efficiency-latency ratio for the specified requirements. The model was trained on a private dataset consisting of 2772 images featuring 16 annotated fish classes. These classes include *Diplodus annularis, Diplodus sargus, Diplodus vulgaris, Epinephelus marginatus, Lithognathus mormyrus, Mugilidae prob Chelon, Oblada melanura, Pomatous salator, Sciena umbra, Seriola dumerili, Serranus, Spicara maena, Spondyliosoma cantharus Chromis chromis, Coris julis, and Dentex dentex.*

The trained neural network underwent evaluation on an independent test split of the dataset, achieving an overall mAP50 of 0.8%.

The trained YOLOv8 model was integrated into the ROS framework as an online inference node, analysing the received images to detect potential Mediterranean fishes. Upon detecting a fish with confidence surpassing a predefined threshold, the fish mask information and the image are published as a custom ROS message, and the image is stored. Then, fish point-clouds are generated by combining the information from the fish mask and the 3D points of the observed scene, obtained from the stereo pair. This automated pipeline allows smart data-storing keeping only sequences containing relevant information and offers 3D information of the detected fish species.

VI. Experiments and Results

In this section experiments and obtained results will be presented. This comprises stereo camera calibration and tests in real-world conditions. It has to be outlined that, prior to deployment in underwater conditions, the system was successfully tested in a hyperbaric chamber, demonstrating its effectiveness up to an equivalent depth of 500 MWC.

A. Stereo camera calibration

The main goal of this experiment was to vary the parameters of the cameras, such as focusing, aperture and exposure time, in order to get smooth and sharped images in a range between 60 cm and 100 cm from lens centre. For this purpose, we used two Snellen Charts, both placed on the lower and upper limits of distance of interest, an we tested which configuration gave better results. All the options tested were under the conditions of maximum brightness from the lights. Fig. 6 contains the images resulting from the distinct configurations checked.

Considering the outcome of the test, we selected the camera configuration shown in Fig. 6-(f). Once the parameters had been set into the camera, we successfully calibrate the SVS through ROS calibration software [16].

B. Field test

The overall system was deployed in real-world conditions during field tests conducted along the coasts of Mallorca (see Fig 7). The SVS operated in automatic mode, capturing images at 2-minute intervals, within 10-minute periods. The point-cloud of the scene was calculated using the $stereo\_image\_proc$ package, and the online inference node identified fish species and published the results as ROS topics [17].

VII. Conclusions and Future Work

This work presented a complete hardware and software solution for an underwater SVS. The developed prototype demonstrated resistance and robustness, establishing its suitability for operation in deep-sea environments up to 500 metres. The power system exhibited durability, enabling continuous recording over several days, with successful validation of the automatic recording mode in a real marine environment.

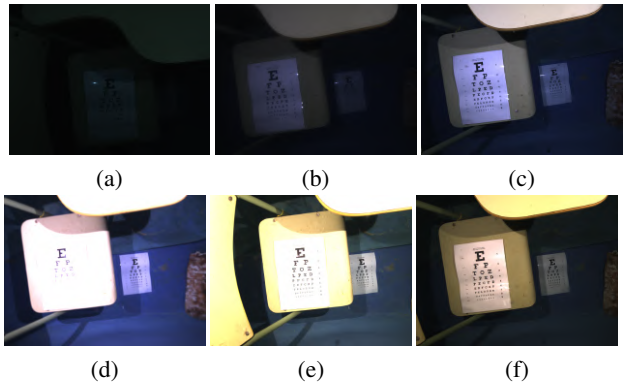
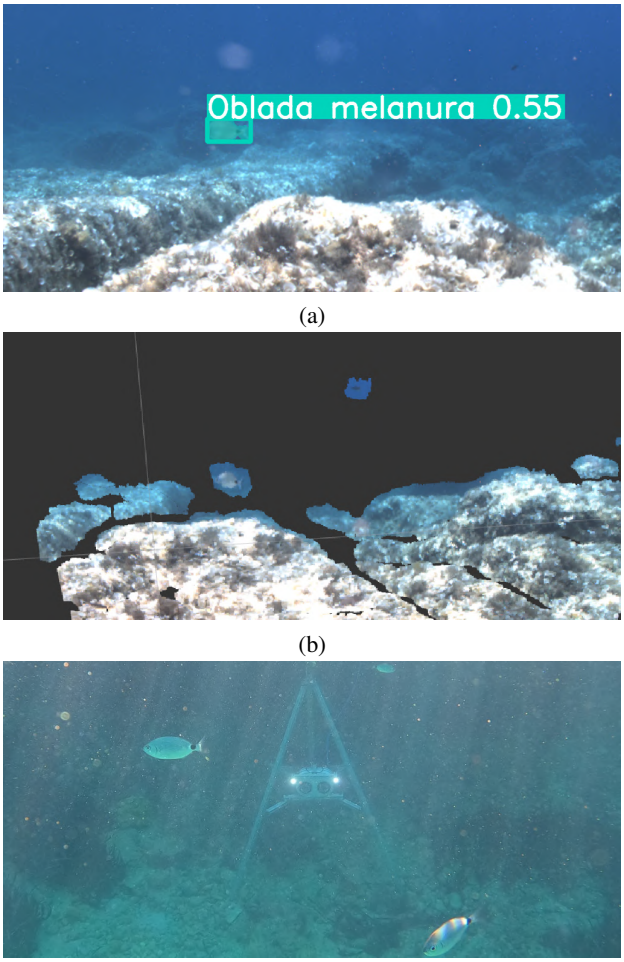


Fig. 6. Focusing test: (a) F-stop: 16, FR: 0.35 and exposure time: 50ms, (b) F-stop: 8, FR: 0.35 and exposure time: 50ms, (c) F-stop: 8, FR: 0.10 and exposure time: 50ms, (d) F-stop: 4, FR: 0.20 and exposure time: 50ms, (e) F-stop: 4, FR: 0.10 and exposure time: 50ms and (f) F-stop: 4, FR: 0.10 and exposure time: 20ms.



(a)

(b)

(c)

Fig. 7. Experiments performed at the coasts of Mallorca: (a) Inference during field test, (b) point cloud of the scene generated in ROS including the detected fish and (c) SVS deployment underwater.

Furthermore, a low-computational-cost neural network, Yolov8-seg, was trained and seamlessly integrated into ROS. This integration allows for efficient data filtering, alleviating the traditionally time-consuming task of manual data review. The application of this network resulted in the extraction of fish masks, which combined with the point cloud information, provided valuable 3D information.

Future work will focus on refining point-cloud information and optimizing its combination with masks in order to achieve accurate fish length and volume estimations. It is also planned to improve the power system by developing a more efficient mode to extend the operation time up to one week. All these improvements will be tested in real deep sea environments.

ACKNOWLEDGMENT

Caterina Muntaner-Gonzalez was supported by the Consejería de Educación y Universidades del Gobierno de las Illes Balears under the contract FPU-2022-010-C CAIB 2022.

This work has been partially sponsored and promoted by the Comunitat Autonoma de les Illes Balears through the Direcció General de Recerca, Innovació i Transformació Digital and the Conselleria de Economia, Hisenda i Innovació via Plans complementaris del Pla de Recuperació, Transformació i Resiliència (PRTR-C17-I1) and by the European Union-Next Generation UE (BIO/002A.1 and BIO/022B.1). Nevertheless, the views and opinions expressed are solely those of the author or authors, and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission are to be held responsible. This work has been partially sponsored and promoted by grant PLEC2021-007525 funded by MCIN/AEI/10.13039/50110001103 and by the European Union NextGenerationEU/PRTR.

REFERENCES

[1] N. Ayache and F. Lustman, "Trinocular Stereo Vision for Robotics", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991.

[2] M. She, Y. Song, J. Mohrmann and K. Koser, "Adjustment and Calibration of Dome Port Camera Systems for Underwater Vision", in Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41 (pp. 79-92). Springer International Publishing.

[3] C.Dominguez, J. Fontes and T. Morato, "A cost-effective video system for a rapid appraisal of deep-sea benthic habitats: The Azor drift-cam", in Methods in Ecology and Evolution, 2021.

[4] B. Phillips, S. Licht, K. Haiat, J. Bonney, J. Allder, N. Chaloux, R. Shomberg and T. Noyes (2019), "DEEPi: A miniaturized, robust, and economical camera and computer system for deep-sea exploration", in Deep-sea research part I: Oceanographic Research Papers, 153, 103136.

[5] J. Giddens, A. Turchik, W. Goodell, M. Rodriguez and D. Delaney (2021), "The national geographic society deep-sea camera system: A low-cost remote video survey instrument to advance biodiversity observation in the deep ocean", in Frontiers in marine science, 1157.

[6] K. Meyer-Kaiser, K. Schrage, S. Suman, J. Bailey and Y. Girdhar, "CATAIN: An underwater camera system for studying settlement in fouling communities at high temporal resolution", in Limnology and Oceanography: Methods, 2023.

[7] A. Delgado, C. Briciu-Burghina and F. Regan (2021), "Antifouling strategies for sensors used in water monitoring: review and future perspectives, in Sensors, 21(2), 389.

[8] F. Oleari, F. Kallasi, D. Lodi-Rizzini, J. Aleotti and S. Caselli (2014), "Performance evaluation of a low-cost stereo vision system for underwater object detection", IFAC Proceedings Volumes, 47(3), 3388-3394.

[9] F. Oleari, F. Kallasi, D. Lodi-Rizzini, J. Aleotti and S. Caselli, "An underwater stereo vision system: From design to deployment and dataset acquisition", in OCEANS 2015-Genova (pp. 1-6), IEEE.

[10] M. Martin-Abadal, G. Oliver-Codina and Y. Gonzalez-Cid, Real-time pipe and valve characterisation and mapping for autonomous underwater intervention tasks", in Sensors 22, no. 21 (2022): 8141.

[11] T. Łuczynski, P. Łuczynski, L. Pehle, M. Wirsum and A. Birk (2019), "Model based design of a stereo vision system for intelligent deep-sea operations", in Measurement, 144, 298-310.

[12] C. Silva, R. Aires and F. Rodrigues, "A compact underwater stereo vision system for measuring fish", in Aquaculture and Fisheries, 2023.

[13] K. Dunkley, A. Dunkley, J. Drewnicki, I. Keith and J. Herbert-Read, " A low-cost, long-running, open-source stereo camera for tracking aquatic species and their behaviours", in Methods in Ecology and Evolution, 2023.

[14] J. Zhang, M. Wang, W. Wang and W. Tang (2017), "Buckling of egg-shaped shells subjected to external pressure", in Thin-Walled Structures, 113, 122-128.

[15] T.S. Solutions (2016), "O-Rings and Back-Up Rings", in Trelleborg Sealing Solutions: Trelleborg, Sweden, 2016.

[16] ROS, "How to calibrate a stereo camera", [online]: http://wiki.ros.org/camera_calibration/Tutorials/StereoCalibration.

[17] ROS, "stereo_image_proc", [online]: http://wiki.ros.org/stereo_image_proc.