An Underwater Stereo Vision System: from Design to Deployment and Dataset Acquisition

Fabio Oleari*, Fabjan Kallasi*, Dario Lodi Rizzini*, Jacopo Aleotti*, Stefano Caselli*
*RIMLab - Robotics and Intelligent Machines Laboratory,
Dipartimento di Ingegneria dell'Informazione, University of Parma, Italy,
Email: {oleari,kallasi,dlr,aleotti,caselli}@ce.unipr.it

***Abstract*—This paper describes the hardware setup of a stereo-vision system for underwater 3D object detection and a camera calibration method that does not require complex underwater procedures. The vision system has been developed within the MARIS Italian national project for underwater cooperative intervention tasks. The hardware architecture is based upon state of the art technologies. The software modules of the system are based on the ROS middleware. A dataset of underwater stereo images of cylindrical pipes has been collected and is made publicly available.**

I. Introduction

Autonomous underwater intervention robotics addresses operations like object grasping, manipulation and transportation. This work describes the vision system developed for the MARIS [3] project (Marine Autonomous Robotics for InterventionS). The vision system has been designed to support object grasping and manipulation from a floating base. Object detection is performed, in the context of the MARIS project, by using a stereo vision system. Stereo vision provides submersible platforms the capability of three-dimensional mapping of sea-floor structures over local areas [9].

While underwater manipulation studies are mainly focused on a single floating vehicle [5], the goal of the MARIS project is the development of technologies and methods for cooperative transportation activities. The reference scenarios are those typical of the off-shore industry, for underwater search and rescue operations, as well as for underwater scientific missions. Since one of the major goals of the MARIS project is the execution of coordinated intervention operations by a team of two AUVs, additional research issues of the MARIS project are those related to mutual exchange of information between the two vehicles and task decomposition. All these tasks require an accurate visual detection and localization of the target object with respect to the AUVs. Besides the vision system, object manipulation in the MARIS project is also supported by reactive control techniques based on force/torque and tactile sensing.

The stereo-vision system presented in this paper includes state of the art components and it extends our preliminary work [10], where a low-cost stereo vision system was developed in order to assess the requirements of underwater vision. Figure 1 shows one of the MARIS AUVs and, in particular, the configuration of the vision canister (which contains the hardware for data processing) and the stereo rig.

The paper also proposes an approach for camera calibration that does not require in-situ underwater procedures [14], [1]. In particular, a two-stage approach for calibration is presented where, in the first stage, a standard in-air calibration is performed. In the second stage, camera parameters are optimized by using prior knowledge about the shape of submerged cylindrical pipes. Moreover, a dataset of about half-an-hour stereo image pairs was acquired by observing pipes laying on the seabed. The dataset comprises image sequences of both single and multiple pipes of different colors.

The paper is organized as follows. Related works about underwater stereo-vision are summarized in Section II. Section III describes the underwater vision system and reports performance results regarding power consumption, thermal dissipation and network bandwidth. Section IV presents the proposed approach for stereo camera calibration. Section V describes the underwater dataset. Section VI concludes the work.

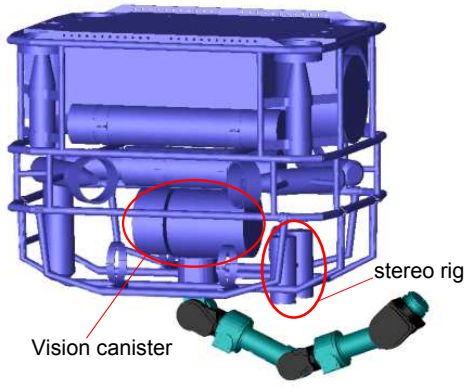

Fig. 1. 3D CAD model of the MARIS AUV. The horizontal stereo rig points to the bottom of the vehicle.

II. Related work

In [16] an embedded stereo-vision system for underwater object detection was presented based on FPGA technology. The system consumed 115.25 mW and achieved a throughput of 26.56 frames per second (800x480 pixels). A standard in-air calibration procedure was adopted. Praveen et al. [12] proposed an adaptive weight-based cross-correlation

method to estimate the disparity map for a pair of underwater images. The authors also pointed out the lack of a dataset for underwater image calibration and evaluation of stereo correspondence algorithms. The underwater dataset collected in this work contributes in this direction. Wehkamp et al. [19] described a workflow for stereoscopic measurements for marine biologists by providing instructions on how to assemble an underwater stereo-photographic system with two digital consumer cameras with underwater calibration. Ortiz et al. [11] proposed a single-camera vision system for real-time underwater cable tracking to detect power cables laid on the seabed. Sarafraz et al. [17] proposed a method to improve the correspondence problem in underwater stereo by exploiting the depth cues in the backscatter components.

Rahman et al. [14] state that in underwater scenario light refraction through multiple media (air, water, lens, etc.) leads to an increment of the radial distortion. In the literature, different approaches for camera calibration correction have been exploited. In particular, starting from the Brown lens distortion model [2] several automatic correction methods have been proposed. Gonzalez-Aguilera et al. [6] presented an iterative numerical approach for the automatic estimation and correction of radial lens distortion using a single image. The proposed method used several geometric constraints as rectilinear lines and vanishing points of a single image acquired in outdoor environment. Although automatic lens-distortion correction has been studied mainly for outdoor environments, methodologies have also been applied in medical and surgical scenarios. Cong et al. [4] proposes a method for radial distortion correction using spherical projection model. A sequence of images of a marker board at different positions are captured by the endoscope to construct a spherical model of the lens. Lee et al. [8] proposed a fully automatic Hough-entropy-based calibration algorithm, which used a Parallel-Line-Pattern to estimate the parameters of the lens distortion model.

III. Overview of the vision system

A. Hardware setup

The underwater vision system was designed balancing power consumption, thermal dissipation and system performance. The hardware setup, shown in Figures 2 and 3, includes two industrial AVT Mako G125C GigE cameras (60.5x29x29mm size, 30 fps, 1292x964 resolution) with ethernet connection. The cameras are housed in two separate small canisters made of black PVC, with a stainless steel back cover and a plexiglass transparent glass. Power supply, camera triggering and data cables are joined into a single 13pin submarine cable that enters the canister from the back steel cover (Figure 4). The vision canister is made of a stainless steel pipe, closed on one side. The other side is sealed by a cover to which internal supports are attached. All I/O connections go through this cover using submarine connectors (Figure 3). Internal supports hold the electronic components and ensure heat dissipation while keeping the system easy to maintain. Each part of the canisters has been worked out to avoid soldering or gluing. Redundant O-rings have been used for junctions thus guaranteeing watertight protection up to 50 m depth.

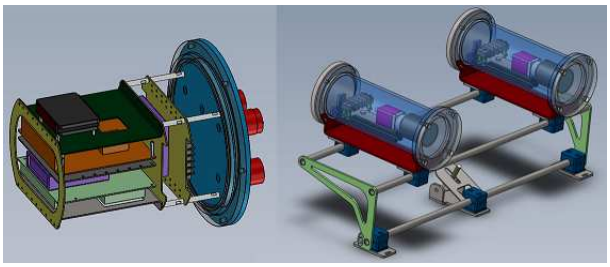

Fig. 2. 3D CAD model of the vision system: vision canister (left) and stereo rig (right).

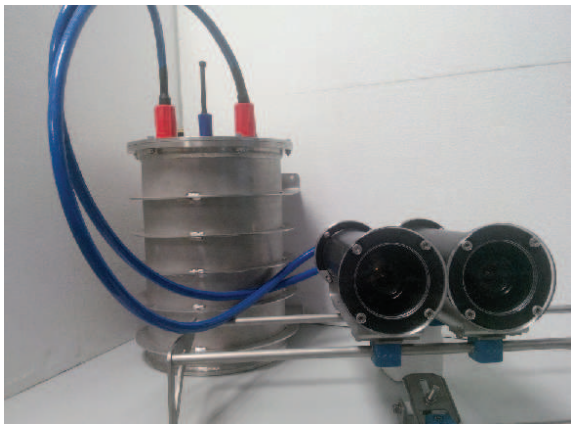

Fig. 3. Underwater vision system: vision canister (left) and the stereo rig (right) with the two camera canisters.

The ECU (Electronic Control Unit) inside the vision canister was designed as a modular system (Figure 4), including two x86 CPUs, one ARM-based board and a microcontroller. The microcontroller (Arduino MEGA 2560) is responsible for camera triggering. The main system relies on an Intel Core i7-4770TE @3.33GHz with 4 physical cores, mounted on an industrial Mini-ITX board. This PC is equipped with 8Gb of DDR3 RAM and two 120Gb Samsung SSDs. The second PC is based on an Intel DN2800MT Mini-ITX mainboard with an ATOM 1.86GHz processor (2Gb RAM, 32Gb mini-PCIe SSD). The CPU has a low TDP of 6.5W. The ARM-based board consists of a Raspberry-Pi, chosen due to its onboard MPEG-2 video coder. This feature will be used for coding images in a standard H264 video streaming for remote monitoring. Network connection is managed by two independent gigabit networks using Netgear GS-105 switches. Due to the high traffic rate generated by the vision system, the two cameras and the i7 ECU are connected by a dedicated network to avoid traffic jams. The second ethernet interface is used to enable communication between the vision system and the UAV.

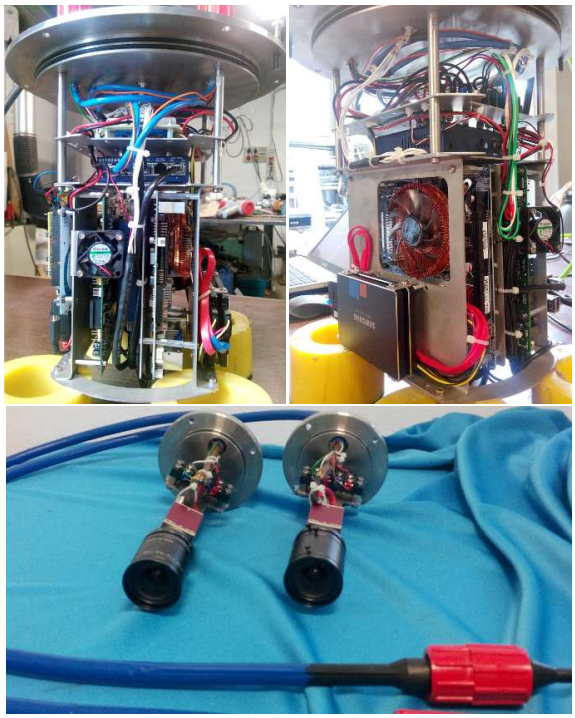

Fig. 4. Components of the vision system: internal view of the vision canister (top images); AVT Mako cameras, optics and cabling (bottom image).

B. Hardware performance evaluation

To evaluate system performance, experiments have been conducted both in a laboratory setup and in a real underwater environment. In particular, once the assembly stage of the hardware was completed water resistance experiments were performed for the vision and the camera canisters. In the laboratory setup canisters have been inflated with air at 6 atm pressure to simulate underwater conditions at 50 m depth. Then, the canisters have been placed for 24 hours inside a water recipient to identify possible leaks. A pressure gauge was used to keep pressure monitored. Underwater tests have been conducted with the canisters sinked at 50 m depth in the see in front of La Spezia (Italy). No water leakages occurred inside the canisters.

Long term experiments have been conducted in the laboratory setup to evaluate power consumption, heat dissipation, working temperature and network performance. Power consumption and heat dissipation have been tested by stressing the CPUs at a high load using the Linux `stress` command. Results are reported in Table I. The power consumption is far below 100 W, thus achieving a good performance for autonomous robot applications. The maximum temperature registered in the vision canister in water with all the CPUs stressed was 60°C, far below a dangerous value. In the laboratory setup, with the vision canister in air, temperature inside the canister after 15 minutes of CPU stressing reached 72°C. Hence, the vision canister can also be operated in air at standard CPU load level for a long time. Bandwidth throughput of both networks has been tested exploiting `iperf` [7], [18], a state of the art tool for network benchmarking. Experiments have been conducted using both TCP and UDP protocols, half and full duplex communication and different packet size. Data transfer rate was about 850 Mbps, full duplex, for both system-to-cameras communication and system-to-vehicle communication. Results were computed from about 1TB of transferred data. The performance of the vision system to reconstruct a disparity map was tested using the standard Absolute Differences (SAD) correlation method A throughput of 12.5 frames per second (1292x964 resolution) was achieved.

Test	Power (W)
Idle	43.2
CPU stress - 1 core - i7	65.0
CPU stress - 2 cores - i7	78.0
CPU stress - 4 cores - i7	79.2 (peak 86.4)
CPU stress - 4 cores/8 threads - i7	79.2 (peak 86.4)
CPU stress - 2 cores/4 threads - Atom	44.9
Stereo Vision frames acquisition	46.6
CPU stress on all cores/CPUs	**81.6 (peak 91.9)**

TABLE I
POWER CONSUMPTION TESTS AT DIFFERENT CPU LOADING LEVELS

IV. CALIBRATION AND SOFTWARE

Due to the difficulty of in-situ underwater camera calibration we propose a two-stage approach for calibration. In the first stage standard in-air calibration is performed. In the second stage camera parameters are optimized exploiting prior knowledge about the 3D shape of the observed submerged pipes. In particular, we provide an automatic method of radial distortion correction based on rectification of cylindrical surfaces. We do not consider the decentering distortion effect on the images as it is not significant. The algorithm proposed in this paper exploits a set of stereo images from the dataset to improve the robustness of radial distortion correction. For image acquisition a camera driver was built upon the AVT SDK integrated with the ROS (Robot Operating System) framework [13]. A synchronization module was also developed so that all images are time-stamped using the ROS timer.

A. Distortion model

The lens distortion model used in this work is called *plumb bob* [2] and describes the correction applied to each image pixel as:

$$\begin{aligned}\Delta_x &= \bar{x}(K_1 r^2 + K_2 r^4 + K_3 r^6 + \cdots) \\ &\quad + [P_1(r^2 + 2\bar{x}^2) + 2P_2\bar{x}\bar{y}][1 + P_3 r^2 + \cdots] \quad (1)\\ \Delta_y &= \bar{y}(K_1 r^2 + K_2 r^4 + K_3 r^6 + \cdots) \\ &\quad + [2P_1\bar{x}\bar{y} + P_2(r^2 + 2\bar{y}^2)][1 + P_3 r^2 + \cdots] \quad (2)\end{aligned}$$

where

$$\begin{aligned}\bar{x} &= x - x_p \\ \bar{y} &= y - y_p \\ r &= [(x - x_p)^2 + (y - y_p)^2]^{\frac{1}{2}} \quad (3)\end{aligned}$$

and (x_p, y_p) are the coordinates of the principal point of the camera. Since relevant effects of the radial distortion in air are determined by the two scalar values K_1 and K_2, the aim of the proposed approach is to achieve an automatic correction of these two parameters. The goal is to estimate the distortion correction $\Delta_K = (\Delta_{K_1}, \Delta_{K_2})$ so that the new parameters K_1' and K_2' defined as

$$\begin{aligned} K_1' &= K_1 + \Delta_{K_1} \\ K_2' &= K_2 + \Delta_{K_2} \end{aligned}$$

get close to the radial distortion parameters that could be obtained with an underwater calibration. If Δ_{K_1} and Δ_{K_2} are set to 0 then no correction is applied to the in air calibration and the result of the reconstruction is that the cylinder is distorted.

B. Calibration algorithm

The algorithm works on a set of N stereo image pairs of the pipe observed at different positions and orientations. A total of 2500 disparity maps are generated from each image pair by sampling the distortion correction in the range $\Delta_K \in [0; 0.300] \times [0; 0.300]$ with a fixed step of 0.006 in both dimensions. Disparity maps are then filtered to segment the pipe from the environment. Segmentation is performed by applying the same binary mask (manually defined) to all disparity maps. A point cloud representing the pipe is computed from each segmented disparity map.

It must be remarked that since the disparity map is noisy and only part of the pipe surface is visible fitting the point cloud with a cylinder to evaluate distortion would not be feasible. Principal Component Analysis (PCA) is applied, instead, to fit each point cloud representing the pipe with a plane translated on the point cloud centroid. The first two eigenvectors (associated to the two greatest eigenvalues) define the direction vectors of the plane. The third eigenvector defines the normal vector of the plane. Next, a point-to-plane distance is computed for each point in the point cloud. The sum of all point-to-plane distances is related to the distortion of the pipe and it is used to evaluate the effectiveness of the Δ_K correction. The greater the sum, the greater the distortion of the pipe.

For each stereo image pair a 50×50 matrix M_i is generated, where each element is the sum of all the point-to-plane distances for any sampled Δ_K correction. Since M_i are computed from different stereo images matrices values are uncorrelated. The best $\overline{\Delta_K^*}$ correction, which minimizes distortion, is estimated as

$$\overline{\Delta_K^*} \triangleq \arg\min_{\Delta_K} \overline{M}(\Delta_K) = \arg\min_{\Delta_K} \sum_{i=1}^{N} M_i(\Delta_K) \quad (4)$$

where $\overline{M}$ is the sum over all the stereo image pairs. Figure 5 shows an example of the sampled distortion function $\overline{M}$. Finally, $\overline{\Delta_K^*}$ is refined by finding the minimum of the paraboloid Δ_K^* fitting $\overline{M}$ in the neighborhood of $\overline{\Delta_K^*}$.

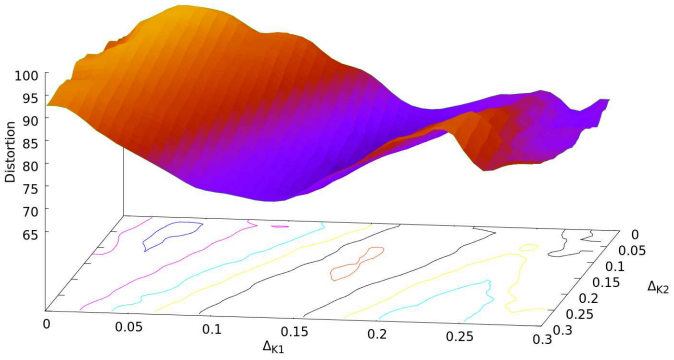

Fig. 5. Example of a sampled distortion function $\overline{M}$.

N	$\overline{\Delta_K^*}$	Δ_K^*
1	$\Delta^*_{K1} = 0.210$ $\Delta^*_{K2} = 0.042$	$\Delta^*_{K1} = 0.213176$ $\Delta^*_{K2} = 0.037045$
2	$\Delta^*_{K1} = 0.078$ $\Delta^*_{K2} = 0.282$	$\Delta^*_{K1} = 0.084486$ $\Delta^*_{K2} = 0.207892$
3	$\Delta^*_{K1} = 0.156$ $\Delta^*_{K2} = 0.150$	$\Delta^*_{K1} = 0.154559$ $\Delta^*_{K2} = 0.150441$
4	$\Delta^*_{K1} = 0.204$ $\Delta^*_{K2} = 0.042$	$\Delta^*_{K1} = 0.204600$ $\Delta^*_{K2} = 0.036582$
5	$\Delta^*_{K1} = 0.156$ $\Delta^*_{K2} = 0.144$	$\Delta^*_{K1} = 0.159081$ $\Delta^*_{K2} = 0.147181$
6	$\Delta^*_{K1} = 0.156$ $\Delta^*_{K2} = 0.150$	$\Delta^*_{K1} = 0.158008$ $\Delta^*_{K2} = 0.150319$
7	$\Delta^*_{K1} = 0.156$ $\Delta^*_{K2} = 0.150$	$\Delta^*_{K1} = 0.158090$ $\Delta^*_{K2} = 0.150308$
8	$\Delta^*_{K1} = 0.156$ $\Delta^*_{K2} = 0.144$	$\Delta^*_{K1} = 0.157092$ $\Delta^*_{K2} = 0.147027$
9	$\Delta_{K1} = 0.156$ $\Delta^*_{K2} = 0.144$	$\Delta^*_{K1} = 0.157127$ $\Delta^*_{K2} = 0.146932$

TABLE II
ESTIMATED CORRECTION PARAMETERS AT DIFFERENT NUMBER N OF STEREO IMAGES.

Table II shows results of the proposed method for automatic calibration. In particular, several tests are reported at different number N of stereo images from the whole dataset. Results show that 5 stereo image pairs are enough to achieve a good estimation of the distortion parameters. Further increasing the number of images does not significantly reduce distortion. Δ^*_{K2} shows a higher variance, thus confirming that it affects radial distortion less than Δ^*_{K1}. The computational time is about 28 minutes on a Intel Core i7-4770 @3.40GHz with 8Gb of DDR3 RAM for each image pair. Figures 6 and 7 show results of the calibration procedure.

V. UNDERWATER DATASET

An underwater acquisition campaign took place near Portofino (Italy) on September 6th, 2014. The underwater stereo vision system was submerged at a depth of 10 m together with a set of cylindrical pipes with different colors and radii ranging from 5 to 6 cm. A team of divers provided lo-

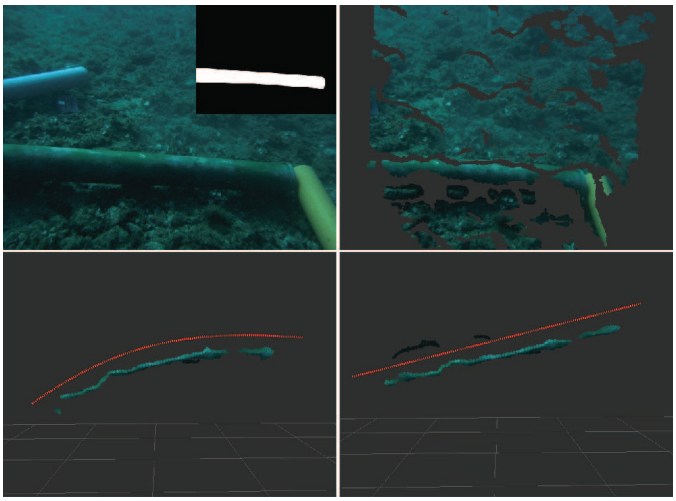

Fig. 6. Example of the automatic calibration approach. Left camera image and segmented pipe (top left). 3D point cloud from disparity map (top right). Distorted pipe using in-air calibration parameters (bottom left). Pipe rectification using optimized camera parameters (bottom right). Red dotted line is displayed to ease the graphical interpretation.

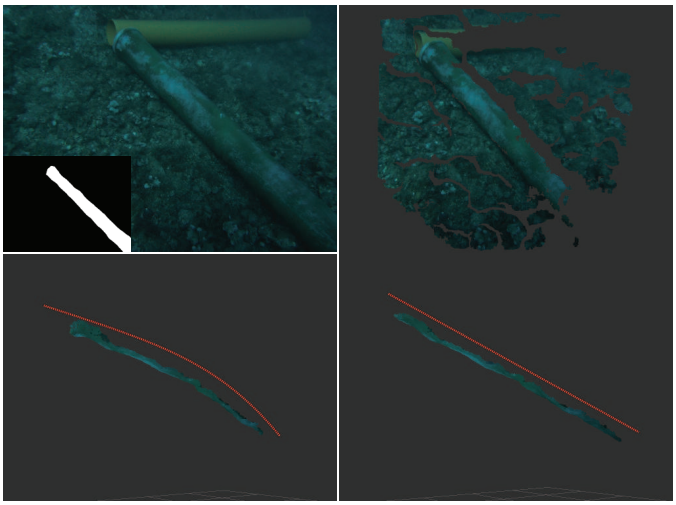

Fig. 7. A second example of the automatic calibration approach. Images are organized as in Figure 6.

gistic and technical support as shown in Figure 8. A dataset of was acquired including stereo image pairs of the pipes laying on the seabed. The dataset, which comprises image sequences of both single and multiple pipes, is available for download at http://rimlab.ce.unipr.it/Maris.html. The dataset include 10123 stereo images in Bayer encoded format (1292x964 resolution), the original in-air camera calibration parameters, and the one obtained after calibration with the method described in Section IV (.yaml files). The dataset was acquired at 15 frames per second.

VI. Conclusions

This paper described the hardware setup of an underwater stereo-vision system for detection and localization of objects laying on the seabed to enable cooperative object transportation tasks. Results of laboratory tests have been reported regarding power consumption, heat dissipation and network performance. A method for camera calibration which exploits the shape of cylindrical submerged objects was also proposed. A dataset of stereo vision images in sea water was also collected and it was made publicly available for download. Future works will involve the integration of the vision system in the MARIS AUVs, the investigation of advanced strategies for image enhancement [15], and the development of algorithms for accurate object pose determination. Further field experiments will also be performed to evaluate the performance of the stereo-vision system in both pool and sea waters.

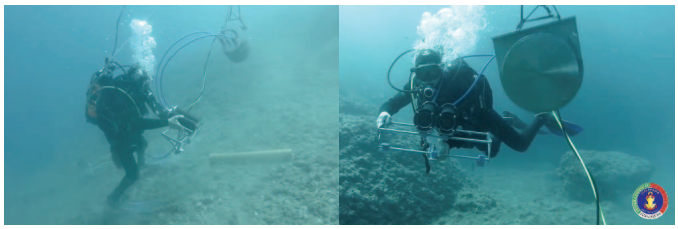

Fig. 8. Vision system submerged at 10 m depth by a team of divers.

Acknowledgment

The system has been developed within the Italian national Project MARIS. The authors would like to thank the divers of Federazione Italiana Attività Subacquee of Parma (FIAS, http://www.fiasparma.it/) for their valuable support. Canisters have been manufactured by G.A.M.P S.r.l, La Spezia, Italy.

References

[1] V. Brandou, A-G Allais, M. Perrier, E. Malis, P. Rives, J. Sarrazin, and P-M Sarradin. 3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS. In *OCEANS 2007*, pages 1–6, June 2007.

[2] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.

[3] G. Casalino, M. Caccia, A. Caiti, G. Antonelli, G. Indiveri, C. Melchiorri, and S. Caselli. MARIS: A national project on marine robotics for interventions. In *22nd Mediterranean Conference of Control and Automation (MED)*, pages 864–869, June 2014.

[4] W. Cong, J. Yang, W. Deng, and J. Zhu. Automatic radial distortion correction for endoscope image. In *Image and Signal Processing (CISP), 2013 6th International Congress on*, volume 2, pages 932–937, Dec 2013.

[5] J.J. Fernandez, M. Prats, P.J. Sanz, J.C. Garcia, R. Marin, M. Robinson, D. Ribas, and P. Ridao. Grasping for the Seabed: Developing a New Underwater Robot Arm for Shallow-Water Intervention. *IEEE Robotics Automation Magazine*, 20(4):121–130, Dec 2013.

[6] D. Gonzalez-Aguilera, J. Gomez-Lahoz, and P. Rodriguez-Gonzalvez. An Automatic Approach for Radial Lens Distortion Correction From a Single Image. *IEEE Sensors Journal*, 11(4):956–965, April 2011.

[7] Lawrence Berkeley National Laboratory. iperf3: A TCP, UDP, and SCTP network bandwidth measurement tool. https://github.com/esnet/iperf, March 2015.

[8] Tung-Ying Lee, Tzu-Shan Chang, Chen-Hao Wei, Shang-Hong Lai, Kai-Che Liu, and Hurng-Sheng Wu. Automatic distortion correction of endoscopic images captured with wide-angle zoom lens. *IEEE Transactions on Biomedical Engineering*, 60(9):2603–2613, Sept 2013.

[9] S. Negahdaripour and H. Madjidi. Stereovision imaging on submersible platforms for 3-D mapping of benthic habitats and sea-floor structures. *IEEE Journal of Oceanic Engineering*, 28(4):625–650, Oct 2003.

[10] F. Oleari, F. Kallasi, D. Lodi Rizzini, J. Aleotti, and S. Caselli. Performance Evaluation of a Low-Cost Stereo Vision System for Underwater Object Detection. In *19th World Congress of the Intl Federation of Automatic Control (IFAC)*, pages 3388–3394, Aug. 2014.

[11] A. Ortiz, M. Simó, and G. Oliver. A Vision System for an Underwater Cable Tracker. *Mach. Vision Appl.*, 13(3):129–140, July 2002.

[12] K.P.U. Praveen and C.J. Prabhakar. Stereo correspondence for underwater images using adaptive weight-based cross-correlation. In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 348–353, Aug 2013.

[13] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. ROS: an open-source Robot Operating System. In *ICRA Workshop on Open Source Software*, 2009.

[14] T. Rahman, J. Anderson, P. Winger, and N. Krouglicof. Calibration of an underwater stereoscopic vision system. In *IEEE Oceans*, pages 1–6, Sept 2013.

[15] M. Roser, M. Dunbabin, and A. Geiger. Simultaneous underwater visibility assessment, enhancement and improved stereo. In *IEEE Intl Conference on Robotics and Automation (ICRA)*, pages 3840–3847, May 2014.

[16] C. Sanchez-Ferreira, J.Y. Mori, C.H. Llanos, and E. Fortaleza. Development of a stereo vision measurement architecture for an underwater robot. In *IEEE Fourth Latin American Symposium on Circuits and Systems (LASCAS)*, pages 1–4, Feb 2013.

[17] A. Sarafraz, S. Negahdaripour, and Y.Y. Schechner. Performance assessment in solving the correspondence problem in underwater stereo imagery. In *OCEANS 2010*, pages 1–7, Sept 2010.

[18] Ajay Tirumala, Feng Qin, Jon Dugan, Jim Ferguson, and Kevin Gibbs. Iperf: The TCP/UDP bandwidth measurement tool. *https://iperf.fr/*, 2005.

[19] M. Wehkamp and P. Fischer. A practical guide to the use of consumer-level digital still cameras for precise stereogrammetric in situ assessments in aquatic environments. *Underwater technology*, 32(2):111–128, July 2014.