

Towards omnidirectional immersion for ROV teleoperation

Josep Bosch, Pere Ridao, Rafael Garcia and Nuno Gracias

Computer Vision and Robotics Institute (VICOROB), Universitat de Girona, 17003

{j.bosch@udg.edu, pere@silver.udg.edu, rafa@silver.udg.edu, ngracias@silver.udg.edu}

Abstract

The use of omnidirectional cameras underwater is enabling many new and exciting applications in multiple fields. Among them, it will allow Remotely Operated Underwater Vehicles (ROVs) to be piloted directly by means of the images captured by omnidirectional cameras through virtual reality (VR) headsets. This immersive experience will extend the pilot's spatial awareness and reduce the usual orientation problems during missions. This paper presents this concept and illustrates it with the first experiments for achieving this purpose.

Keywords: Immersive teleoperation, ROV, omnidirectional image, VR.

1 INTRODUCTION

In the last few years, omnidirectional cameras have received increasing interest from the computer vision community, in tasks such as mapping, augmented reality, visual surveillance, motion estimation, and simultaneous localization and mapping (SLAM). The use of omnidirectional cameras in underwater environments opens the door to several new technological applications in fields as diverse as underwater robotics, marine science, oil and gas industries, underwater archeology and public outreach. However, due to the challenges of the underwater medium, the use of these cameras is still very limited compared with land and air applications.

The integration of omnidirectional cameras with underwater robots is expected to have a large impact in both Autonomous Underwater Vehicles (AUVs) and Remotely Operated Vehicles (ROVs) [1]. For AUVs, the wide field of view of the cameras is very convenient for mapping tasks and visual SLAM [2], especially in confined or cluttered environments. In the case of ROVs, it will allow them to be piloted directly, using the images captured by omnidirectional cameras through virtual reality headsets. This immersive experience will extend the pilot's spatial awareness and reduce the usual orientation problems during missions.

In this paper we focus on this concept, analysing the advantages and challenges of this new approach, and present the first experiments towards achieving a prototype of this system.

At present the exploration of the seafloor using an ROV in an oceanographic campaign requires one or more professional ROV pilots who are only responsible for commanding the ROV, and a group of scientists who are interested in exploring specific areas of the sea according to their disciplines. Both the pilots and scientists observe display monitors with live video feed from one or multiple conventional cameras located in the robot, to decide where to move the robot and which areas to explore. On frequent occasions the cameras show scenes with no recognisable landmarks, which affects the pilot's spatial orientation and the awareness of the area surveyed. The use of a virtual reality headset with a 360° coverage of the area surveyed can help both pilots and scientists to have better spatial awareness and save time and resources while performing missions.

Teleoperation of underwater robots is typically performed using joysticks or gamepads, which allow high precision commands with a user-friendly interface for pilots. When using VR headsets, the user is not able to see the real world while wearing the headset, as he only visualizes the live feed coming from the vehicle. This fact could make the task of operating the vehicle without removing the VR headset very complex. In order to simplify the teleoperation under such conditions a new approach has been studied, which takes advantage of the recent developments in the gesture-tracking field for VR. Furthermore, the use of these techniques can facilitate the teleoperation task within the limitations that non-expert pilots such as archaeologists, biologists or geologists could operate the robot by themselves.

Figure 1 illustrates the concept presented in this paper: an inexperienced pilot uses a VR headset to supervise a mission performed by an ROV in a full-immersive experience while being able to operate it easily by sending commands through hand-gesture recognition.



Figure 1: A pilot tele-operating a ROV using a VR headset for achieving an immersive experience and sending commands using gesture recognition.

A selection of the related work can be found in section 2. The equipment used for experiments is described in section 3. Section 4 and section 5 present the composition of the panoramas and the teleoperation technique used, respectively. Section 6 presents practical results. Finally, section 7 draws some brief conclusions and makes suggestions for future work.

2 SELECTED RELATED WORK

One of the most known and widely used applications of omnidirectional cameras on land, is Google Street View [3]. This globally known technology featured in Google Maps provides panoramic views from different positions, on streets from thousands of cities around the world and has proved the value of omnidirectional cameras for practical applications. Following the same concept, the authors of this paper presented the concept of an underwater version of this product in a previous work [4].

On the other hand, the problem of developing new tools for improving locational awareness for ROV pilots is an issue that has attracted attention in the underwater scientific community [5]. With the rapid evolution of VR headsets in recent years, many works related to immersive virtual reality have been published recently [6, 7] and some authors have applied this concept to the underwater domain. Domingues et al. [8] proposed a semi-immersive Human Robot Interface (HMI), using a large screen with active stereoscopy and a force feedback device for teleoperation. Garcia et al. [9] proposed, for the first time, the use of a VR headset and teleoperation using a gesture-based method and tested it in a virtual environment using a simulator. Our work has similarity with Garcia's, but it presents results with omnidirectional video collected from experiments in real environments, and real tests with gesture-based teleoperation on an ROV. At the same time different authors [10] worked on data-compression techniques

in order to send panoramic video in real time for diverse applications.

3 EQUIPMENT

In order to illustrate the concept, preliminary experiments have been carried out during 2015 using the equipment described in this section.

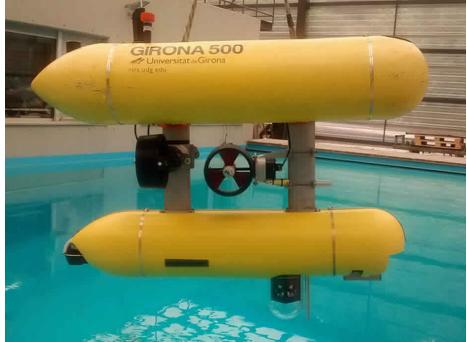
3.1 GIRONA500 AUV

The Girona500 [11] (Fig. 2a) is a reconfigurable vehicle which can be used in autonomous mode (AUV mode) or remotely operated mode (ROV mode). It is rated for 500 m depth and is hovering-capable with 5 thrusters actuated in 4 DOFs: surge, sway, heave and yaw. It is stable in pitch and roll. Its navigation sensor suite includes GPS, a pressure sensor, a Doppler Velocity Log (DVL), and an attitude and heading reference unit (AHRS) that makes it able to navigate precisely. In addition, for long underwater missions, the vehicle can be equipped with an ultra short base line (USBL) device to avoid drifting. The navigation is based on an extended Kalman filter (EKF) that combines input from the sensors to obtain a robust estimation of the AUV position and velocity. The vehicle has a large payload capacity (up to 35 liters) which allows it to carry extra equipment. In the case of the work described in this paper, this area was used to accommodate the omnidirectional camera.

3.2 OMNIDIRECTIONAL CAMERA

The omnidirectional multi-camera system (OMS) is based on a Point Grey's Ladybug 3 [12] which comprises 6 individual cameras and is designed for land applications. A custom housing was designed to make it submersible up to 60 meters depth (Fig. 2b). The housing is composed of a transparent poly methyl methacrylate (PMMA) dome, which encapsulates the camera, and a body made of aluminum alloy, which contains a small form factor computer, dedicated to processing the video feed. The camera outputs 6 separate images that can be combined to create an hemispherical panorama (Fig. 3) using the calibration parameters of the camera. Due to the waterproof housing and the changes in the optics required to obtain a complete hemispherical image in water, the factory-provided calibration is not valid, and a custom calibration procedure for underwater OMS was developed.

Calibration of multi-camera systems typically covers two different sets of parameters: intrinsic parameters, concerning the image formation geome-



(a) The omnidirectional camera integrated with the Girona500 AUV.



(b) Omnidirectional underwater camera used for the survey.

Figure 2: Equipment used for data collection.

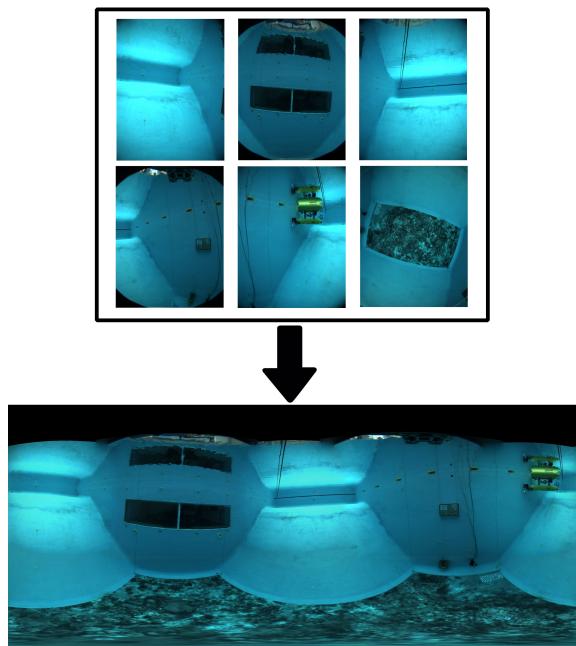


Figure 3: The omnidirectional camera outputs six individual images that are joined into a panorama.

try for each individual camera, and extrinsic parameters, which describe the relative positions and orientations between cameras. In omnidirectional multi-camera systems, the calibration of the extrinsic parameters is an important challenge, due to the usually small overlap between neighboring cameras. Furthermore, for underwater cameras it must be taken into account that the direction of the rays of light changes at every medium transition along the path from a point in water to the imaging sensor inside the camera. In order to model accurately the distortion due to this effect, it becomes essential to explicitly model and simulate the intersection of each light ray with the different media. Due to this phenomenon, any small variation in the estimated relative position of the housing can significantly affect the final direction of the rays and generate projection errors.

The calibration procedure was done in three different stages. The first consisted of the estimation of the intrinsic parameters, which was done separately for each single camera in air and without the waterproof housing. The second stage consisted of the estimation of the extrinsic parameters, carried out under the same conditions as the first step. Finally, the last stage took place underwater and estimated the camera pose with respect to the waterproof housing. The full design and calibration of the camera can be found in [13].

3.3 VIRTUAL REALITY HEADSET

A virtual reality headset is a stereoscopic head-mounted display able to track the head motion using multiple tracking sensors for the purpose of computer games, 3D simulations or any other virtual immersive experience. The Virtual Reality headset used in this paper is the Oculus Rift Development Kit 2 (DK2) [14], a development version of the final Oculus Rift headset that will be released during 2016. The DK2 version has a field of view of 100° with a resolution of 960x1080 pixels per eye and a refresh rate of 75Hz with a pixel persistence of 3ms. The device is equipped with gyroscope, accelerometer and magnetometer updating at 1000Hz that allow it to precisely track the motion of the user's head.

3.4 HAND GESTURE RECOGNITION DEVICE

A new approach for robot teleoperation has been studied using a controller able to track the hands of a user at a high frequency. The Leap Motion controller [15] is a small peripheral device which is designed to be placed on a physical desktop, facing upward (Fig. 4), or it can be mounted onto a virtual reality headset, which would be very useful



Figure 4: Hands motion being tracked by the Leap Motion Device.

in the concept presented in this paper. The device uses two monochromatic infra-red cameras and three infrared LEDs to observe a roughly hemispherical area up to a distance of about 1 meter. The LEDs generate pattern-less IR light and the cameras generate almost 200 frames per second of reflected data.

4 PANORAMA COMPOSITION

As seen in section 3.2 the camera outputs six separate images that must be combined appropriately into a panoramic or omnidirectional image for their correct interpretation (Fig. 3). Making use of the calibration of the cameras, each pixel of the output images can be associated with a 3D ray in the underwater space. Except for the small area where there is image overlap, it is not possible to estimate the distance to the objects from just one set of images acquired at a single location. For this reason, for visualization purposes, the world around the camera is assumed to be a sphere, where all of the points sensed by the camera are at a constant preselected distance. Once the sphere radius is defined, a spherical point cloud is quick to compute, and it can be easily loaded into a computer 3D-viewer or re-projected into a 2D image; the equi-rectangular projection being the most widely used.

Given a world point in Cartesian coordinates $Q = (X, Y, Z)$, it can be converted to spherical coordinates (Figure 5) $Q = (\theta, \phi, R)$ through Equations 1–3.

$$R = \sqrt{X^2 + Y^2 + Z^2} \quad (1)$$

$$\theta = \text{atan}2(Y, X), -\pi \leq \theta \leq \pi \quad (2)$$

$$\phi = \text{acos} \left(\frac{Z}{R} \right), 0 \leq \phi \leq \pi \quad (3)$$

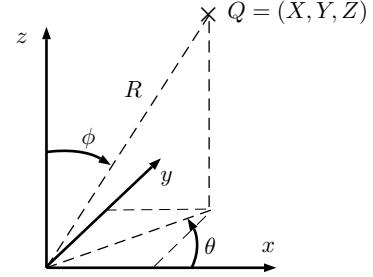


Figure 5: Conversion from Cartesian to spherical coordinates.

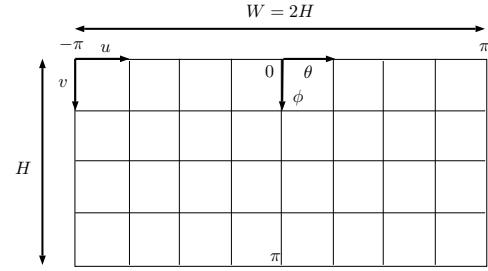


Figure 6: Scheme and notation for the equirectangular projection.

The equirectangular projection projects a given point Q to a cylinder (Figure 6) through Equations 4 and 5:

$$u = \frac{\theta + \pi}{2\pi} \cdot W \quad (4)$$

$$v = \frac{\phi}{\pi} \cdot H \quad (5)$$

The inverse equations are:

$$\theta = \frac{u \cdot 2\pi}{W} - \pi \quad (6)$$

$$\phi = \frac{v \cdot \pi}{H} \quad (7)$$

The first step when composing a panorama is choosing its parameters: projection type, projection distance and final size. For every pixel of the panorama, the 3D point it represents is computed according to the inverse equations of its projection (Equations 6, 7). This 3D point is then projected to each one of the six cameras according to the calibration. If the point is only within the FOV of one camera, we give to the pixel of the panorama the same intensity values as the pixel corresponding to the projection of the 3D point into the camera. In the case of overlapping regions, a blending criterion is needed, to establish the value of the panorama pixel. The blending criteria chosen for the panoramas displayed in this paper is known as gradient blending. Further details of this technique can be found in [16].

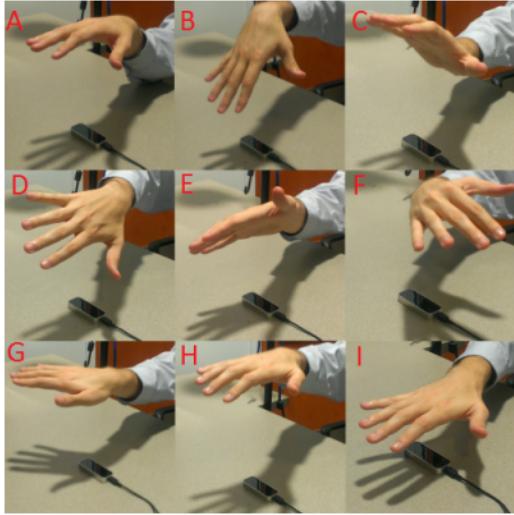


Figure 7: Gestures for moving the robot. A: Stay still. B: Move forward. C: Move backwards. D: Sway left. E: Sway right. F: Turn left. G: Turn right. H: Move upwards. I: Move downwards

5 TELEOPERATION

The Leap Motion Controller is able, out of the box, to track the position of both hands of the pilot when they are in the field of view of the cameras, and output their relative pose with respect to the frame of the device. A high-level controller has been written to translate the gestures of the pilot into commands that the robot can understand. The gestures have been chosen in so that they are both easy to perform and intuitive to inexperienced users. The main idea behind this is that during a scientific mission, scientists such as biologists or archaeologists could wear the VR headset and pilot the vehicle by themselves in order to examine with more detail regions with high interest, rather than having to indicate to a professional pilot the regions they want to explore in detail.

The set of gestures used is called "First person" and consists of using one hand as if it was the robot. As seen in Figure 7 tilting the hand slightly forward is sent to the robot as a move forward command, rotating the hand (with respect to an imaginary vertical axis) to the left is sent as a left rotation, etc. The only commands which require movement translations are the upwards and downwards movements that require the pilot to slightly move his hand in the desired direction.

6 RESULTS

In order to obtain preliminary results from testing the concept presented here, the setup was tested

separately in two sets of experiments. In the first set, the omnidirectional camera was integrated with the Girona500 AUV and tested during mapping missions. In the other set, the experiments were focused on tele-operating the Girona500 using hand-gesture techniques.

The omnidirectional camera has already been integrated with the Girona500 AUV (Fig. 2a) and tested both in a water tank and under real sea conditions. The equipment was tested repeatedly in mapping missions in the framework of MORPH EU-FP7 [17] project experiments that were carried out in Horta, Azores Islands during September 2015. The data was recorded onto the camera's disk memory for later post-processing. Once the data was retrieved, all the images captured in the same time instant were composed into an equi-rectangular panorama, as explained in section 4, before combining them into a 360° video (Fig. 8). The final video can be played in a spherical panorama viewer such as Kolor Eyes [18], for visualizing it in VR headsets such as the Oculus Rift. For making it available to a larger public, the video can be uploaded to platforms such as Youtube [19] or Facebook [20], which provide the spectators with spherical vision.

The users that tried the Oculus Rift headset were impressed by the immersive sensation achieved, and they all agreed that the capability to have online streaming of omnidirectional video from the robot could help both pilots and scientists to perform better during their missions. Nevertheless, testers said that the continuous use of the headset caused them a slight feeling of dizziness, and that the experience would be even better with a higher screen resolution. These are issues that with the rapid evolution of VR technologies are expected to be solved in the next release of the headset.

Besides this, a hand motion interpreter based on the Leap Motion Device was developed and integrated into the robot software architecture using ROS [22]. The tests consisted in deploying the Girona500 AUV in a water tank and performing simple trajectories, using the gesture-based interface with direct vision of the robot's movements. The pilot tried on several occasions to achieve a specific position in the water tank, departing from different locations in order to test the precision of the technique. People with different training levels tested the robot teleoperation with both a gamepad and the gesture-based technique proposed. All users reported that it was easy to learn and perform the gestures, and that the technique was useful even though for operations requiring high precision the use of a gamepad was recommended. It was also said that it was tiring to

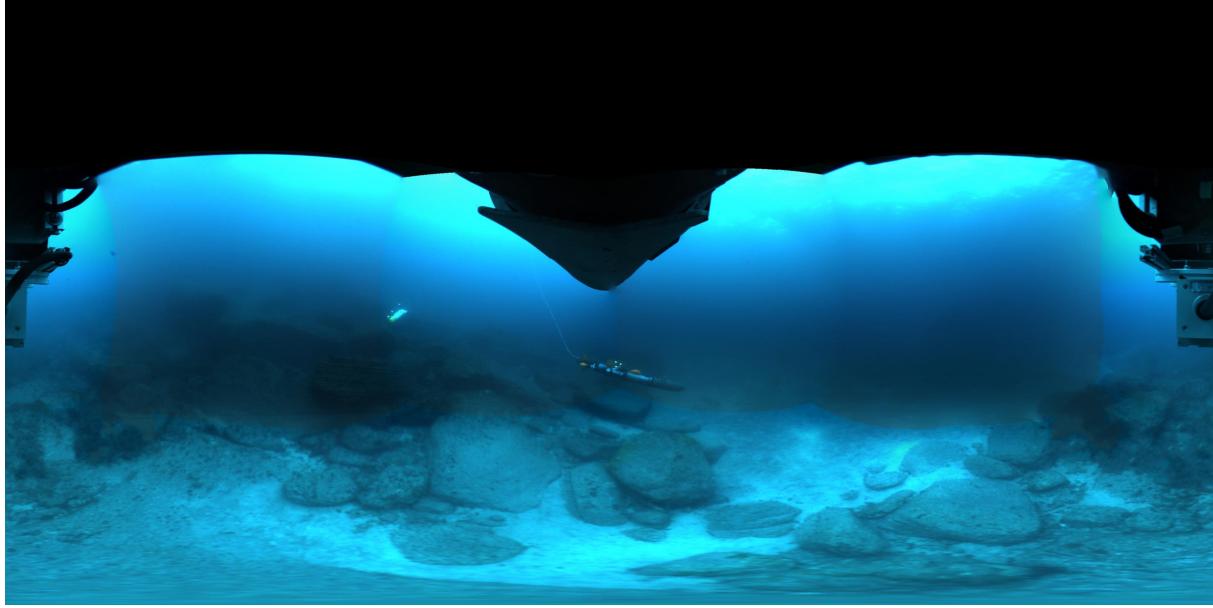


Figure 8: Equi-rectangular panorama captured during a mission in Horta, Azores Islands, September 2015, where the omnidirectional camera was used to track AUVs navigating at close range [21].

command the robot for a long period, as the hands are not used to be in a holding position for a long time.

7 CONCLUSIONS AND FUTURE WORK

The concept of immersive teleoperation presented in this paper shows great potential in terms of increasing the scientific capabilities in oceanographic expeditions, even though it is yet not fully implemented. It could help in achieving a better spatial awareness of the seafloor during explorations, and in saving both time and resources for pilots and scientists. The next step towards achieving a fully-working system will focus on developing a fiber-optics cable to allow higher speeds of data transmission between the ROV and the control station, and studying data-compression methods for the transmission of the images.

Acknowledgments

The authors wish to thank Ruben Daenen from University of Hasselt who worked on the Leap Motion during his stay in University of Girona. This research was supported by the Spanish National Projects ARCHROV (Marine ARChaeology through HROV/AUV cooperation) under the agreement DPI2014-57746-C3-3-R and OMNIUS under the agreement CTM2013-46718-R, the Generalitat de Catalunya through the ACCIO/TecnioSpring program (TECSPR14-1-0050) (to N. Gracias), and "la Secretaria d'Universitats i Recerca del Departament d'Economia i Coneix-

ement de la Generalitat de Catalunya" (to J. Bosch).

References

- [1] S. Negahdaripour, H. Zhang, P. Firoozfam, and J. Oles, "Utilizing panoramic views for visually guided tasks in underwater robotics applications," in *OCEANS, 2001. MTS/IEEE Conference and Exhibition*, vol. 4, 2001, pp. 2593–2600 vol.4.
- [2] A. Rituerto, L. Puig, and J. Guerrero, "Visual slam with an omnidirectional camera," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, Aug 2010, pp. 348–351.
- [3] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, "Google street view: Capturing the world at street level," *Computer*, no. 6, pp. 32–38, 2010.
- [4] J. Bosch, P. Ridao, D. Ribas, and N. Gracias, "Creating 360° underwater virtual tours using an omnidirectional camera integrated in an auv," in *OCEANS 2015 - Genova*, May 2015, pp. 1–7.
- [5] R. Chellali and K. Baizid, "What maps and what displays for remote situation awareness and rov localization?" in *Proc. of the Symposium on Human Interface 2011, Held as Part of HCI International 2011*, Orlando, FL, July 2011, pp. 364–372.

- [6] D. Pamungkas and K. Ward, "Immersive teleoperation of a robot arm using electro-tactile feedback," in *Automation, Robotics and Applications (ICARA), 2015 6th International Conference on*, Feb 2015, pp. 300–305.
- [7] H. Martins and R. Ventura, "Immersive 3-d teleoperation of a search and rescue robot using a head-mounted display," in *2009 IEEE Conference on Emerging Technologies Factory Automation*, Sept 2009, pp. 1–8.
- [8] C. Domingues, M. Essabbah, N. Cheaib, S. Otmane, and A. Dinis, "Human-Robot-Interfaces based on Mixed Reality for Underwater Robot Teleoperation," in *9th IFAC Conference on Manoeuvring and Control of Marine Craft (MCMC 2012)*, vol. 9, no. PART 1, Arenzano, Italy, Sep. 2012, pp. 212–215. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00761894>
- [9] J. C. Garcia, B. Patrao, J. Perez, J. Seabra, P. Menezes, J. Dias, and P. J. Sanz, "Towards an immersive and natural gesture controlled interface for intervention underwater robots," in *OCEANS 2015 - Genova*, 2015, pp. 1–5.
- [10] Y. Sanchez, R. Skupin, and T. Schierl, "Compressed domain video processing for tile based panoramic streaming using hevc," in *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept 2015, pp. 2244–2248.
- [11] D. Ribas, N. Palomeras, P. Ridao, M. Carreras, and A. Mallios, "Girona 500 AUV: From Survey to Intervention," *Mechatronics, IEEE/ASME Transactions on*, vol. 17, no. 1, pp. 46–53, Feb 2012.
- [12] I. Point Grey Research, "360 Spherical - Ladybug3 - Firewire Camera," <http://www.ptgrey.com/products/ladybug3/>.
- [13] J. Bosch, N. Gracias, P. Ridao, and D. Ribas, "Omnidirectional underwater camera design and calibration," *Sensors*, vol. 15, no. 3, pp. 6033–6065, 2015. [Online]. Available: <http://www.mdpi.com/1424-8220/15/3/6033>
- [14] Oculus VR, Inc., "Oculus Rift - Virtual Reality Headset for 3D Gaming," <http://www.oculusvr.com/>.
- [15] Leap Motion, Inc., "Leap Motion - Mac & PC Motion Controller for Games, Design, Virtual Reality & More," <http://www.leapmotion.com/>.
- [16] R. Prados, R. Garcia, N. Gracias, J. Escartin, and L. Neumann, "A novel blending technique for underwater giga-mosaicing," *IEEE Journal of Oceanic Engineering*, vol. 37, no. 4, pp. 626–644, 2012.
- [17] J. Kalwa, M. Carreiro-Silva, F. Tempera, J. Fontes, R. Santos, M.-C. Fabri, L. Brignone, P. Ridao, A. Birk, T. Glotzbach, M. Caccia, J. Alves, and A. Pascoal, "The MORPH concept and its application in marine research," in *OCEANS - Bergen, 2013 MTS/IEEE*, June 2013, pp. 1–8.
- [18] KOLOR, SARL, "Kolor - Kolor Eyes," <http://www.kolor.com/kolor-eyes/>.
- [19] CIRS UDG, "MORPH FP7-360° video of wall following," <https://www.youtube.com/watch?v=SCHDMbRr3WI>.
- [20] Facebook, Inc, "Facebook," <http://www.facebook.com>.
- [21] J. Bosch, N. Gracias, P. Ridao, K. Istenič, and D. Ribas, "Close-range tracking of underwater vehicles using light beacons," *Sensors*, vol. 16, no. 4, p. 429, 2016.
- [22] 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.