

Simultaneous Underwater Navigation and Mapping

Ana Rita Silva Gaspar¹, Aníbal Matos²

¹Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Porto, Portugal (up201402645@fe.up.pt) ORCID [0000-0002-8925-3447](https://orcid.org/0000-0002-8925-3447); ²Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Porto, Portugal (anibal@fe.up.pt) ORCID [0000-0002-9771-002X](https://orcid.org/0000-0002-9771-002X)

Abstract

The use of underwater autonomous vehicles has been growing, allowing the performance of tasks that cause inherent risks to Human, namely in inspection processes near to structures. With growth in usage of systems with autonomous navigation, visual acquisition methods have also gotten more developed because, they have appealing cost and they also show interesting results when operate at a short distance. It is possible to improve the quality of navigation through visual SLAM techniques which can map and locate simultaneously and its key aspect is the detection of revisited areas. These techniques are not usually applied to underwater scenarios and, therefore, its performance in environment is unknown. The paper presents a more reliable navigation system for underwater vehicles, resorting to some visual SLAM techniques from literature. The results, conducted in a realistic scenario, demonstrated the ability of the system to be applied to underwater environment.

Author Keywords. Loop-Closure, Navigation, SLAM, Underwater Environment, Vision.

Type: Research Article

 Open Access  Peer Reviewed  CC BY

1. Introduction

The growing use of underwater autonomous vehicles (AUV) is related with the fact of its actual features ([Wynn et al. 2014](#)) allow its application in tasks that may involve risks for Human, such as environment monitoring, inspection and demining. However, to ensure the use of an AUV in diverse applications, these vehicles must be able to navigate autonomously and the data must be obtained accurately ([Bosch et al. 2016](#)).

Since the underwater environment is unknown, we have an increase effort in exploitation and development of techniques to allow the navigation of underwater vehicles. In this context, Simultaneous Localization and Mapping (SLAM) techniques ([Paull et al. 2014](#)) has been arise to help in autonomous navigation namely in unstructured environments and when the initial information is poor. In this technique the robot constructs a coherent map of its environment while, at the same time, determines its location within that same map. However, this approach have to deal with some problems ([Aulinas et al. 2008](#)), namely the data association that occur when the number of possible hypotheses that identify the landmark grow.

The adversities of the underwater environment reduce the sensors to use. In this environment, the radio signal only propagates in short distances that makes the using of Global Positioning System (GPS) and techniques based in Wi-Fi communications less appealing. Although expensive and with some possible errors, the acoustic sensors present better performance in this environment. However, the autonomous vehicles are crucial in depth applications, such as monitoring and inspection of underwater structures. In this

context, the visual sensors have been an important focus to ensure that the vehicles can execute close-range missions autonomously and in real-time. Although the images could be affected by the poor visibility and light attenuation (Prados Gutiérrez 2013), have been applied some techniques to restore and to enhance the quality of imaging (Corchs and Schettini 2010; Pérez-Alcocer et al. 2016).

The SLAM approach with visual sensors (Pi et al. 2014) has the main goal to estimate the camera motion while reconstructs the environment. This technique assumes the extraction of features to estimate the position of the vehicle (Taketomi, Uchiyama, and Ikeda 2017) and for that presents two main methodologies based in:

- Filters – the motion is estimated with all frames, processing by a filter;
- Keyframes – the motion is estimated based some frames previously selected.

In addition, is common to use Bundle Adjustment (BA) to adjust the trajectory in certain moments, depending of the implementation. Moreover, to limit the estimation error, the vehicle must be able to recognize revisited areas - loop closure detection (Lowry et al. 2016). However, with visual systems is considered the greatest challenge since it implies the association of nonconsecutive images. As solution comes up the visual vocabulary concept, considered an efficient approach (Nister and Stewenius 2006), to determine the similarity between images, namely the Bag-of-Words (BoW) technique (Nicosevici and Garcia 2012; Law, Thome, and Cord 2014). The vocabulary can be created offline or online. In the first case, it is constructed a priori from a large set of training images. The online approach does not require human intervention: it is constructed according to robot motion. This approach stands out by accurately modeling the objects and scenes present in the surroundings and by being constructed as visual information becomes available.

This paper presents a robust, accurate and efficient visual system for simultaneous navigation and mapping in underwater environment. Thus, it is possible to contribute for a location of vehicles more reliable and safer and, therefore, to allow the use of AUV in long-term operations in an unknown environment.

Thus, this the paper is organized as follows: section 2 presents the steps and its description to obtain the developed system. Afterwards, the results obtained with the system in underwater environment are illustrate in section 3. Moreover, this section presents a comparative analysis between vocabulary and Bundle Adjustment approaches to detect revisited areas. Finally, in section 4 are discussed the major conclusions of this work.

2. Visual Navigation in Close-Range Scenarios

To develop the visual navigation system for underwater vehicles, the more appropriate SLAM technique of the literature was selected. After that, to recognize revisited areas and, so, to allow a more reliable motion estimation, a vocabulary method was developed. Figure 1 presents the modules that compose the developed system.

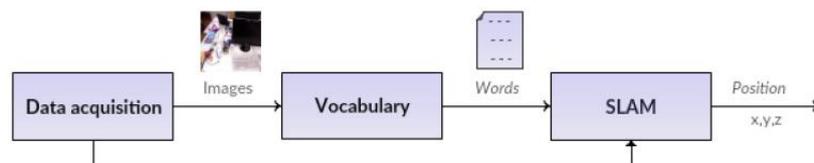


Figure 1: Overview of the steps for final system

A set of visual stereo SLAM techniques was analyzed based on the performance that is expected from the literature. These techniques are the following: RTAB-Map (Labbé and Michaud 2014), S-PTAM (Pire et al. 2015), and ORB-SLAM2 (Mur-Artal, Montiel, and Tardos 2015). The first implementation allows the use of external odometry methods, such as viso2,

to estimate 6DOF motion, but is not fit to pure rotations and it assumes that the camera is always at the same high. Moreover, it resorts to bucketing technique to ensure that the features extraction is uniform over entire image. The S-PTAM divides the SLAM problem into parallel tasks: tracking (detects and matches features to estimate the camera motion) and map optimization (removes the outliers). The ORB-SLAM2 uses the same features not only to map and tracking but also for local recognition, that allows that the system could be more efficient and reliable. These two methods use a keyframe-based approach for motion estimation, that avoids a computational growing. For loop-closure detection, the RTAB-MAP and ORB-SLAM2 use vocabulary approaches, that evaluate the similarity between the current frame and the others. The S-PTAM only performs an iterative bundle adjustment. The ORB-SLAM2 implementation uses an offline vocabulary and RTAB-Map constructs a vocabulary as the images become available.

These implementations were compared qualitatively and quantitatively to determine what of them is more fit to the intended context (real-time conditions). Thus, the quality of the motion estimation (Euclidean error), Central Processing Unit (CPU) and Random-Access Memory (RAM) utilization and the processing time were evaluated. For that some datasets online available were used, namely KITTI dataset (Geiger, Lenz, and Urtasun 2012) that depicts the outdoor environment (acquired by a car in urban areas and highway), see Figure 2. This dataset includes ground-truth of some trajectories and loop-closure situations.



Figure 2: Illustrative example of the KITTI dataset: (a) urban area and (b) highway (Gaspar et al. 2018)

Figure 2 illustrates the behavior obtained by the implementations to the KITTI dataset (sequence 05).

The ORB-SLAM2 is the only that estimates all camera positions, detects the loop-closures and, consequently, adjusts its trajectory. This implementation presents lower errors between estimated trajectory and ground-truth. The RTAB-Map, with viso2 to motion estimation, tries to replicate the motion but with some deviations, that can be explained by the susceptibility to sudden rotations of the camera. The S-PTAM is far from to achieve the intended trajectory, since frames at the beginning were lost. Thus, it is not possible to conclude about the loop-closure detection to relation these two last implementations.

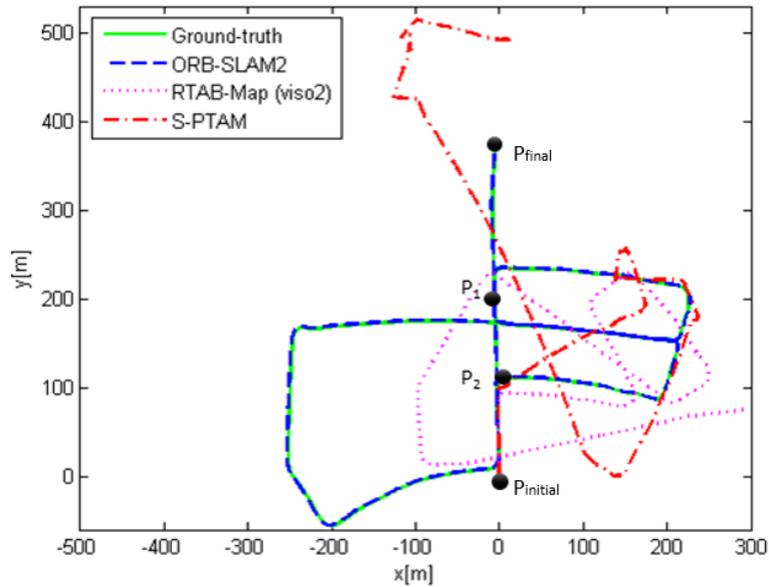


Figure 3: Trajectories obtained with KITTI dataset (sequence 05) in real-time conditions by visual SLAM techniques (Gaspar et al. 2018)

Based on the comparative analysis (Gaspar et al. 2018), with others sequences of KITTI and with others datasets online available (Stereo_20Hz and MIT Stata Center), was evident that the ORB-SLAM2 implementation was the best in the motion estimation and computational requirements – crucial in real-time operations. In terms of processing time, RTAB-Map and S-PTAM give better results. However, as they do not use all the frames for the motion estimation, the obtained values are insignificant. Moreover, these implementations have worst performance in Euclidean error and CPU. The S-PTAM only estimated correctly the motion in simple trajectories (without direction changes) and it is the approach that presents the highest computational requirements. Finally, in a qualitative way, the ORB-SLAM2 is the only that is suitable for the reliable recognition of revisited areas (loop-closure detection). Table 1 presents an overview of the performance obtained by the different methods.

	RTAB-Map	S-PTAM	ORB-SLAM2
Euclidean error	-	+	++
Processing time	-	+	++
CPU	+	-	++
RAM	+	++	-

Table 1: Performance obtained by the implementations related to some parameters. (++) represents the better result and (-) the worse

Moreover, the behavior in ideal conditions - offline processing - was also considered. As expected, when the ratio between the processed frames and the acquired frames is higher, the performance is better. Over these conditions, RTAB-Map and S-PTAM significantly improved their behavior by detecting existing loops, see Figure 4.

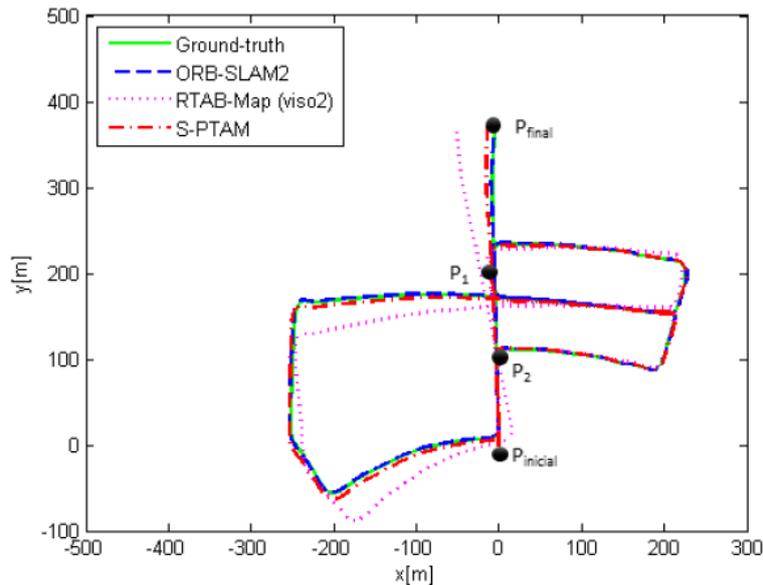


Figure 4: Trajectories obtained with KITTI dataset (sequence 05) in ideal conditions by visual SLAM techniques (Gaspar et al. 2018)

In general, the ORB-SLAM2 and S-PTAM implementations are the most complete. The S-PTAM emphasizes the minimization of the dependency between two threads and the use of binary features that decrease the computational requirements (increasing the detection process). The ORB-SLAM2 proves the efficiency and effectiveness of the quality of the loop detection. The possibility to adapt the vocabulary to the context and the performance of motion estimate in real-time, highlights the ORB-SLAM2 implementation. Thus, according to intended applicability this implementation was selected to underwater context.

To evaluate the selected SLAM technique and, consequently, the impact of the visual vocabulary in underwater environment, data were acquired by a visual system (composed by two cameras and two lasers) in a scenario that simulates this environment. Moreover, two anchors were placed as a control point and distance information points to evaluate (quantitatively) the obtained motion estimation were marked. Figure 5 depicts the data acquisition scenario.



Figure 5: Data acquisition system and scenario

Thus, some trajectories to evaluate the behavior of the ORB-SLAM2 in underwater environment were acquired. The trajectories present some direction changes and distance to the ground was kept, as possible. Loop-closure situations were included to assess the visual vocabulary method developed and, consequently, understand the impact of these approaches in the performance of the motion estimation.

3. Results and Discussion

Figure 6 illustrates a realized trajectory in the underwater scenario to verify the performance of the ORB-SLAM2 with the developed vocabulary method in this environment. The system starts in left side - point i - and it moves about 2 meters (passing by the anchors). Here, it performs a direction change (loop-closure situation) and, after, moves in diagonal direction with a direction change up to near to initial position - point f – but without loop situation in this point.

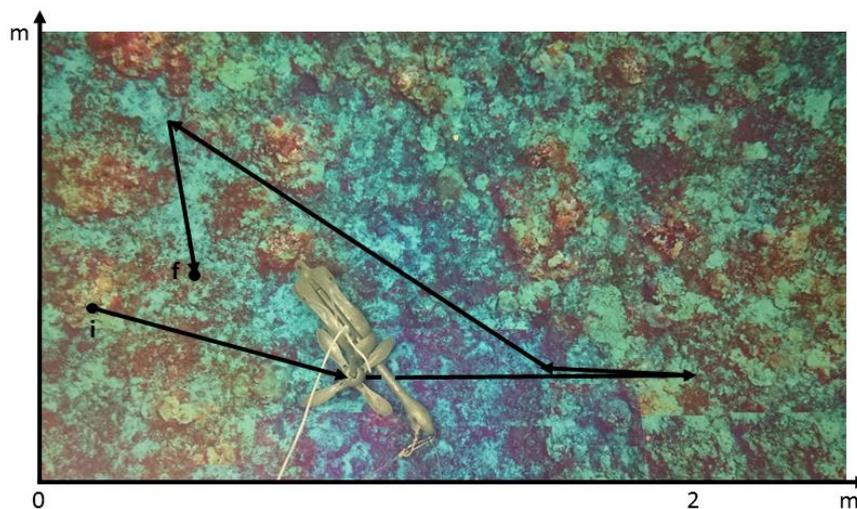


Figure 6: Illustrative example of the realized trajectory by visual system in underwater environment

In Figure 7 is visible a good motion estimation obtained by the ORB-SLAM2. However, this estimation does not include the intended initial point, because the initial requirements of this implementation, at this moment, were not satisfied. It is also noticeable that a direction change – point mid - near to final position was wrong. However, this aspect does not present relevance, because the trajectory ends close to initial point without loop detection, as expected.

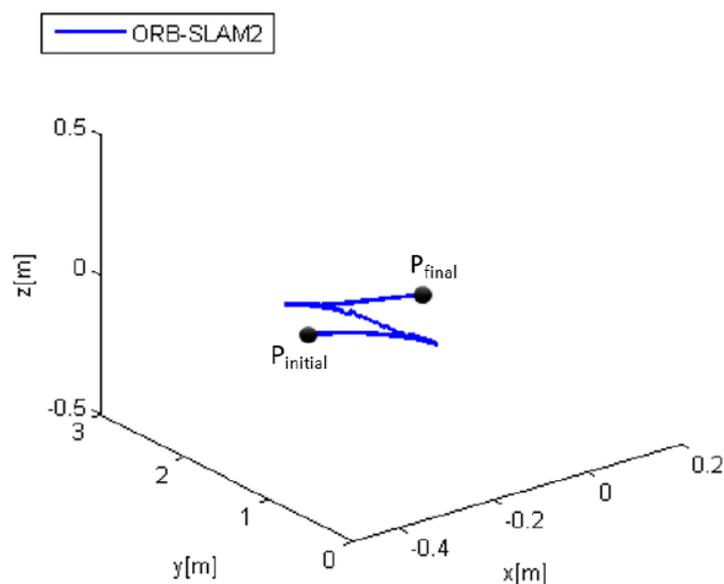


Figure 7: Underwater xyz trajectory obtained by ORB-SLAM2 method

The loop-closure situations present in this trajectory were detected (after the direction changes). In addition, the z-axis value is kept and the obtained scale was suitable as visible in Table 2.

	Real distance (m)	Estimated distance (m)
P_{initial} - P_{mid}	2.00	1.60
P_{initial} - P_{final}	0.50	0.40

Table 2: Quantitative performance obtained by the ORB-SLAM2 method in underwater environment

To analyze the influence of a vocabulary approach against the use of bundle adjustment, this experiment was conducted with S-PTAM implementation, see [Figure 8](#).

As visible, a low performance was obtained, since the motion estimation was not as desired and the revisited areas were not detected. Therefore, the obtained scale is far to real values and it was not intended that, in the ends, the x and y values been similar. Besides that, the final point is closer to the initial point than desired. Thus, to use only iteratively Bundle Adjustment is not able to detect revisited areas along to the camera motion.

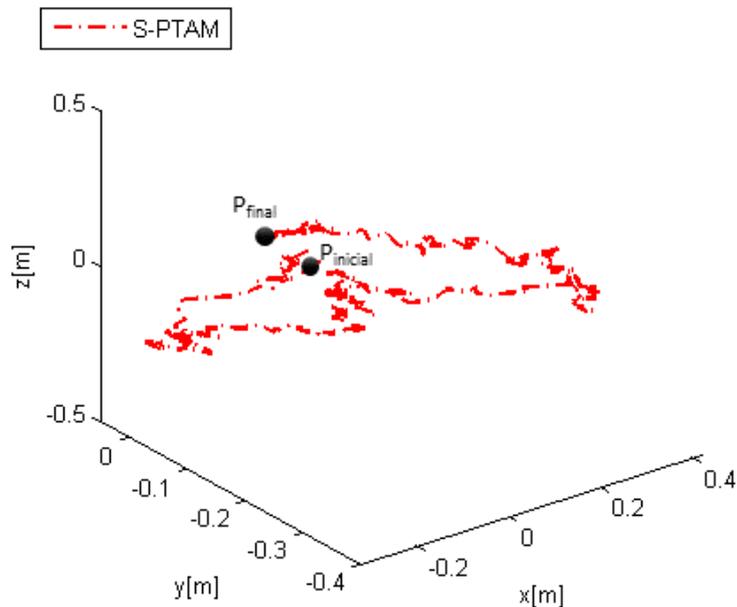


Figure 8: Underwater xyz trajectory obtained by S-PTAM method

4. Conclusions

Several visual SLAM methods that are typically applied to the navigation of mobile robots were analyzed in this paper. Results conducted that the ORB-SLAM2 is the more suitable technique. The comparative study of these methods was important to develop the final system, since that allow to achieve better results. The ORB-SLAM2 also proved its good performance in underwater scenario, since that could be following the motion with correct quantitative information. Presented low errors, namely 40cm between the initial and mid points and 10cm between initial and final positions. Moreover, it detected all loop-closure situations, adjusting its trajectory at these moments. This fact proved that the developed vocabulary method presented a good performance.

In addition, it was visible that the iteratively Bundle Adjustment is not enough to recognize revisited areas since that the S-PTAM does not detected any loop. Thus, the importance of visual vocabulary approach to increase navigation performance was proved. So, it is possible to use the presented system in a real environment, allowing operational flexibility.

In future, it is intended to increase the robustness of the system, increased the number of experiences in underwater environment and to improve the computational requirements of the ORB-SLAM2 method.

References

- Aulinas, J., Y. Petillot, J. Salvi, and X. Lladó. 2008. "The SLAM problem: A survey". In *Frontiers in Artificial Intelligence and Applications*, 363-71. Volume 184: Artificial Intelligence Research and Development. <https://doi.org/10.3233/978-1-58603-925-7-363>.
- Bosch, J., N. Gracias, P. Ridao, K. Istenič, and D. Ribas. 2016. "Close-range tracking of underwater vehicles using light beacons". *Sensors (Switzerland)* 16, no. 4 (march): 1-26. <https://doi.org/10.3390/s16040429>.
- Corchs, S., and R. Schettini. 2010. "Underwater image processing: State of the art of restoration and image enhancement methods". *Eurasip Journal on Advances in Signal Processing* 2010: Article number 746052. <https://doi.org/10.1155/2010/746052>.
- Gaspar, A. R., A. Nunes, A. M. Pinto, and A. Matos. 2018. "Urban@CRAS dataset: Benchmarking of visual odometry and SLAM techniques". *Robotics and Autonomous Systems* 109 (november): 59-67. <https://doi.org/10.1016/j.robot.2018.08.004>.
- Geiger, A., P. Lenz, and R. Urtasun. 2012. "Are we ready for autonomous driving? The KITTI vision benchmark suite". In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3354-61. <https://doi.org/10.1109/CVPR.2012.6248074>.
- Labbé, M., and F. Michaud. 2014. "Online global loop closure detection for large-scale multi-session graph-based SLAM". In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2661-66. <https://doi.org/10.1109/IROS.2014.6942926>.
- Law, Marc T., Nicolas Thome, and Matthieu Cord. 2014. "Bag-of-words image representation: Key ideas and further insight". In *Fusion in Computer Vision: Understanding Complex Visual Content*, 29-52. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05696-8_2.
- Lowry, S., N. Sunderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford. 2016. "Visual place recognition: A survey". *IEEE Transactions on Robotics* 32, no. 1 (february): 1-19. <https://doi.org/10.1109/TRO.2015.2496823>.
- Mur-Artal, R., J. M. M. Montiel, and J. D. Tardos. 2015. "ORB-SLAM: A versatile and accurate monocular SLAM System". *IEEE Transactions on Robotics* 31, no. 5 (october): 1147-63. <https://doi.org/10.1109/TRO.2015.2463671>.
- Nicosevici, T., and R. Garcia. 2012. "Automatic visual bag-of-words for online robot navigation and mapping". *IEEE Transactions on Robotics* 28, no. 4 (august): 886-98. <https://doi.org/10.1109/TRO.2012.2192013>.
- Nister, D., and H. Stewenius. 2006. "Scalable recognition with a vocabulary tree". In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2161-68. <https://doi.org/10.1109/CVPR.2006.264>.
- Paull, L., S. Saeedi, M. Seto, and H. Li. 2014. "AUV navigation and localization: A review". *IEEE Journal of Oceanic Engineering* 39, no. 1 (january): 131-49. <https://doi.org/10.1109/JOE.2013.2278891>.
- Pérez-Alcocer, R., L. A. Torres-Méndez, E. Olguín-Díaz, and A. A. Maldonado-Ramírez. 2016. "Vision-based autonomous underwater vehicle navigation in poor visibility conditions using a model-free robust control". *Journal of Sensors* 2016: Article number 8594096. <https://doi.org/10.1155/2016/8594096>.
- Pi, S., B. He, S. Zhang, R. Nian, Y. Shen, and T. Yan. 2014. "Stereo visual SLAM system in underwater environment". In *OCEANS 2014 - TAIPEI*. <https://doi.org/10.1109/OCEANS-TAIPEI.2014.6964369>.

- Pire, T., T. Fischer, J. Civera, P. De Cristoforis, and J. J. Berlles. 2015. "Stereo parallel tracking and mapping for robot localization". In *IEEE International Conference on Intelligent Robots and Systems*, 1373-78. <https://doi.org/10.1109/IROS.2015.7353546>.
- Prados Gutiérrez, Ricard. 2013. "Image blending techniques and their application in underwater mosaicing". PhD diss., Departament d'Arquitectura i Tecnologia de Computadors, Universitat de Girona. <http://hdl.handle.net/10803/111333>.
- Taketomi, T., H. Uchiyama, and S. Ikeda. 2017. "Visual SLAM algorithms: a survey from 2010 to 2016". *IPSI Transactions on Computer Vision and Applications* 9, no. 1 (june): 16. <https://doi.org/10.1186/s41074-017-0027-2>.
- Wynn, R. B., V. A. I. Huvenne, T. P. Le Bas, B. J. Murton, D. P. Connelly, B. J. Bett, H. A. Ruhl, et al. 2014. "Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience". *Marine Geology* 352 (june): 451-68. <https://doi.org/10.1016/j.margeo.2014.03.012>.

Acknowledgments

The authors would like to thank FEUP and INESC TEC for the conditions created for the development of the present work.