

Three-Dimensional Mapping in Underwater Environment

Alexandra Nunes¹, Aníbal Matos²




¹Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Porto, Portugal (up201402644@fe.up.pt) ORCID [0000-0002-4142-3144](https://orcid.org/0000-0002-4142-3144); 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

Autonomous underwater vehicles are applied in diverse fields, namely in tasks that are risky for human beings to perform, as optical inspection for the purpose of structures quality control. Optical sensors are more appealing cost and they supply a larger quantity of data. Lasers can be used to reconstruct structures in three dimensions, along with cameras, which create a faithful representation of the environment. However, in this context a visual approach was used and the paper presents a method that can put together the three-dimensional information that has been harvested over time, combining also RGB information for surface reconstruction. The map construction follows the motion estimated by a odometry method previously selected from the literature. Experiments conducted using real scenario show that the proposed solution is able to provide a reliable map for objects and even the seafloor.

Author Keywords. Object Reconstruction, Odometry, Underwater, Visual Mapping.

Type: Research Article

 Open Access  Peer Reviewed  CC BY

1. Introduction

Nowadays, the acquisition and analysis of ocean information are crucial steps to biological research and, simultaneously, to support different maritime industry applications. In several underwater tasks, namely inspection of structures and tracking moving objects ([Galceran and Carreras 2013](#); [Srividya and Shobha 2014](#)), the use of visual information has been investigated. In addition, it is important in close range inspection, when the proximity to submerged structures is required ([Kocak et al. 2008](#)).

The image acquisition is a quite complex task due, for example, the attenuation existent in this environment and the shadows created from the artificial illumination, that can induce false motion ([Prados Gutiérrez 2013](#)). However, actually some methods have been developed to improve the image quality after acquisition ([Corchs and Schettini 2010](#)). Otherwise, for the close-range inspection it is important that the acquisition of images occurs as close as possible to the ground that reduces the field of view, that is a challenge in terms of computational processing and memory management since implies storage of more data.

In this type of applications, the vehicles have to deal with unknown and uncontrolled environments that make it important to construct three-dimensional maps with texture information, not only on the ocean floor but also on structures that are submerged ([Campos et al. 2015](#)). For the mapping process it is necessary the correct motion estimation, where visual odometry techniques can be used throughout input images from one or multiple cameras ([Scaramuzza and Fraundorfer 2011](#); [Fraundorfer and Scaramuzza 2012](#)). These methods present some advantages in relation to the traditional approaches (with IMUs and laser, for example) namely in situations where the GPS signal is not enough (like underwater

environments). However, this process has to deal with different problems that can cause errors, namely with parallax problems caused by the approximation to the ground or the structure when the data are acquired. There are three common crucial techniques for obtaining the data to build the 3D maps:

- Triangulation;
- Time-of-flight sensors;
- Structured light source.

The last two are important due to their high precision and resolution levels to map structures of any size and shape. But the use of cameras has become attractive for this reconstruction to be possible with high resolution and low cost. The map it is possible to use to identify some objects inside and, therefore, this field is in study to achieve better results in object identification and their reconstruction with high quality (Massot-Campos and Oliver-Codina 2015; Chen et al. 2017).

Taking into account that the total perception of the ocean seafloor is still reduced, this paper presents the development of a method that improves and allows a 3D underwater mapping from fusion of the images with 3D information. Thus, the main contributions of this work are the selection of the most appropriate techniques from the literature of odometry for underwater applications and the development of the method able to map with texture information an area of the seafloor. Experiments conducted using real scenario show that the proposed solution is able to provide a reliable map only used visual sensors.

Thus, the paper is divided in 4 sections. Section 2 describes the approaches used to developed the final method. Next, section 3 presents some results and analysis the quality of them to mapping areas and objects reconstruction. Finally, the conclusions are presented in section 4.

2. Visual Mapping with Texture Information

The presented method has the main goal to allow the accurate mapping of the structures or the seafloor with images and 3D information coming to the stereo visual system. In this way, it is purely visual approach and the egomotion must be estimated in the first step. Thus, Figure 1 presents an overview of the system, which is composed by three independent modules.

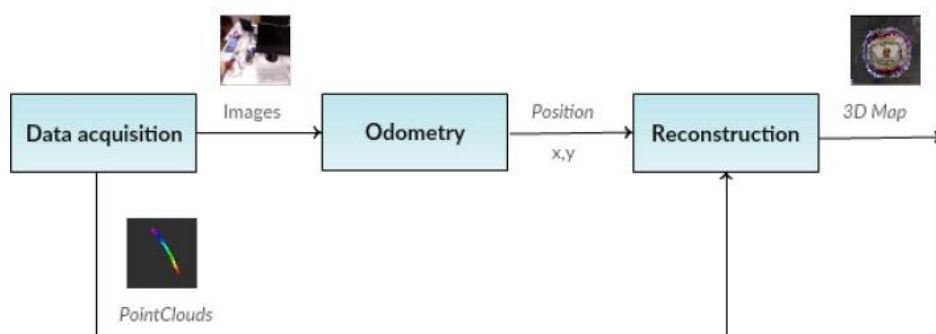


Figure 1: Block diagram of the system

Simultaneously with the building of the map it is necessary to estimate the motion of the camera. Thus, the quality of the map is dependent of the performance of the motion estimation and so, to achieve an appealing map in the final result is necessary to use an appropriate method of the egomotion estimation with the lower errors, as possible.

So, in a first phase a comparative analysis it was done between different methods of the literature, namely mono-vo (Singh 2015), viso2 (Geiger, Ziegler, and Stiller 2011), ORB-SLAM (Mur-Artal, Montiel, and Tardos 2015), to select the most appropriate monocular odometry

method. The first one although implements an outlier's removal mechanism is a simple implementation. Thus, this method discards the features that have erroneous correspondence between two consecutive frames. The scale factor is obtained by an external source to be possible to correct the estimations. The viso2 implementation has a high number of configurable parameters and uses a bucketing technique to allow that features are distributed uniformly into image. This implementation also presents an outlier's mechanism. In the final, the ORB-SLAM uses keyframes to egomotion estimation and although it is a SLAM implementation only the motion estimations were verified, without any revisiting areas or regions. This method resorts to ORB descriptor (binary features) that decreases the time of features extraction and allows the real-time operations.

This type of analysis is crucial to verify the advantages and disadvantages of each implementations in similar test conditions. The methods were compared taking into account some important features for typical robotics applications such as: the Euclidean error in motion estimation (that is normalized because the scale factor of each implementation is calculated in the different way), the Central Processing Unit (CPU) utilization and the processing time. For this analysis different available online scenarios were chosen (indoor and outdoor), namely the KITTI dataset (Geiger, Lenz, and Urtasun 2012), that are commonly used to develop and test visual processing algorithms, see Figure 2. This dataset presents ground truth, calibration parameters and different trajectories with different dimensions. In the sequences that present loops these are excluded because the main goal of this comparison is only estimating the motion.



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

Figure 3 presents an example of this comparative study with sequence 07 (Gaspar et al. 2018). It is possible to see that the mono-vo follows correctly the motion, but an incorrect direction change causes a deviation between the real and the estimated trajectory. This can be explained because their heuristic that only estimates the motion that are majority "forward". The ORB-SLAM estimated correctly all direction changes and presented lower errors in the egomotion estimation. However, this implementation presents higher processing time. The viso2 presented also a good estimation only with some deviations in the initial part.

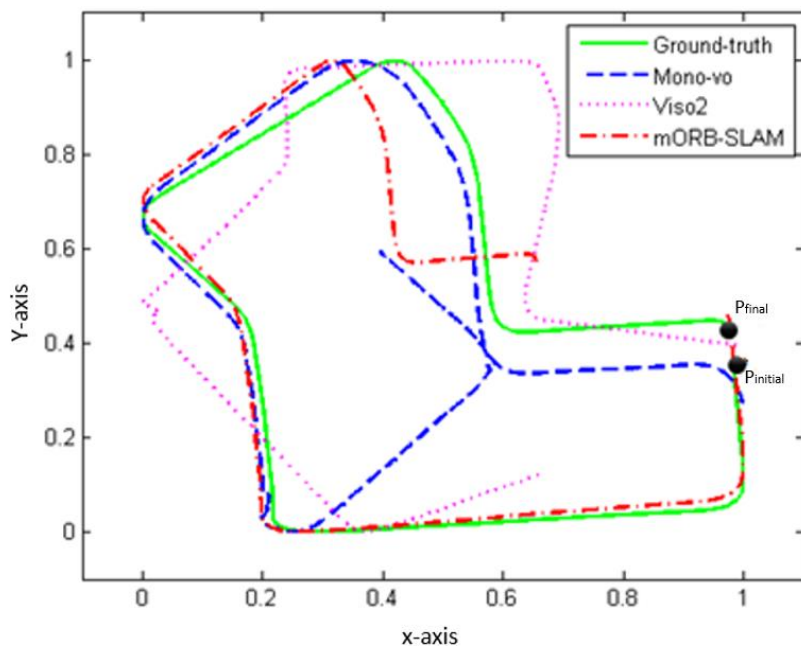


Figure 3: Normalized trajectories obtained with KITTI dataset (Gaspar et al. 2018)

Table 1 summarizes the obtained results in terms of evaluated features considering other datasets (Gaspar et al. 2018).

	Mono-vo	Viso2	ORB-SLAM
Euclidean error	Low	Moderate	High
Processing time	Moderate	High	Low
CPU	Low	High	Moderate

Table 1: Performance of the implementations with the evaluated parameters

So, it was verified that the ORB-SLAM implementation was the best in the motion estimation. However, this implementation uses an offline vocabulary provided by the authors who makes it difficult to adapt to different scenarios, namely in the underwater environments such as intended. It is also important to stress that frames were only discarded when location was lost or the initial moment could not be initialized. In terms of processing time and computational requirements the viso2 implementation was the best. This presents good estimation results during the study except situations where the height of the camera changes in relation to the ground or in the presence of abrupt and pure rotations. Although the mono-vo tries to replicate the motion of the camera, it has a heuristic that becomes a limitation when there are more complex trajectories. Any external motion is detected as a motion of the camera (even when the camera is not moving) which causes diverse incorrect direction changes.

From the analysis done, it was clear that the ORB-SLAM and viso2 implementations are, in general, the most appropriate. However, it is important to refer that the viso2 implementation is quite configurable. In the intended context there is no suitable vocabulary for the desired scenario, so the viso2 implementation stands out, opting for its use.

Afterwards, the final map is obtained through the 3D data (pointcloud) by the stereo system, considering a preliminary calibration step. These 3D data are processed over the time considering the egomotion estimated with the implementation selected. Thus, before the accumulation points it is necessary that each pointcloud changes its orientation and position into relation with the previous. In large scale scenarios the final result can have large dimensions and, in these cases, it is relevant to take into account the hierarchical approaches. Thus, in this work octrees were used that are able to organize the 3D information efficiently, reduce the cost of the allocation and deallocation memory steps for cases where it is

necessary that the construct rate of the octree to be high. The use of this structures presents as advantages search procedures more efficient and automatic adjust in relation to final pointcloud size.

To test and to improve developed method it was used a visual system with two Mako G-125 cameras. The data are collected and rectified with parameters provided by a preliminary calibration step, that is appropriate for stereo systems. The presented scenario simulates the real environment in controlled conditions (pool for experiments), see [Figure 4](#). This scenario presents about 5.5x4.5 meters and 1.7 meters of depth. To verify the quality of the final map, with developed method, two anchors were placed on the ground.

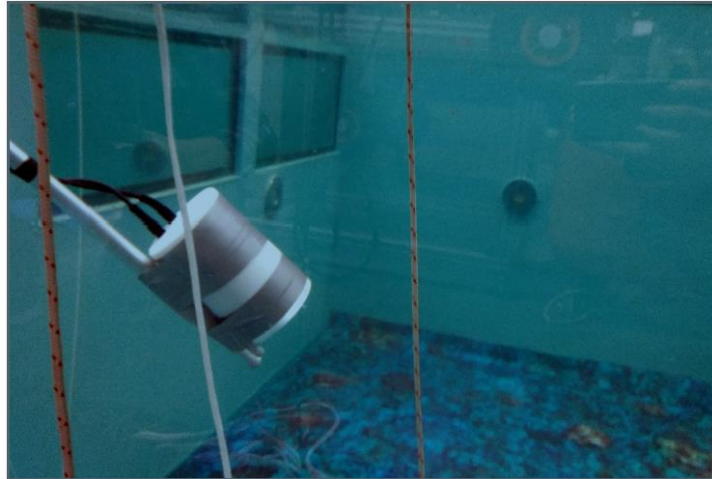


Figure 4: Data acquisition system

3. Results and Discussion

Before testing the final map, it was possible to verify the capability of the octrees in the memory management and the advantages in the use of them. It was perceptible that the accumulation time of different pointclouds is very lower than when we use traditional methods (to accumulate five million of points it was necessary only 12 seconds in contrast to 90 seconds for the traditional methods – sum operator). It is also important to refer that the time used in each iteration is the same along the entire process (never more than 0.07 seconds in each operation), that does not occur in the other methods where this time increases on each iteration. Thus, the efficiency of the use of octrees as structures for storage and accumulation of pointclouds was accomplished, avoiding memory management problems unnecessarily.

To evaluate and verify the quality of the underwater map creation with the odometry method chosen previously a trajectory was performed around the object. The viso2 implementation estimated correctly the trajectory with some small deviations, can be caused by the oscillations on the acquisition process or due changes in terms of the distance to the ground (susceptibilities of the implementation already verified in the literature).

[Figure 5](#) demonstrates that the final map was well constructed, following the motion estimated by the odometry method and with the pointclouds in the same plane.

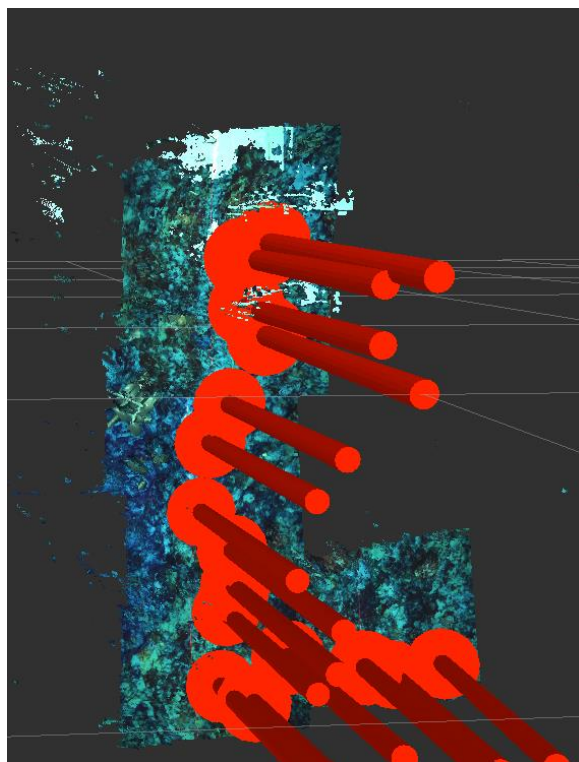


Figure 5: Map construction with odometry information

With this method it is also possible to see and to recognize the objects that is intended to map and, therefore the anchors on the ground are a good tool to verify the map quality. [Figure 6](#) demonstrates that the object is perceptible and in this way the reconstruction of small objects is also possible with this method.

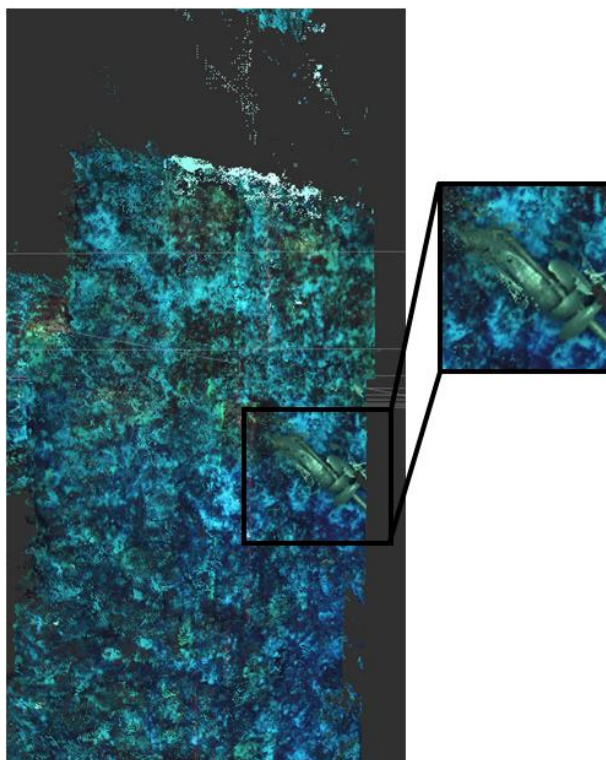


Figure 6: Map with the object well reconstructed

In [Figure 7](#) it is possible to observe the complete map of the trajectory as well as to capture the underwater seafloor and the two anchors placed on the ground.

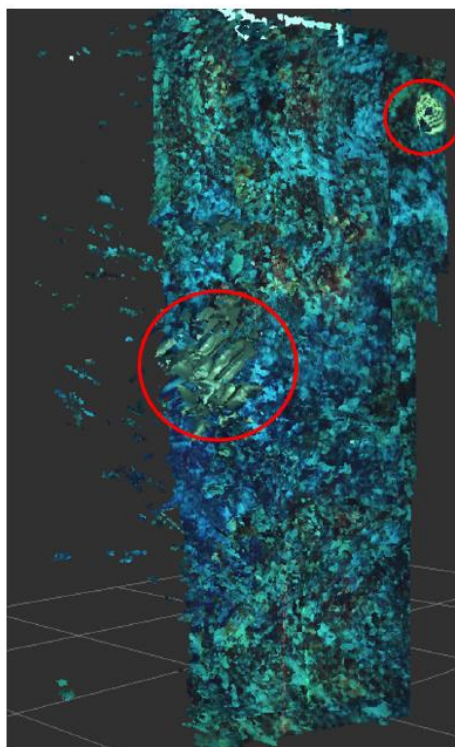


Figure 7: Complete map with the objects and structures well reconstructed

Thus, it has been proved that if the motion estimation is correct, the final map presents coherence and data quality in the mapping of generic areas. However, some errors can occur when the motion is incorrectly estimated, namely: repetition of regions or objects and the incorrect orientation of the pointcloud before the accumulation on the map even if the error occur only one time.

Although the quality of each pointcloud at the time of acquisition is high it is possible to see the existence of small points out of the usable area that can be considered noise.

When it is necessary accumulate these structures the quantity of noise will increase and, consequently, the final map going to be unpleasant in relation to the intended. In that case, the application of a filter voxel grid in the z axis solves the problem and builds a more visually pleasant map, see [Figure 8](#).

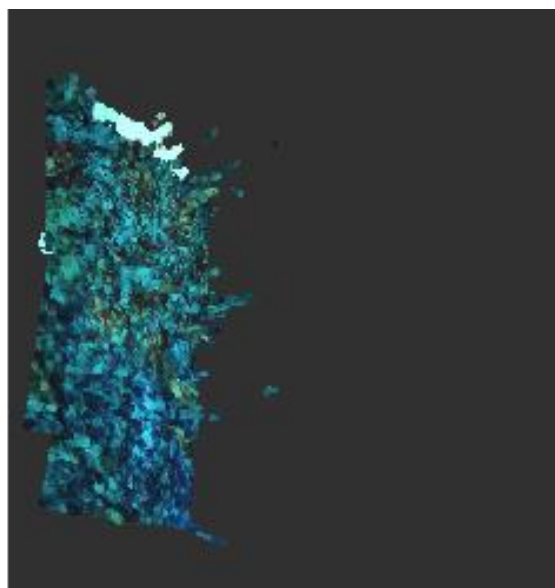


Figure 8: Final map with voxel grid filter

4. Conclusions

In the first step a comparative study was realized that allow to test some odometry techniques of the literature. The results demonstrated that viso2 is the most appropriate and, thus, was possible to obtain better results in the accumulation of the pointclouds to construct the final map. This step demonstrated that is the most important for a good map quality (an error in motion estimation causes an inappropriate accumulation).

The use of octrees proved to be the most suitable way to store data and manage memory. In the underwater context the developed system proved to be able to reconstruct the environment and the final map presents enough quality to recognize the objects as intended. Thus, it is possible to use the system in real environment, namely in inspection and monitoring operations.

In terms of the future work the number of experiments in the underwater environments with other objects and conditions should be increased to improve the robustness of the developed method. In addition, it is also important to develop a filtering method that can be applied in the odometry data to avoid some consecutive errors.

References

- Campos, R., R. Garcia, P. Alliez, and M. Yvinec. 2015. "A surface reconstruction method for in-detail underwater 3D optical mapping". *The International Journal of Robotics Research* 34, no. 1 (october): 64-89. <https://doi.org/10.1177/0278364914544531>.
- Chen, Z., Z. Zhang, F. Dai, Y. Bu e H. Wang. 2017. "Monocular vision-based underwater object detection". *Sensors (Switzerland)* 17, no. 8 (august): Article number 1784. <https://doi.org/10.3390/s17081784>.
- 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>.
- Fraundorfer, F., and D. Scaramuzza. 2012. "Visual odometry: Part II: Matching, robustness, optimization, and applications". *IEEE Robotics and Automation Magazine* 19, no. 2: 78-90. <https://doi.org/10.1109/MRA.2012.2182810>.
- Galceran, Enric, and Marc Carreras. 2013. "Automating seafloor inspection using autonomous underwater vehicles". Paper presented in Martech 2013 5th International Workshop on Marine Technology, Girona. <http://hdl.handle.net/2099/13918>.
- 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., J. Ziegler, and C. Stiller. 2011. "StereoScan: Dense 3d reconstruction in real-time". In *IEEE Intelligent Vehicles Symposium, Proceedings*, 963-68. <https://doi.org/10.1109/IVS.2011.5940405>.
- 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>.
- Kocak, D. M., F. R. Dalglish, F. M. Caimi, and Y. Y. Schechner. 2008. "A focus on recent developments and trends in underwater imaging". *Marine Technology Society Journal* 42, no. 1 (march): 52-67. <https://doi.org/10.4031/002533208786861209>.

- Massot-Campos, M., and G. Oliver-Codina. 2015. "Optical sensors and methods for underwater 3D reconstruction". *Sensors (Switzerland)* 15, no. 12 (december): 31525-57. <https://doi.org/10.3390/s151229864>.
- 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>.
- 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>.
- Scaramuzza, D., and F. Fraundorfer. 2011. "Visual odometry: Part I: The first 30 years and fundamentals [Tutorial]". *IEEE Robotics & Automation Magazine* 18, no. 4 (december): 80-92. <https://doi.org/10.1109/MRA.2011.943233>.
- Singh, Avi. 2015. "Monocular visual odometry". Undergraduate report, Indian Institute of Technology, Kanpur, India. <http://avisingh599.github.io/assets/ugp2-report.pdf>.
- Srividya, M. S., and G. Shobha. 2014. "A survey on: Underwater video processing for detecting and tracking moving objects" *International Journal of Soft Computing and Engineering* 4, no. 1 (march): 15–17. <https://www.ijscce.org/download/volume-4-issue-1/>.

Acknowledgments

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