

The Impact of Convergence Cameras in a Stereoscopic System for AUVs

João Aguiar, Andry Maykol Pinto, Nuno A. Cruz, and Anibal C. Matos

INESTEC and Faculdade de Engenharia da Universidade do Porto,
Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal
`ee10018@fe.up.pt`, `andry.pinto@fe.up.pt`, `nacruz@fe.up.pt`, `anibal@fe.up.pt`

Abstract. Underwater imaging is being increasingly helpful for the autonomous robots to reconstruct and map the marine environments which is fundamental for searching for pipelines or wreckages in depth waters. In this context, the accuracy of the information obtained from the environment is of extremely importance. This work presents a study about the accuracy of a reconfigurable stereo vision system while determining a dense disparity estimation for underwater imaging. The idea is to explore the advantage of this kind of system for underwater autonomous vehicles (AUV) since varying parameters like the baseline and the pose of the cameras make possible to extract accurate 3D information at different distances between the AUV and the scene. Therefore, the impact of these parameters is analyzed using a metric error of the point cloud acquired by a stereoscopic system. Furthermore, results obtained directly from an underwater environment proved that a reconfigurable stereo system can have some advantages for autonomous vehicles since, in some trials, the error was reduced by 0.05m for distances between 1.125 and 2.675m.

Keywords: Underwater perception; Stereoscopic; Convergence cameras; Autonomous underwater vehicle

1 Introduction

Several researchers are developing novel vision systems for perceiving the aquatic environment using the UAVs (autonomous underwater vehicles). The idea beyond this effort is to enhance the ability of these vehicles to understand and to navigate in harsh environments, especially, for operating close to man-made structures and the sea-floor. Nonetheless, the perceptual information is not collected as easily as expected by underwater robotic applications due to severe conditions related with the light propagation in deep waters. These conditions affect the accuracy and the detail with the environment is perceived which is quite relevant for mapping underwater structures, obstacle avoidance and precise location for docking.

Nowadays, there are technological solutions already available for robotic applications such as the Sound Navigation and Ranging (SONAR), that calculates the distance using sound waves. Although the SONAR evidencing significant

advantages for greater distances, the multiple reflections of the sound waves often result in a noisy sensor data for smaller distances. Therefore, vision-based systems enhance the capabilities of AUVs by complementing the sensor information available from the SONAR. In this context, the stereoscopic systems are suitable vision-based sensors since they are a cost-effective solutions and provide rich information of texture; however, their effectiveness depends on the visibility of the scene and is affected by suspensoids, backscattering and light absorption. Conventional stereoscopic systems for underwater environments often resort to small baselines and to a pre-defined camera poses (usually, parallel configurations) which may not be the most flexible approach for measuring 3D information for an autonomous vehicle because the objects of interest could have different relative distances to the UAV which influences the accuracy of the sensor data that is possible to be retrieved.

This paper introduces the concept of a Reconfigurable stereoscopic system based on SELF-adjusting cameras (RSECam) that converge or diverge according to the average distance to the scene. Besides this novel concept, the research provides also a preliminary study about the effect of the baseline and the pose of cameras in the performance of the system, by analyzing the accuracy of the point cloud that is obtained on underwater scenarios.

A set of experiments are conducted as a proof-of-concept for the RSECam. These validations provide a practical and quantitative analysis of the perceptual system in a real water tank, with several distinct objects at different ranges and including two different baselines (11cm and 29cm).

Therefore, contributions of this article include:

- A novel concept of a Reconfigurable stereoscopic system based on SELF-adjusting cameras (RSECam);
- A study about the impact of convergence cameras, in metric units, for the proposed underwater stereoscopic system;
- Qualitative evaluations for a proof-of-concept of the RSECam: have mainly considered realistic scenarios.

The article is organized as follows: section 2 presents a brief review of systems that perceive distance information from the underwater environment. Section 3 shows the concept of the RSECam technique. Afterwards, section 4 presents more practical results for the proof-of-concept. The experiments showed that reconfigurable stereoscopic configurations based on convergence cameras increase, in some cases, the accuracy of the 3D information when compared to conventional stereoscopic systems (with parallel or pre-determined configurations).

Finally, section 5 presents the most important conclusions of this research.

2 Related works

Intelligent and autonomous robotic vehicles have been developed with advanced visual systems [10] and they are capable to perform tasks that would otherwise be too dangerous, expensive or even impossible to achieve. The most often used

sensors for underwater applications are the acoustic sensors due to the advantages of the sound propagation in water. Acoustic sensors are used in several applications like, for example, obstacle detection [4]. The work [12] presents an experimentally study of capabilities and limitations of underwater localization methods based on a robotic fish equipped with small, low-power sounder (buzzer) and microphones. In this experience, the authors achieved an underwater localization resolution of 20cm over a range of 10m. Other sensors used in range measurement are the LiDAR (Light Detection And Ranging) sensors despite of the light attenuation and absorption problems when submerged [3]. They are very useful for small and accurate distances and several investigators are already trying to decrease the undesirable backscattering problem [11]. LiDAR sensors can be separated in two different groups: the ones that work with triangulation and those that use ToF (time of flight). The former have higher resolution (less than 1mm) than ToF but only for short ranges (less than 1m). The latter are better for distances greater than 2.5m and they have a 5mm range precision for 8m distances. In the case of [6] and considering a range of 10m, the prototype based in ToF technology achieved a precision of 30mm (Jerlov Type III). They also presented some 3D reconstructions on-the-fly. The LiDAR-based techniques are discussed with detail in [7]. Another possible approach for distance measurement is the visual triangulation which can be grouped in active or passive techniques. The first needs a light source like a laser [8] or even structured light [5]. Two cameras are used at least for the second approach in order to determine the 3D information from the scene. The work presented in [1] concludes that structured light can be used in underwater environments and authors reported good results for a 3D reconstruction in low turbidity waters. However the solution presented in the research requires a projector and a structure capable of protecting both the camera and the projector itself when submerging the rig in water. The presence of a projector in this kind of system makes the range measurement easier due to the fact that it normally projects a well-defined pattern however, the power required by the projector device is substantial which reduces the autonomy of the robotic vehicle. Passive techniques require a minimum number of cameras of two to successfully measure the distance of a certain object within the scene and through a stereo correspondence - features captured on the left camera are found in the right camera and then, the disparity (difference between right and left pixels) can be determined. The biggest disadvantage of this technique is related to underwater imaging. Some image preprocessing algorithms are being created to mitigate some of the issues, for instance, the backscattering problem [13]. However, these methods usually affect the distance that is measured to the objects, for instance, the research [13] presented an average error of 28cm for an object at 246cm due to the presence of scattering. Two different cameras were used in [12] for underwater ranging measurement. This solution presented good results since the average error was about 10% for distances under 5m. Finally, the research presented in [13] demonstrates a similar comparison and the influence of the baseline distance is also studied. As can be noticed, the range error increases with the distance of the cameras to the object and, in this particular case, the

results were obtained with 50cm of baseline and considering distances greater than 10 meters.

Currently, the limited visual perception capability of AUVs restricts the use these vehicles in purely autonomous operations with a medium complexity. Therefore, it becomes crucial the development of vision-based methods for understanding the sea-floor to allow a more efficient use of such robotic systems in real environments [9].

3 The Reconfigurable stereoscopic system based on SELF-adjusting cameras (RSECam)

This research presents the concept of a Reconfigurable stereoscopic system based on SELF-adjusting cameras (RSECam). In its essence, the RSECam adjusts the pose of the cameras to rearrange the configuration of the stereo rig, for achieving a more accurate 3D measures when compared to conventional solutions.

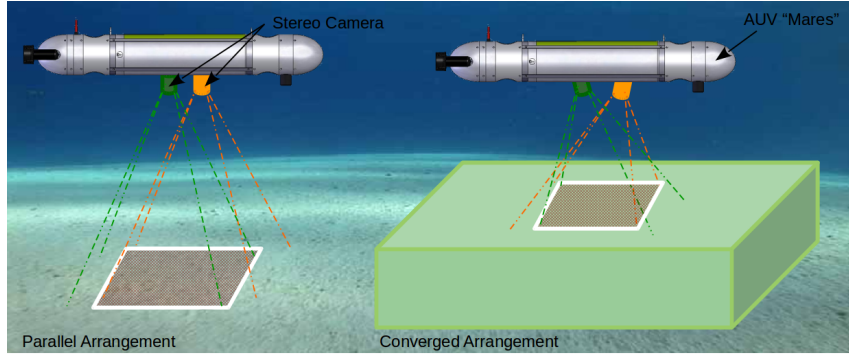


Fig. 1. The concept of RSECam.

The system should adjust the configuration of its camera according to the distance between the vision sensor to the scene. This changes the region where the two cameras' fields of view converge or intersect, often called by stereo window. Thus, the stereo window is modified by converging or diverging the cameras based on the rule that nearing objects lead to higher convergence angle values. The convergence is the angle formed by the cameras¹ and it is sometimes called by "toe-in". Moreover, the convergence point determines where the object appears in relation to the stereo window. Figure 1 presents the concept and the example of a practical application of the stereoscopic system that is proposed in this paper. Closer the convergence point means that the horopter² gets closer

¹ For simplicity, this paper assumes that a stereo rig is formed by 2 cameras.

² Horopter is the range of depth values within which objects can be measured by the stereo vision.

and shrinks the depth range and obviously the area or volume. Thus, the resolution of depth estimation should improve in theory. The RSECam could be extremely relevant for applications where the distance to the scene is constantly changing, like the AUV (autonomous underwater vehicles).

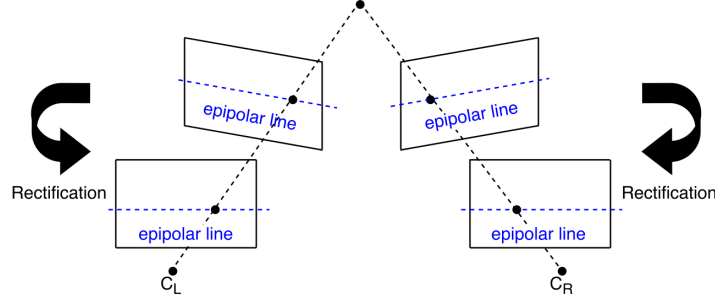


Fig. 2. The rectification process for convergence of the cameras.

Considering a parallel stereo rig, the distance to an object can be determined through a triangulation principle and can be approximated by: $Z(x, y) = f \cdot \beta / \text{disp}(x, y)$, where f is the focal length, β is the baseline and $\text{disp}(x, y)$ is the disparity between corresponding points (x, y) . Although the parallel configuration denotes a particular case for the stereoscopy and it is difficult to be achieved in real life applications due to mechanical misalignments of cameras, the large majority of stereoscopic sensors resort to this kind of configuration. Besides that, the extraction of 3D information from converged cameras can be as simpler as the parallel cameras by considering properties of the epipolar constraint, since the possible locations of a point that is seen in one image is a line on the conjugated image. Thus, a rectification process can be performed which defines a transformation of each image plane in the way that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes, see figure 2. This is intrinsically related to the extrinsic configuration of the stereo rig (mutual position and orientation).

A reliable extrinsic calibration is crucial for range measurements using the RSECam concept since these parameters are extremely important for the rectification procedure of converged cameras (a more detailed justification is presented in [2]). Obviously, new extrinsic parameters are required when the pose of cameras is modified. Therefore, this paper proposes a quadratic approximation of discretized and manually obtained extrinsic configurations. This quadratic approximation adjusts automatically the extrinsic configuration that is currently used by the RSECam, where a set of conventional calibrations (using the chessboard method) are conducted for different convergence angles. Thus, a set of translation matrices and the angular values can be modeled as a function of the convergence angle which is constrained by a maximum and minimum distance

to the targets. Besides the advantages of the RSECam that were already discussed in this paper, the concept is translated into a substantial increase of the mechanical complexity when compared to a conventional stereo sensors. In addition, the downside of converged cameras is related with a distortion effect caused by the two views (called the "keystone effect") and with mechanical tolerances and calibration errors. Therefore, a set of experiments must be conducted to evaluate the principles that support the RSECam concept before a complex and expensive prototyping of a setup that fulfill the requirements needed by AUVs (to be properly installed).

4 Results

A set of comprehensive experiments were conducted as part of this research in order to analyze the proposed RSECam. These experiments provide a preliminary study about the accuracy that could be expected from a self-reconfigurable stereo sensor under realistic underwater scenarios. The major goal is to validate the concept of the RSECam, which means to conduct a proof-of-concept regarding the dynamic convergence of cameras according to the average distance to a target (9×9 neighbors are considered). Therefore, the accuracy of the system is evaluated using the point cloud as a function of some factors such as the convergence angle (pose of cameras) and the baseline.

A simple but effective setup was developed for these trials: a stereo rig installed outside of a water tank with an acrylic window. This configuration is composed of two Mako-125C cameras with a 6mm lens, two stepper motors and one micro-controller. Each camera is supported by a base that is attached to a stepper motor and, moreover, the baseline of cameras could be modified by a horizontal slide. The micro-controller is responsible for controlling motors and, consequently, to converge or diverge the cameras according to the desired pose. The setup of the hardware and a diagram of the process that originates the point cloud is depicted in figure 3. First, the cameras configuration is defined and extrinsic parameters are obtained. Then, images are rectified, the stereo correspondence is performed and the disparity is converted to 3D points. The dynamic configuration of the RSECam is obtained by a quadratic interpolation of the extrinsic parameters that were obtained by manual calibration procedures from a set of 4 convergence angles (-1.1 , 6.3 , 12.3 and 16.8 degrees).

A set of objects were placed inside the water tank at known distances: $1.125m$, $2.125m$, $2.625m$ and $3.125m$. The distance error was analyzed in four distinct convergence angles (0° , 4° , 11° and 19.7°) and two baselines $0.11m$ and $0.19m$: figure 4 a) and b), respectively.

As expected, the error increases with the distance to a target. Figure 4 a) represents the results for a $0.11m$ of baseline and it shows that the object at higher distance could only be seen for a convergence angle between 0° and 4° . In general, the error evolution for the two closest objects demonstrate a consistent and significant decay for higher convergence angular values. The object at $2.625m$ showed a high error for an angle of 11° which could be interpreted as an

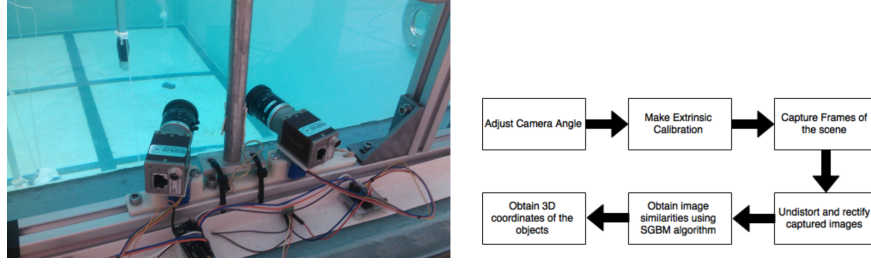


Fig. 3. The setup of the RSECam: a proof-of-concept (on the left). The architecture for extracting 3D information from the RSECam (on the right).

outlier since this error reduced to $0.22m$ when the angle was 19.7° (following the same error profile as the trial with the target at $2.125m$). This outlier could be motivated by the keystone effect, the difficulty in calibrating underwater stereo sensors and errors on the extrinsic parameters caused by a poor interpolation of the calibration samples (6.3° and 12.3°).

The trials conducted with a baseline value of $0.29m$ follow a profile similar to figure 4 a) however, with a higher average error. Figure 4 b) demonstrates that the best convergence angle was 19.7° for the closest object and about 4° for objects at $2.125m$ and $2.675m$. An increase of error (up to $0.25m$) was also verified in the 11° (for objects at $2.675m$) which decreases for a convergence angle of 19.7° . This can be justified by the increase of the baseline and the difficulties in calibrating the stereo system in underwater. Please noticed that object at larger distance cannot produce a relevant discussion since it goes outside the field of view of the sensor for half of the experiments. Additional trials contemplating more convergence angles must be conducted in the future.

Therefore, the results conducted in this research prove that the simultaneously manipulating of both the convergence and the interaxial gives control over the depth and the placement of objects within that 3D space which can improve the accuracy of 3D information retrieved from the environment. This could lead to significant advances for the navigation of AUVs however, more detailed experiments should be performed before the installation of a such sensor on a realistic robotic platform.

5 Conclusion

This paper presented a Reconfigurable stereoscopic system based on Self-adjusting cameras (RSECam) which is a stereo system that converges or diverges the cameras according to the average distance to the scene. A preliminary study presents the impact of modifying the convergence point (by changing the pose of the cameras and the baseline) on the quality of the information. A set of experiments were conducted as a proof-of-concept and to quantify if the self-adjusting capability represents any advantage when compared to conventional stereoscopic

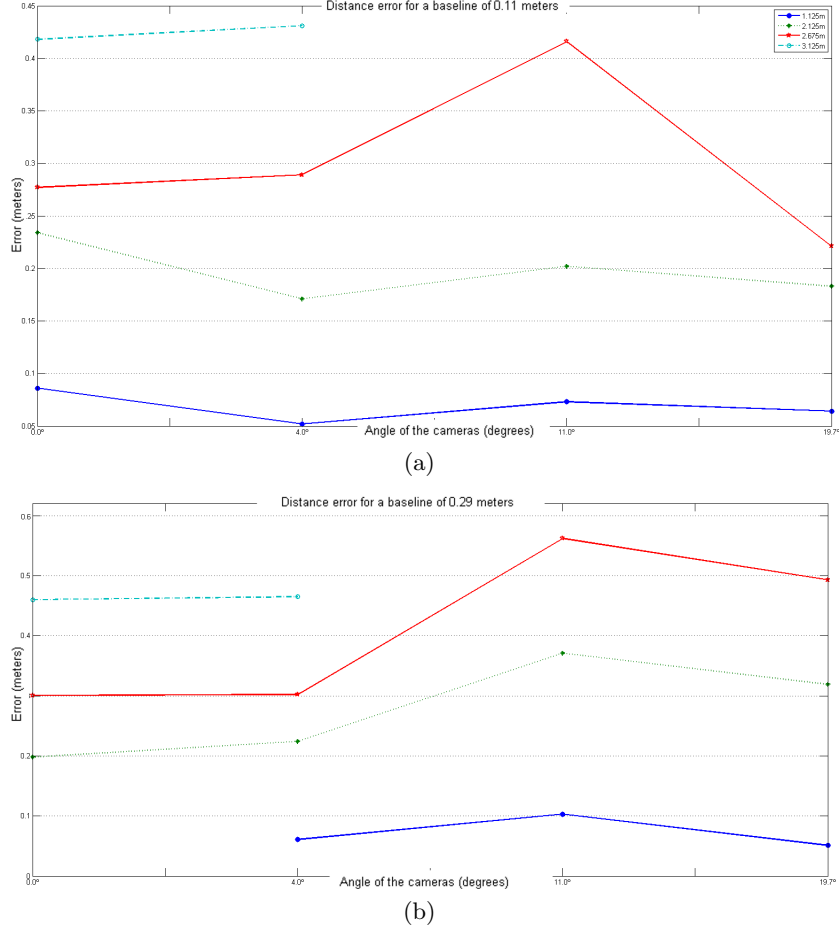


Fig. 4. The average error (in metric units) of 20 surrounding measures for the distance between the target to the stereo sensor as a function of the converged angle: considering a baseline of 0.11 meters and 0.29 meters, respectively. The absence of data represents objects outside the field of view of the sensor for that specific convergence angle.

systems. The results obtained from realistic underwater sequences proved that the accuracy of the 3D process was improved in some convergence angles. This is more evident for smaller distances (to the target) and for a stereo configuration with small baseline values since the error of the distance measured was decreased up to $0.10m$. This means that a vision sensor based on the RSECam will outperform conventional stereoscopic systems based on pre-defined configurations of the cameras.

For future activities, the RSECam technique will be tested in other realistic scenarios and, in addition, a navigation methodology will be developed to provide helpful and complementarity information for close range operations of AUVs.

Acknowledgments. This work is financed by the ERDF through COMPETE 2020 Programme within project "POCI-01-0145-FEDER-006961", and by National Funds through the FCT Portuguese Foundation for Science and Technology as part of project UID/EEA/50014/2013. Additionally, by the project "NORTE-01-0145-FEDER-000036" is financed by the NORTE 2020, and a H2020 project, under grant agreement No.692427.

References

1. F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A. Razionale. Experimentation of structured light and stereo vision for underwater 3d reconstruction. *Journal of Photogrammetry and Remote Sensing*, 66(4):508518, 2011.
2. Andrea Fusiello, Emanuele Trucco, and Alessandro Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.
3. J. S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101111, 1990.
4. J.-H. Li, M.-J. Lee, W.-S. Lee, J.-T. Kim, H.-J. Kang, and J.-H. Suh. Real time obstacle detection in a water tank environment and its experimental study. In IEEE, editor, *IEEE/OES Autonomous Underwater Vehicles (AUV)*, pages 1–5, 2014.
5. M. Massot-Campos and G. Oliver-Codina. Underwater laser-based structured light system for one-shot 3d reconstruction. In IEEE, editor, *IEEE Sensors*, pages 1138–1141, 2014.
6. D. McLeod, J. Jacobson, M. Hardy, and C. Embry. Autonomous inspection using an underwater 3d lidar. In IEEE, editor, *IEEE Oceans-San Diego*, page 18, 2013.
7. D. Moroni, M. A. Pascali, M. Reggiannini, , and O. Salvetti. Underwater scene understanding by optical and acoustic data integration. In Acoustical Society of America, editor, *Proceedings of Meetings on Acoustics*, pages 70–85, 2014.
8. K. Muljowidodo, M. A. Rasyid, N. SaptoAdi, and A. Budiyo. Vision based distance measurement system using single laser pointer design for underwater vehicle. *Indian journal of marine science*, 38(3):324331, 2009.
9. Andry Maykol Pinto, Miguel V. Correia, A. Paulo Moreira, and Paulo G. Costa. Unsupervised flow-based motion analysis for an autonomous moving system. *Image and Vision Computing*, 32(6-7):391–404, 2014.
10. Andry Maykol Pinto, A. Paulo Moreira, and Paulo G. Costa. An Architecture for Visual Motion Perception of a Surveillance-based Autonomous Robot. In IEEE, editor, *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 205–211, 2014.
11. L. K. Rumbaugh, E. M. Boltt, W. D. Jemison, and Y. Li. A 532 nm chaotic lidar transmitter for high resolution underwater ranging and imaging. In IEEE, editor, *IEEE Oceans-San Diego*, page 16, 2013.
12. S. Shatara, X. Tan, E. Mbemmo, N. Gingery, , and S. Henneberger. Experimental investigation on underwater acoustic ranging for small robotic fish. In IEEE, editor, *IEEE International Conference on Robotics and Automation*, pages 712–717, 2008.
13. B. Zheng, H. Zheng, L. Zhao, Y. Gu, L. Sun, , and Y. Sun. Underwater 3d target positioning by inhomogeneous illumination based on binocular stereo vision. In IEEE, editor, *IEEE Oceans*, page 14, 2012.