Feature extraction for underwater visual SLAM

Josep Aulinas, Marc Carreras, Xavier Llado
Joaquim Salvi, Rafael Garcia and Ricard Prados
Computer Vision and Robotics Group
Institute of Informatics and Applications
University of Girona
17071 Girona, Spain
{jaulinas,marcc,llado,qsalvi,rafa,rprados}@eia.udg.edu

Yvan R. Petillot
Oceans Systems Lab
School of Engineering and Physical Sciences
Heriot Watt University
Edinburgh EH14 4AS, United Kingdom
Y.R.Petillot@hw.ac.uk

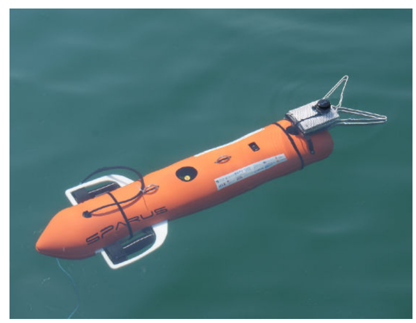
***Abstract*—Detecting and selecting proper landmarks is a key issue to solve Simultaneous Localization and Mapping (SLAM). In this work, we present a novel approach to perform this landmark detection. Our approach is based on using three sources of information: 1) three-dimensional topological information from SLAM; 2) context information to characterize regions of interest (RoI); and 3) features extracted from these RoIs. Topological information is taken from the SLAM algorithm, i.e. the three-dimensional approximate position of the landmark with a certain level of uncertainty. Contextual information is obtained by segmenting the image into background and RoIs. Features extracted from points of interest are then computed by using common feature extractors such as SIFT and SURF. This information is used to associate new observations with known landmarks obtained from previous observations. The proposed approach is tested under a real unstructured underwater environment using the SPARUS AUV. Results demonstrate the validity of our approach, improving map consistency.**

I. Introduction

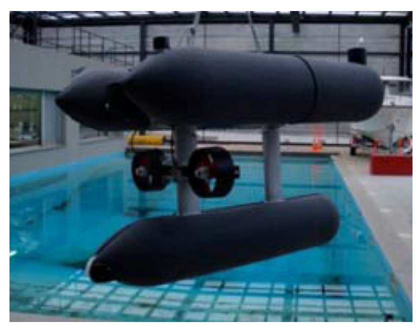
Many areas in oceans, seas and lakes are still completely unknown. For this reason, underwater research is nowadays gaining importance within the scientific community and underwater technology developers. Different underwater vehicles have been developed in order to explore underwater regions, specially those of difficult access for humans. Some examples of these vehicles are the ones developed at the Underwater Robotics Research Center (CIRS) at the University of Girona: ICTIENU (2006), SPARUS (2010) and GIRONA-500 (2011) (see Fig. 1). These vehicles are necessary not only to reach difficult places in the hydrosphere, but also to intervene in maintenance or reparation of structures, facilities or vehicles that work in the water. In order to conduct autonomously such tasks, it is vital to have precise and accurate information about the scene and the vehicle location.

To achieve this goal, the proposed system must enable 3D reconstruction from underwater imagery. For this reason, it is necessary to develop algorithms to take profit from available data. This is: 1) to use navigation and sensor data to improve the localization estimates, 2) to extract robust features from underwater imagery, 3) to develop reliable feature matching algorithms, 4) to determine the uncertainties associated with the process to allow loop detection, and 5) to make it a real-time feasible solution for large missions.

(a) ICTINEU.



(b) SPARUS.



(c) GIRONA-500.

Fig. 1. AUVs developed by CIRS at the University of Girona.

Autonomous Underwater Vehicles (AUVs) are equipped with on-board sensors, which provide valuable information about the vehicle state and the environment. This information is used to build an approximate map of the area and to calculate the approximate position of the vehicle within this map using the so called Simultaneously Localization and Mapping (SLAM) techniques [1]. SLAM is a process by which a mobile robot can build a map of an unknown environment and at the same time use this map to deduce its location. Initially, both map and vehicle position are not known. The vehicle has a known kinematic model and it is moving through the unknown environment, which is populated with several landmarks. The vehicle is equipped with sensory systems capable of taking measurements of the relative location between landmarks and the vehicle itself. SLAM techniques use the navigation and sensor data to improve localization estimates, while determining associated uncertainties.

In underwater robotics, the most commonly used sensors to measure navigation data on AUVs are the Inertial Measurement Unit (IMU) and the Doppler Velocity Log (DVL), while acoustic sensors are used to gather data from the environment, for instance, imaging sonar [2] or side-scan sonar [3]. However, the use of such acoustic devices does not give any intensity information, which might be necessary on intervention missions to detect specific objects, or might be useful when navigating through shallow waters. In addition, extracting robust features from acoustic images is a complex task due to the fact that the data is considerably noisy. Instead, optical video cameras provide further information that can be used to extract robust features.

Following this idea, in this work we propose a SLAM algorithm that uses optical video underwater images to perform navigation and to build a 3D map of the environment. These images contain regions of interest (RoIs) with salient features, which are useful to determine landmarks. Within SLAM framework, a landmark is understood as part of the map information and is used to update the map and localize the vehicle when observed a second time. Being able to identify when a landmark is reobserved is very important in order to close a loop. Notice that, closing a loop is important to improve map consistency and localization accuracy. For this reason, it is necessary to provide the system with algorithms capable to identify when a new observation corresponds to a new landmark or an already seen one. Therefore, detecting and selecting proper landmarks is a key issue to solve within the SLAM problem.

In this work, we present a novel approach to perform underwater landmark detection. Our approach is based on using three sources of information: 1) three-dimensional topological information from the SLAM problem; 2) image contextual information to characterize RoIs; and 3) features extracted from these RoIs. Topological information is taken from the SLAM algorithm, i.e. the three-dimensional approximate position of the landmark with a certain level of uncertainty. Contextual information is obtained by segmenting the image into background and RoIs. Features extracted from points of interest are then computed by using common features extractors such as Scale-invariant Feature Transform (SIFT) [23] and Speeded Up Robust Features (SURF) [24]. In our approach, SURF is used because is faster than SIFT. This information is then used to associate new observations with known landmarks obtained from previous observations. The method is evaluated through experimental validation, on a real unstructured underwater environment using the SPARUS AUV.

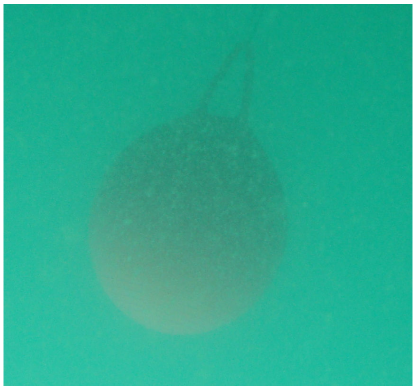
The paper is organized as follows: Section II presents the background behind this work by summarizing the most representative works on underwater computer vision; Section III presents the feature extraction and matching procedure; experimental validation is presented in Section IV, while Section V discussed the conclusions of this work.

II. Underwater Optical Imaging

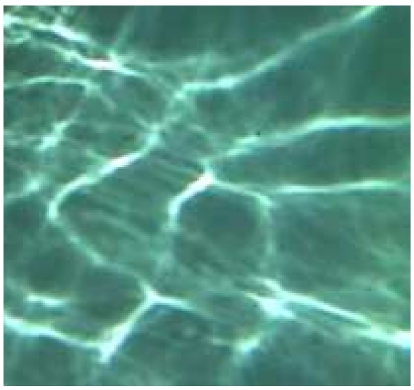
The interest on using optical cameras under the water increases with hardware improvements. Optical cameras provide high resolution imaging of the sea floor which is easy to interpret by operators and scientists. These images are useful for many applications, such as, inspection and maintenance of underwater man-made structures [4], wreck localization [5], mine countermeasures and seabed surveying [6]. In these applications, computer vision algorithms might be useful on station keeping [7], [8], cable tracking [9], [10], motion estimation (as a navigation aid) [12], localization [11] and/or mosaicking [13], [14]. Mosaicking strategies normally assume planarity, which in large scale mosaicking is not very realistic. Large areas can contain very rugged terrain, therefore, it is necessary to account for three-dimensional structure. In [Hartley 2000], the authors study extensively the theory to convert optical imagery to three-dimensional representations. Recent works use optical cameras to generate underwater 3D reconstruction of the scenario [6], [15], [16]. In all these approaches, computer vision algorithms are used to segment and interpret images, extract features, and perform the detection and classification of objects.

Features are selected to provide robustness in front of certain degree of distortion, so that the same point can be detected when observed from a different view point. Underwater images are very challenging, because apart from changes caused by camera motion, they normally suffer from specific artifacts due to the medium (see Fig. 2). These distortions are caused by diffusion, which produces low contrast, scattering, blur and loss of colour (see Fig. 2(a)); sun flickering, which depending on the shape of the water surface produces patterns randomly in all directions (see Fig. 2(b)); and also by non-uniform lighting (see Fig. 2(c)). Several approaches propose image processing algorithms to address these issues, for instance [17] presents an approach to correct lighting effects and [18] presents a technique to filter flickering.

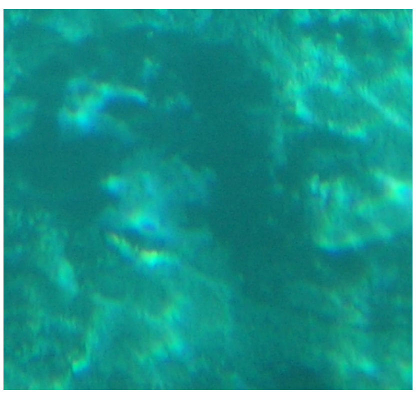
In the SLAM context, features must be distinguishable in order to simplify the association of new observations to corresponding map features. In general, SLAM approaches use features that can be detected by their location. For instance, features that are far apart from other features within the map. However, in underwater environments, it is interesting to have as much features as possible, and observe them repeatedly, in order to reduce the uncertainty caused by significant vehicle drift. In this sense, features from optical images are used either to estimate the motion in a frame to frame basis, but also as



(a) Diffusion.



(b) Sun flickering.



(c) Non-uniform light.

Fig. 2. Artifacts that appear on underwater images.

landmarks for the SLAM problem. These landmarks have to be very robust and features are commonly used to characterize them. Several methods to extract features from optical images exist. Edges, corner and contour detectors are commonly used in computer vision, for instance the well-known Canny edge detector [19], or the Harris corner detector [20]. These features are commonly used on cable tracking approaches [9] and on mosaicking [13]. In addition, texture patches are used to provide more information on the interest points, and to improve the matching step [14]. However, image patches show poor robustness to viewpoint changes and scale. A different invariant approach is to use moment based descriptors [21].

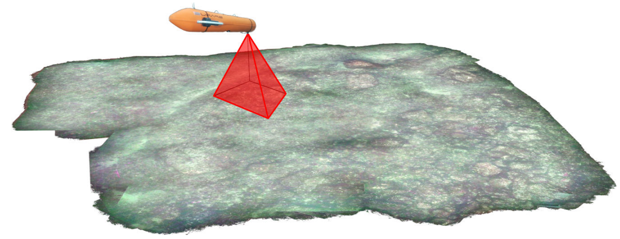


Fig. 3. Working principle for the SPARUS down-looking camera.

For instance, [22] uses Zernike moments, which are robust to scale and rotation. More robust approaches are SIFT [23] and more recently SURF [24], which produce rotation and scale invariant features. SIFT and SURF features are becoming important features in recent approaches [25], [26].

In most of these approaches, the output of the feature extraction step is a set of keypoints with its features and descriptors for every image. Feature matching algorithms are necessary to allow proper data association. Traditionally, the cross correlation between two image patches was used, but this metric is weak in front of slight rotations or scale variations. A common practice is to match these keypoints between two images bases on the similarity of their descriptors, i.e. the Euclidean distance between descriptor vectors. This approach is prone to find correct pairings, however, many features will not have a match because either they belong to the background or they were not detected in the second image. For this reason, SIFT and SURF matching algorithms use the same distance together with a comparison between neighbouring features, making the matching more robust [23].

III. Down-looking Optical Camera

AUVs are gaining importance on intervention missions. In order to conduct autonomously such tasks, it is necessary to have precise and accurate information about the scene. To achieve this goal, computer vision algorithms are necessary to enable 3D reconstruction from underwater imagery. These algorithms must extract robust features from underwater imagery and perform reliable feature matching.

SPARUS AUV is equiped with a down-looking camera, as shown in Fig. 3. This camera acquires three frames per second, like the one shown in Fig. 4(a). These images contain regions of interest with salient features, as shown in Fig. 4(b). These salient features will then be used as landmarks in the SLAM algorithm.

The idea behind the landmark detection used for SPARUS dataset is based on using context information to characterize a RoI and SURF features extracted from these RoIs. This information is then used together with the topological location of these landmarks to match new observations with known landmarks obtained from previous observations.

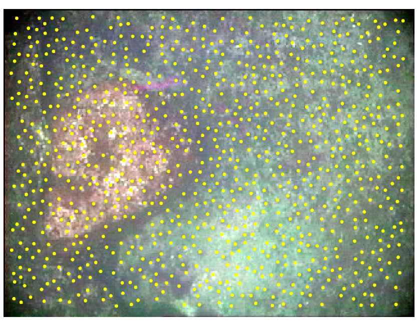
In what follows we will describe the details of each step:

A. Feature Extraction

Features are selected to provide robustness in front of certain degree of distortion, so that the same point can be detected

(a) Original image

(b) Salient features



(c) SURF features

Fig. 4. Underwater image and its salient features.

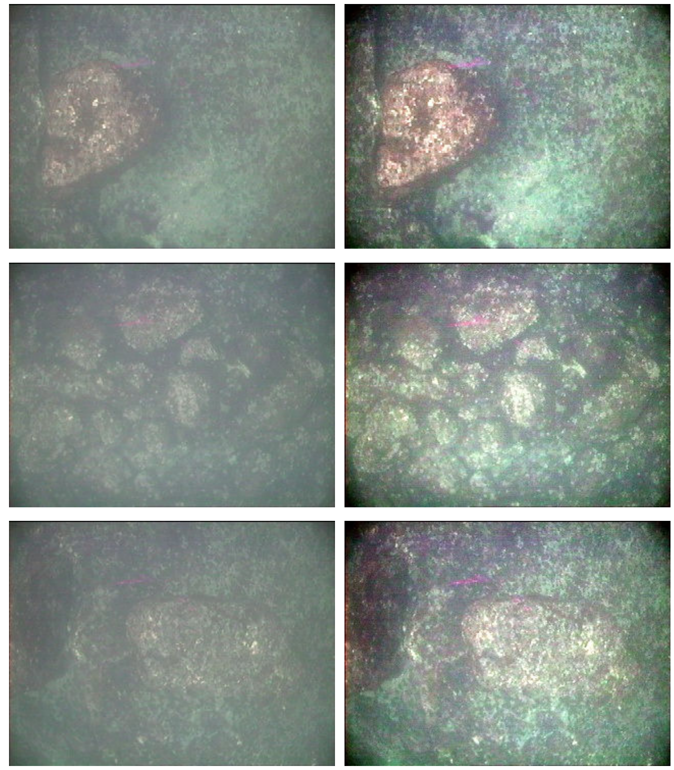


Fig. 5. Three different examples showing original images taken by SPARUS' camera (left column) and its corresponding preprocessed image (right column).

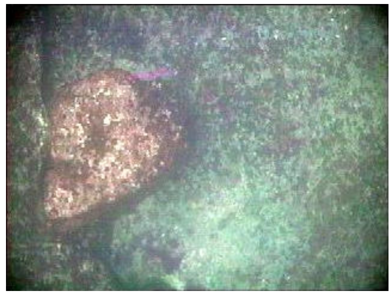
when observed from a different view point. The feature extraction procedure is show in Fig. 6. The process starts with an image preprocessing stage. Preprocessing consists of single channel selection, i.e., gray, followed by non-uniform light correction and a normalization. These preprocessing steps are done by common filtering techniques, in this particular case the ones presented in [18] are used. Results from this preprocessing step are shown in Fig. 5. In addition, lens distortion is corrected, using the calibration obtained with the well known Bouguet's calibration toolbox [27].

The second stage is focused on detecting RoIs within these images, i.e. segmenting RoIs. In order to do so, two parallel segmentation processes are computed. Both of them are based on image processing common operations. The first process starts with edge detection, producing the binary image shown in Fig. 6(b). Afterwards, erode/dilate operations are conducted, joining regions and eliminating insignificant spots (see Fig. 6(c)). Next step is a region search within this black and white image, producing the segmentation shown in Fig. 6(d). On the other hand, the second process uses the Hue channel (see Fig. 6(e)). This channel is then blurred in order to smooth the whole image. Afterwards, a threshold is applied, giving the results shown in Fig. 6(f)). This threshold is selected according to the mean value of the Hue image. Afterwards, a region search is conducted, producing the results shown in Fig. 6(g). At this point both processes are fused: a RoI is selected through the intersection of both segmentations (see Fig. 6(h)).

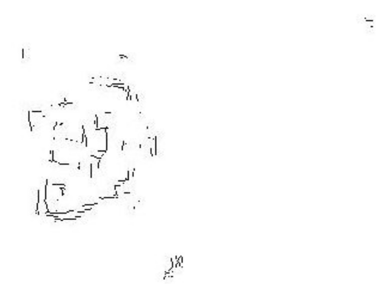
The third stage uses SURF features (see Fig. 6(i)). Depending on the previous step, if a RoI exists, then SURF features are extracted within this RoI and associated with it. Otherwise, if no RoI was segmented, SURF features are extracted within the whole image, and stored according to the camera pose from the moment they are extracted, for further matching when the camera revisits the same area.

B. Feature Matching

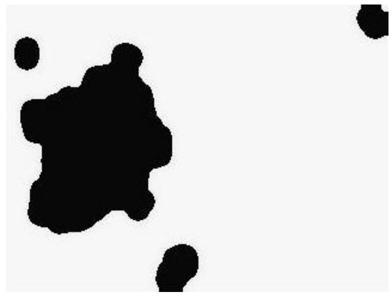
The matching approach used in this work is as follows. First, map information is used to obtain a first approximate of pairing candidates. The tree-dimensional position of a landmark and its uncertainty are the first constraint. Therefore, only new observations whose uncertainty intersects with known landmarks' uncertainty, are checked as possible pairing candidates. Initially, only few landmarks are in the map and their uncertainties might be small, producing only one candidate. However, as the mission goes on, more landmarks



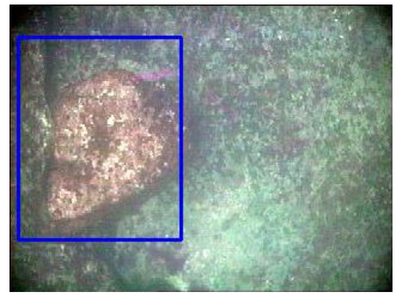
(a) Pre-processed image



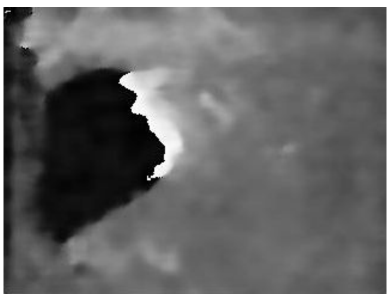
(b) Edge detection



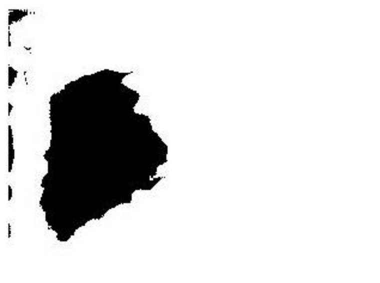
(c) Erode/dilate



(d) RoI-1



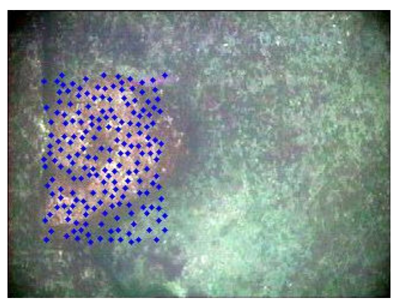
(e) Hue channel



(f) Binary image

(g) RoI-2

(h) RoI intersection



(i) SURF features

Fig. 6. Procedure to extract regions of interest (RoI). The final selected RoI is the one shown in red in h), and its SURF features are shown in blue in i).

are added in the map, and uncertainties may be larger. At this point, more than one pairing candidate will be found, and more information is necessary to find the correct match. Therefore, SURF matching algorithm is used to discard false matchings (see Fig. 7), together with the so called Joint Compatibility Branch and Bound (JCBB) algorithm [28]. JCBB addresses the issue of having multiple hypothesis by considering the compatibility of all the pairings globally. JCBB is very robust because it considers relative locations between features.

IV. Experimental Validation

Experiments were conducted on a sequence of images acquired by a down-looking camera on-board of the SPARUS AUV. These sequence was composed of 3199 frames of 320×240. SPARUS AUV was initially design to participate in the Student Autonomous Underwater Challenge – Europe (SAUC-E) competition. SPARUS is a torpedo shaped AUV that won the 2010 SAUC-E edition. SPARUS is equiped with several sensing devices: DVL, IMU, down-looking camera,

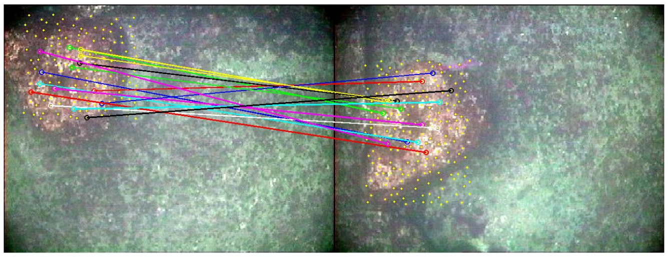


Fig. 7. This example shows two different observations of a stone. SURF features extracted from the ROIs are ploted in yellow. The ones that match in both images are connected with a line. The stone is observed from different angles, i.e. is rotated. This fact explains why the lines are not parallel.

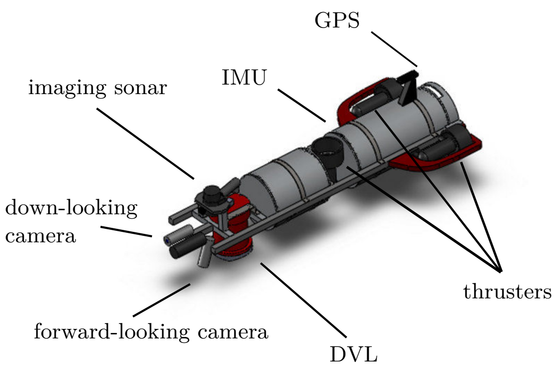


Fig. 8. SPARUS 3D model with its sensors.

forward-looking camera, imaging sonar and GPS (see Fig. 8). In this work, only DVL, IMU and down-looking camera were used, producing information about velocities, orientations and about sea floor.

Firstly, camera calibration parameters were obtained using a set of 30 images containing a chess board calibration pattern. Some of these images are shown in Fig. 9. Available images for calibration were considerably noisy, producing large calibration uncertainty, such as about ten pixels uncertainty for the principal point location.

Secondly, RoIs and features were extracted from the sequence of frames. The two process to detect RoIs were based on edge detection procedures and on hue channel selection, and afterwards the results from both processes were fused. The first process produced 5284 regions, the second process found 2908 regions, while the fusion of both processes defined a total of 1307 regions. Thirdly, the matching process found 627 matches. The performance of this matching process was evaluated through the SLAM approach presented in [29]. Correct matching produces proper SLAM updates, thus, improves vehicles trajectory estimates as shown in Fig. 10. Fig. 10(a) shows vehicle's estimated trajectory using dead reckoning, while Fig. 10(b) is the same trajectory estimated by SLAM. The main difference between these two figures is the significant improvement produce by SLAM, i.e. the first figure

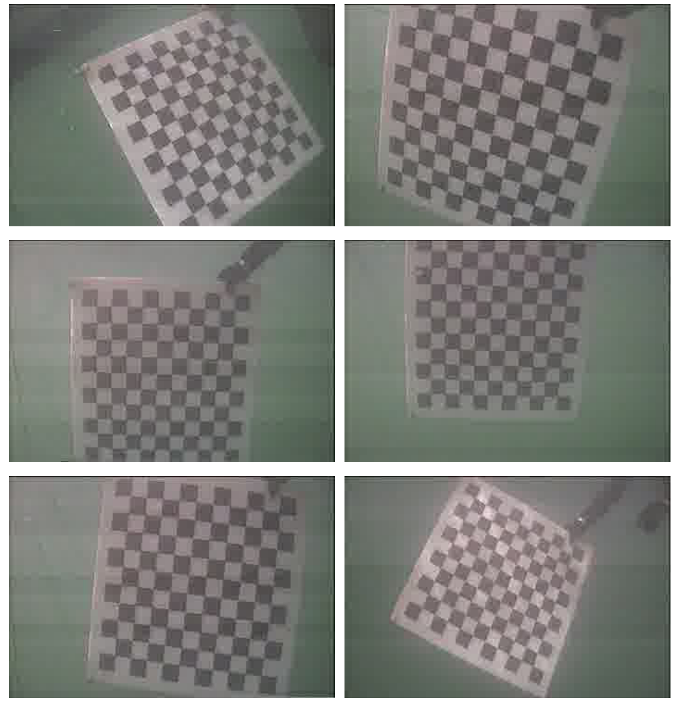


Fig. 9. A subset of the 30 images used to calibrate the camera on-board of SPARUS AUV.

shows the drift suffered when using only dead reckoning, which is addressed in the second figure.

V. Conclusion

In this paper, a novel approach to detect robust features from underwater optical imaging was presented. The idea behind feature extraction and matching on optical camera images was described step by step. The core of this approach relies on using SURF feature extraction and its corresponding matching algorithm, combined with common image processing techniques to determine regions of interest. After analysing the results, one can say that the method performs satisfactorily, because there is a significant improvement on the final SLAM estimate. Notice that these tests were conducted off-line, therefore they need further improvement to become real-time solutions.

Acknowledgment

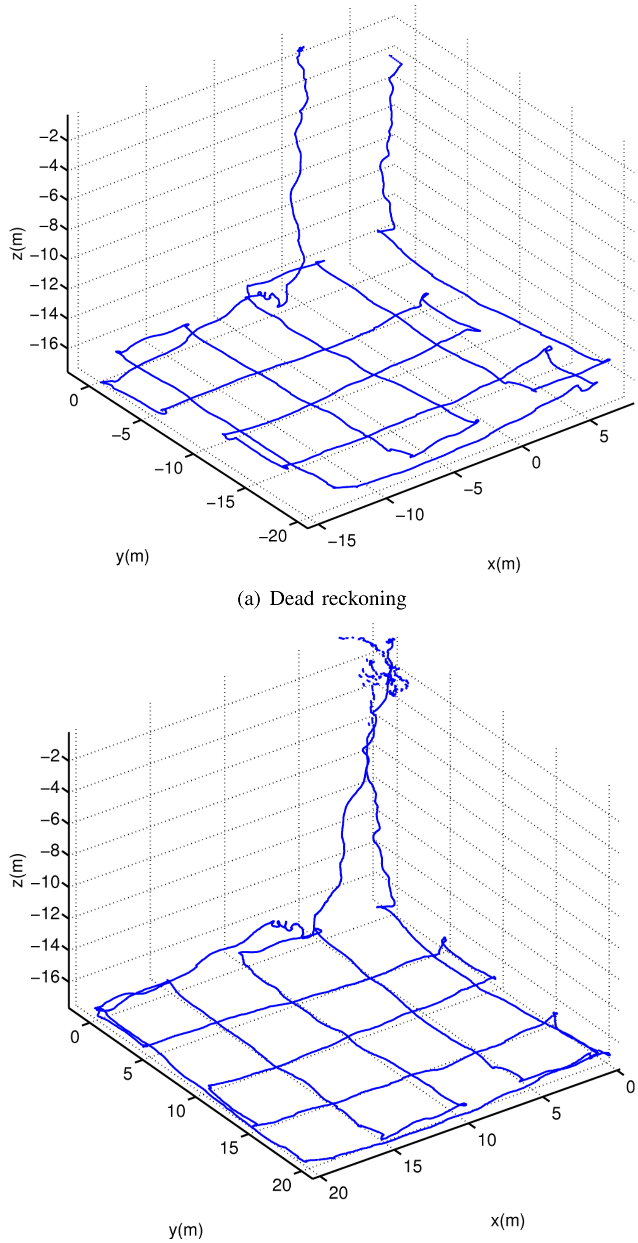
This work was partially funded through the Spanish Ministry of Education and Science (MCINN) under grant CTM2010-15216 and the EU under grant FP7-ICT-2009-248497.

References

[1] H. Durrant-Whyte and T. Bailey. *Simultaneous localization and mapping (SLAM): Part I The Essential Algorithms*. IEEE Robotics and Automation Magazine, vol. 13, no. 2, pages 99–108, 2006.

[2] D. Ribas, P. Ridao, J.D. Tardós and J. Neira. *Underwater SLAM in Man Made Structured Environments*. Journal of Field Robotics, vol. 25, no. 11–12, pages 898–921, 2008.

[3] J. Aulinas, X. LLadó, J. Salvi and Y. Petillot. *Feature based SLAM using Side-Scan salient objects*. In MTS/IEEE Oceans (OCEANS'10), Seattle (USA), September 2010.



(a) Dead reckoning

(b) SLAM

Fig. 10. 3D view of vehicle's trajectory. In a) one can observe the drift suffered during the mission, as the ending point is far from the starting point, while in b) this drift has been correct by the means of SLAM.

[4] M. Walter, F. Hover and J. Leonard. *SLAM for ship hull inspection using exactly sparse extended information filters*. In International Conference on Robotics and Automation, pages 1463–1470, Pasadena, CA, 2008.

[5] R. Eustice, H. Singh, J. Leonard, M. Walter and R. Ballard. *Visually Navigating the RMS Titanic with SLAM Information Filters*. In Proceedings of Robotics Science and Systems, pages 57–64, June 2005.

[6] M. Johnson-Roberson, O. Pizarro, S.B. Williams and I.J. Mahon. *Generation and Visualization of Large-Scale Three-Dimensional Reconstructions from Underwater Robotic Surveys*. In Journal of Field Robotics, vol. 27, no. 1, pages 21–51, 2010.

[7] X. Cufi, R. Garcia and P. Ridao. *An approach to vision-based station keeping for an unmanned underwater vehicle*. In IEEE/RSJ International Conference on Intelligent Robots and System, volume 1, pages 799 – 804, 2002.

[8] S. Negahdaripour and P. Firoozfam. *An ROV stereovision system for ship-hull inspection*. In IEEE Journal on Oceanic Engineering, volume 31, pages 551–564, 2006.

[9] A. Ortiz, J. Antich and G. Oliver. *A particle filter-based approach for tracking undersea narrow telecommunication cables*. Machine Vision and Applications, vol. 22, no. 2, pages 283–302, 2008.

[10] S. Wirth, A. Ortiz, D. Paulus and G. Oliver. *Using Particle Filters for Autonomous Underwater Cable Tracking*. In IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles, volume 2, pages 221–280, Killaloe (Ireland), 2008.

[11] R. Garcia, J. Batlle, X. Cufi and J. Amat. *Positioning an underwater vehicle through image mosaicking*. In IEEE International Conference on Robotics and Automation, volume 3, pages 2779 – 2784, 2001.

[12] R. Garcia, X. Cufi, P. Ridao and M. Carreras. Constructing photo-mosaics to assist uuv navigation and station-keeping, chapitre 9, pages 195–234. Robotics and Automation in the Maritime Industries, 2006.

[13] N. Gracias, S. van der Zwaan, A. Bernardino and J. Santos-Victor. *Mosaic-based navigation for autonomous underwater vehicles*. IEEE Journal of Oceanic Engineering, vol. 28, no. 4, pages 609–624, 2003.

[14] N. Gracias and S. Negahdaripour. *Underwater Mosaic Creation using Video sequences from Different Altitudes*. In Proceedings of MTS/IEEE OCEANS, volume 2, pages 1295–1300, 2005.

[Hartley 2000] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, 2000.

[15] J.M. Sáez, A. Hogue, F. Escolano and M. Jenkin. *Underwater 3D SLAM through Entropy Minimization*. In Proceedings of IEEE International Conference on Robotics and Automation, numéro 1642246, pages 3562–3567, 2006.

[16] T. Nicosevici and R. Garcia. *On-line robust 3D Mapping using structure from motion cues*. In MTS/IEEE Techno-Ocean Conference (Oceans'08), Kobe (Japan), April 2008.

[17] R. Garcia, T. Nicosevici and X. Cufi. *On the Way to Solve Lighting Problems in Underwater Imaging*. In IEEE OCEANS Conference (OCEANS), pages 1018–1024, 2002.

[18] N. Gracias, S. Negahdaripour, L. Neumann, R. Prados and R. Garcia. *A motion compensated filtering approach to remove sunlight flicker in shallow water images*. In Proceedings of the MTS/IEEE Oceans 2008 Conference (OCEANS), pages 1018 – 1024, 2008.

[19] J. Canny. *A Computational Approach To Edge Detection*. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pages 679–698, 1986.

[20] C. Harris and M. Stephens. *A combined corner and edge detector*. In In Proceedings of the 4th Alvey Vision Conference, pages 147–151, Manchester (UK), 1988.

[21] F. Mindru, T. Moons and L. Van Gool. *Recognizing color patterns irrespective of viewpoint and illumination*. In In Proceedings of Conference on Computer Vision and Pattern Recognition, pages 368–373, 1999.

[22] O. Pizarro and H. Singh. *Toward Large-Area Underwater Mosaicking for Scientific Applications*. IEEE Journal of Oceanic Engineering, vol. 28, no. 4, pages 651–672, 2003.

[23] D. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, vol. 60, no. 2, pages 91–110, 2004.

[24] H. Bay, T. Tuytelaars and L. J. Van Gool. *SURF: Speeded up robust features*. In In Proceedings of the ECCV'06: European Conference on Computer Vision, volume 3952, pages 404–417, 2006.

[25] T. Nicosevici, R. Garcia, S. Negahdaripour, M. Kudzinava and J. Ferrer. *Identification of Suitable Interest Points Using Geometric and Photometric Cues in Motion Video for Efficient 3-D Environmental Modeling*. In IEEE International Conference on Robotics and Automation, pages 4969–4974, Roma (Italy), April 2007.

[26] J. Salvi, Y. Petillot, S.J. Thomas and J. Aulinas. *Visual SLAM for Underwater Vehicles using Video Velocity Log and Natural Landmarks*. In MTS/IEEE OCEANS, pages 2082–2089, Quebec City (Canada), September 2008.

[27] Jean-Yves Bouguet. *Camera Calibration Toolbox for Matlab*, 2009. Retrieved 02 December 2009.

[28] J. Neira and J.D. Tardós. *Data Association in Stochastic Mapping Using the Joint Compatibility Test*. IEEE Transactions on Robotics and Automation, vol. 17, no. 6, pages 890 – 897, December 2001.

[29] J. Aulinas, X. LLadó, J. Salvi and Y. Petillot. *Selective Submap Joining for underwater Large Scale 6-DOF SLAM*. In IEEE/RSJ International conference on Intelligent Robots and Systems (IROS), pages 2552–2557, Taipei (Taiwan), October 2010.