Visual Characterization and Automatic Detection of Posidonia Oceanica for Meadows Mapping using an AUV*

Francisco Bonin-Font†1, Aleksandar Ćosić‡2, Miquel Massot-Campos§1, and Gabriel Oliver¶1

1 Systems, Robotics and Vision Group, University of the Balearic Islands (UIB), , Ctra Valldemossa km 7.5, Palma de Mallorca (07122) (Spain)
2 Robotics Laboratory, Institute Mihailo Pupin, University of Belgrade, (Serbia).

Abstract — *Detection and mapping of Posidonia Oceanica (P.O.) meadows in the Mediterranean sea is crucial to preserve the endemic ecosystems and to underscore the negative impact of many anthropogenic activities. Furthermore, navigating Autonomous Underwater Vehicles (AUV) in marine environments where the ground is densely colonized with P.O. is nowadays a challenging task, since this sea-grass generates dynamic and extremely textured areas, difficult to match in consecutive or loop-closing images. This paper presents two important contributions: a) a complete assessment to characterize areas of P.O. with visual features, in order to obtain visual localization data as accurate as possible; b) the automatic detection of P.O. in video sequences grabbed from the Sparus II AUV (*Autonomous Underwater Vehicle*) [1] in coastal areas; the images containing the detected sea-grass are used later to build visual mosaics showing the real extension and form of the P.O. meadow.*

Keywords — *Posidonia Oceanica, Visual Localization, Mosaicing Underwater Environments, Autonomous Underwater Vehicles.*

I. INTRODUCTION AND RELATED WORK

P. O. is an endemic sea-grass of the Mediterranean that forms large meadows. P.O. meadows play an extremely important role in the maintenance of the subsea ecosystems because they contribute on the sediment deposition, they attenuate currents and wave energy and stabilize unconsolidated sediments. They are also directly related to the abundance of biodiversity being a source of food and refuge of numerous animal species. Furthermore, recent studies have demonstrated that P.O. meadows generate and release greats amounts of oxygen to the water, augmenting its quality and transparency. The state of the P.O. meadows is directly affected by human activities. Since their extension is, in general, declining, new strategies are needed to monitor and control benthic habitats, because their preservation will affect, directly, critical activities for the tourist and fishing industries. The control of posidonia meadows is typically done using divers, who install certain structures in the perimeter of the meadow and certain markers in between to control height and extension. Sometimes, divers are tracked with acoustic localizers to build a georeferenced survey [1]. However, this process is slow, tedious, imprecise and limited by the autonomy of the scuba tanks. For that reason, and in the context of the ARSEA national project (TIN2014-58662-R), we propose to visualize, build and characterize 2D models of marine areas (also know as photo-mosaics) with P.O., using the SPARUS II AUV equipped with a camera looking downwards and navigating at a constant depth. These models will be associated to absolute locations and can be used, firstly, to measure the extension of the meadow, and afterwards, to control its state and evolution in time.

Robot autonomous navigation in marine environments with P.O. using visual sensors becomes an ambitious objective, due to the slight swaying of the sea-grass and the difficulty to find stable and trackable features. Mosaicing any kind of environment requires a robust image characterization to find the relative motion and position between overlapping images. But in underwater environments with P.O. this issue becomes strictly crucial, since the difficulty of finding, matching and tracking visual keypoints increases considerably with respect to other scenarios in land and water.

Results obtained with different visual key-point detectors and descriptors are different, in terms of computation time and accuracy in the matching and tracking procedures. Different features can lead to results with different precision, when computing visual odometry or visual loop-closure detection in a certain environment. In consequence, the use of those visual key-points that provide the most accurate localization and navigation data in underwater scenarios with P.O., is mandatory for the reconstruction of the 2D mosaics and their maximum reliability.

The literature is very scarce in visual approaches for P.O. automatic visual detection or for navigating underwater vehicles in scenarios with the presence of this seagrass. Authors of [2] describe the construction of photo-mosaics of P.O. meadows, built from images captured by a diver equipped with an integral system, which includes a Nikon camera and several sensors to measure the orientation of the camera, the depth, the water pH, oxygen, temperature and electrical conductivity. In [3], various techniques for texture segmentation are used to fit a line of a dominant P.O. texture in the image plane. This fitted line marks the separation between the P.O. meadow and the rest of the sea bed, and it can be used to guide an AUV along the boundary of a P.O. meadow. The analysis and exploitation of multispectral images provided by the satellites can provide accurate maps of P.O.: for instance, using images provided by SPOT 5 satellite [4], or, in [5], where 7 different classes of sea bottom are discriminate from images provided by LandSat 7.

II. VISUAL CHARACTERIZATION OF POSIDONIA OCEANICA

Vision-based real-time AUV self-localization and navigation in underwater environments with P.O. depends entirely in the capacity of the visual algorithms to find and track artificial landmarks (i.e. visual features) for odometry computation, or to register images that represent a loop-closing when computing visual SLAM [6] or mosaicing. Therefore, the success of both processes (feature tracking and image registration) applied in these environments will be strongly conditioned to the use of the appropriate feature detectors

*This work is partially supported by Ministry of Economy and Competitiveness under contracts TIN2014-58662-R and FEDER funds

†francisco.bonin@uib.es
‡aleksandar.cosic@pupin.rs
§miquel.massot@uib.es
¶goliver@uib.es

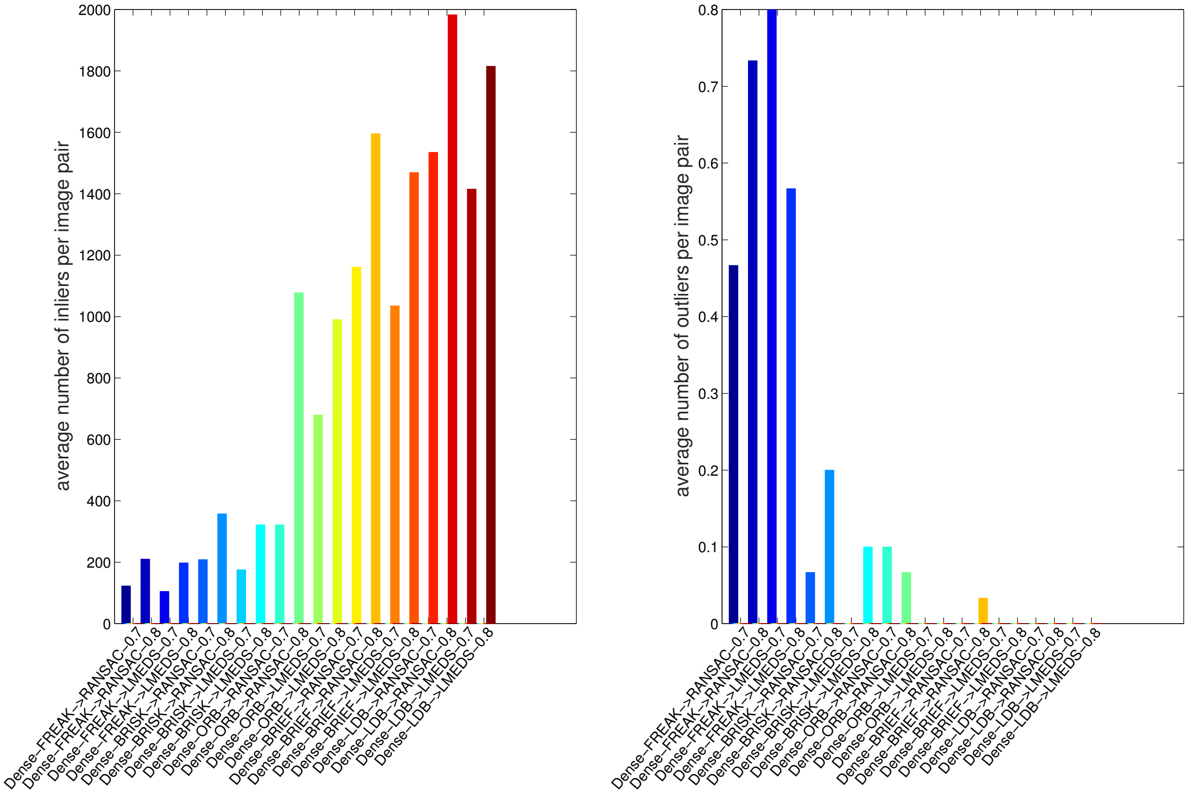

Fig. 1: Left: Mean of number of inliers in consecutive frames for the best 5 combinations detector/descriptor. Right: Mean of number of outliers.

and descriptors to get the best localization data. On the other hand, getting reliable and accurate mosaics of the underwater environment from all the images taken during the trajectory also depends directly on the quality of the visual key-points used to characterize the images. Ten different visual feature detectors (FAST [7], STAR [8], SIFT [9], SURF [10], ORB [11], BRISK [12], MSER [13], GFTT [14], Harris [15] and *OpenCv Dense* [16]) were combined (all with all) with seven different descriptors (SIFT, SURF, BRIEF [17], BRISK, ORB, FREAK [18] and LDB [19]) and tested on several video sequences grabbed at 10 fps from the AUV Sparus II in some coastal areas of Mallorca full of P.O.. The performance of every combination detector/descriptor has been assessed in two different kinds of tests:

A) All the combinations detector/descriptor are used to find and track the visual key-points (so called features) in consecutive images of the video sequences. Features are matched/tracked using a cross-matching process with two different matching thresholds (0.7 and 0.8). Then, outliers are filtered out imposing the epipolar constraint between the set of key-points of both images [20] and using two different types of iterative algorithms, LmedS [21] (Least Median of Square regression) and RANSAC [22] (Random Sample Consensus). The combination detector/descriptor that gives the highest number of inliers will be used to compute the visual odometry. The more inliers in the matching process across consecutive images, the more accurate will be the computed camera displacement.

Figure 1 shows, on the left, the mean of the number of inliers per pair of consecutive images across a sequence grabbed in the coastal waters of Mallorca, where P.O. is very extended. The mentioned data is shown for the best 5 combinations (detector/descriptor), that is, the combinations that generate a largest number of inliers between consecutive images. *OpenCV Dense* is the detector that generates the highest and almost constant number of features in all the images of the sequence. This algorithm establishes keypoints distributed densely and regularly over the image in a multilayer space grid, avoiding the need to apply additional bucketing functions to obtain features spread over the image uniformly. However, these features lack of any remarkable visual property or characteristic, which imposes its linkage with a robust descriptor. Figure 1 shows, on the right, the mean of outliers for the same combinations shown on the left. Both plots suggest that, being *OpenCV Dense* the detector that returns more key-points uniformly located in the image, the descriptors LDB and BRIEF, both using a cross matching threshold of 0.8 and RANSAC, are the descriptors that return the highest number of trackable inliers. Nevertheless, given the high number of inliers, both combinations will be equally useful for odometry computation. Besides, for these kind of scenarios, FRISK and BRISK are the descriptors that, being between the best 5, present a worse performance in the sense that they return the highest number of outliers, which means that, some of the features find by the detector are useless to be tracked/matched. With all that, BRISK with RANSAC is still a good option since the number of inliers in high enough (400) to get an adequate homography, and the number of outliers is very low. Note that, a) all the best descriptors are binary, decreasing considerably the execution time dedicated to the matching process, and, b) descriptors such as SIFT, SURF, STAR or FAST have to be avoided: SURF provides much less number of features, FAST generates a low number of inliers, and SIFT increases considerably the detection/matching running time; all these weaknesses lead to worse results when calculating the camera displacement.

B) The accuracy and reliability in the constructed photomosaics also depends completely on the performance of the visual feature detector/descriptor used to characterize the images. 2D photomosaics are build from images of each video sequence using the approach detailed in [23]. One of the main steps of this algorithm is the computation of a homography that relates two consecutive images and images that most likely overlap. Then, when all these homographies are computed, a process of bundle adjustment is run in order to recalculate and refine all relative transformations with respect to a global mosaic frame. A homography matrix is an affine transform that relates a set of co-planar points viewed from two different images. The homography contains the camera motion (rotation and translation) between the two images. And this motion is fundamental to obtain an accurate and reliable mosaic. If the height at which the underwater scene has been grabbed is much greater than the maximum vertical dimension of the relief of the seabottom, all visual key-points can be considered coplanar, thus they can be related via a homography. A homography between two images is calculated by trying to solve a system of multiple equations that relate the visual features from both images, and applying RANSAC [20]. Homographies reflect a more accurate and reliable motion of the camera as the robustness and the number of inliers matching between both overlapping images increases. So, the same combination of feature detector/descriptor that gives the best results in matching the inliers for the visual odometry calculation is also used to mosaic the areas with P.O. imaged using the SPARUS II.

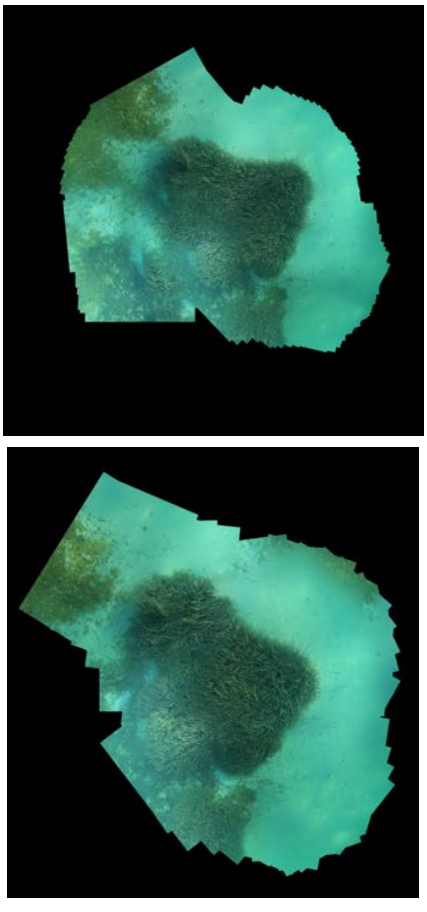

Fig. 2: Mosaic of a marine area with P.O. in the coastal area of Mallorca. Top: build with LDB features. Bottom: build with BRISK features.

Figure 2 shows 2 different photomosaics of the same marine area. The one on the top has been performed using the last version of LDB-32 binary descriptors [24] (which is invariant to in-plane rotations since it incorporates a dominant orientation for every patch and aligns the patch to this orientation before computing the descriptor). The mosaic on the bottom of Figure 2 has been build using BRISK descriptors, which are also, to a significant extent, invariant to rotation and scale changes, as the last version of LDB. Both mosaics incorporate 447 frames and have been created with the settings that give a result closer to the reality. The one on the bottom fits better to the real environment, and the one on the top presents slight errors in the compositions of the successive images, since the left upper part showing some algae is partially duplicated. BRISK is inherently more robust to rotation changes than LDB. This increases the accuracy in the mosaicing process, since it favors the detection of images that overlap and close a loop, although they have been taken in different time instants and from different view points. Note how the partially duplicated area located in the upper-left part of the top picture is located on the left-bottom in the bottom picture, and it is not duplicated.

III. AUTOMATIC DETECTION OF P.O.

The automatic detection of P.O. in the images will be useful to, a) correct dynamically the programmed mission of the AUV, in order to survey only areas with denser grass, and, b) process the video sequences directly from an automatic software module to calculate their extension and evolution. The P.O. is automatically detected in each image sequence using the approach presented in [25]. The method is based on Law's energy filters and statistical descriptors of the Gray Level Co-occurrence Matrix. A classifier is trained using a Logistic Model Tree, before its application in a certain environment. Figure 3 shows, in the bottom row, two images with P.O. corresponding to the scene mosaiced in Figure 2. The upper row of Figure 3 shows the detected areas of P.O. represented with superimposed gray patches. Patches with darker levels of gray indicate higher probabilities to contain P.O., and areas with lower levels of gray indicate less probability.

IV. CONCLUSIONS AND FUTURE WORK

The study of presence, state and evolution of P.O. grasslands must be a crucial activity in the way to preserve the marine ecosystems. Nowadays, this task is done using divers equipped with some technology to increase their imaging and localization capabilities, but with the typical limitations of operability time and security inherent to them. It is also possible, to a certain extent and with important limitations, to discriminate, in satellite images, the P.O. meadows present in coastal areas.

We propose the use of a lightweight hovering AUV equipped with cameras to survey marine areas in order to detect and quantize the presence of this sea-grass. The use of an AUV to perform this kind of tasks increments the operative time, it releases all the security measures addressed to humans and increases the sensorial capabilities of the system.

This paper has presented a robust way to characterize visually the P.O. in the images. An extensive set of feature detectors and descriptors has been applied in video sequences with the presence of P.O. to assess their performance in terms of number of features and their robustness in the tracking process across consecutive video frames. The combination that gives better results in terms of number of matches and number of inliers is later used to create photomosaics of the environment. A larger number of key points with an increasing robustness leads to a more accurate camera motion calculated between consecutive images, or between images that close a loop. Experiments on video sequences grabbed in coastal areas of Mallorca show how the combination of *OpenCV Dense* features described according to the last version of the LDB and BRISK approaches give the best result in terms of number of trackable inliers for motion calculation and in the later mosaicing process. SIFT and

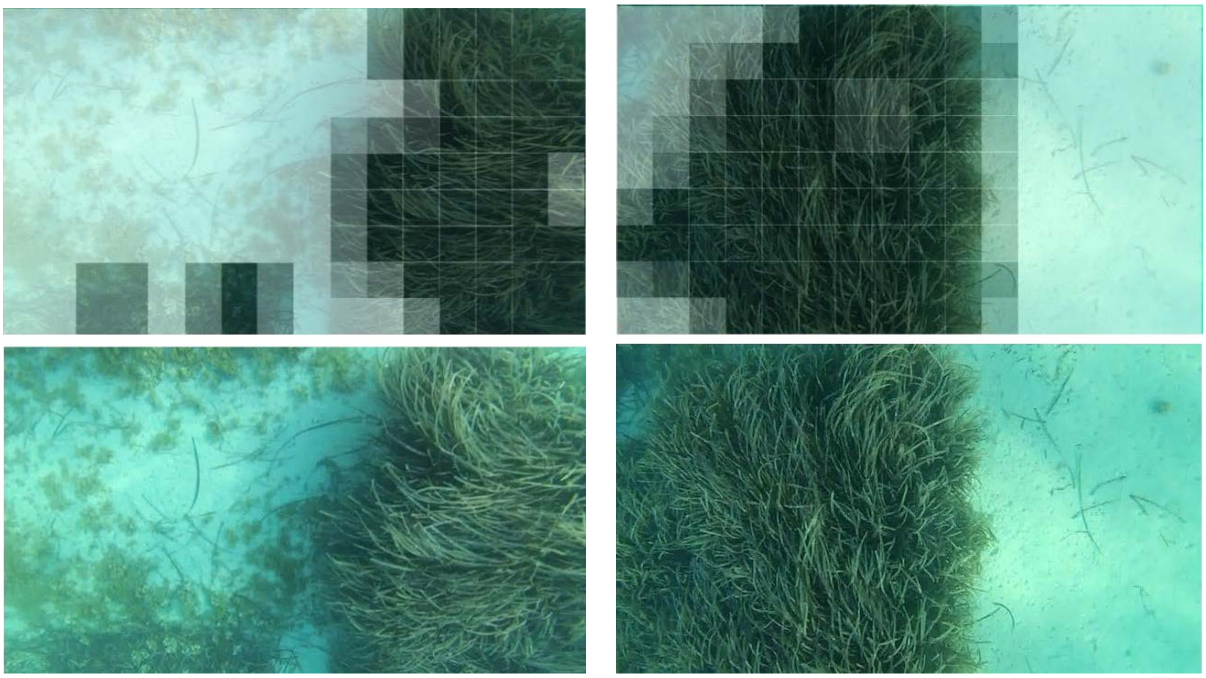

Fig. 3: Bottom: Images with P.O.. Top: Images with gray patches indicating areas which most likely contain P.O., super-imposed on the original frame.

SURF produce less feature inliers with a matching time much longer than LDB.

Our forthcoming work includes the construction of 2D photomosaics with all the images that contain the areas of P.O. automatically detected with [25], and the automatic calculation of the meadow extension.

REFERENCES

[1] M. Carreras, C. Candela, D. Ribas, A. Mallios, L. Mag, E. Vidal, N. Palomeras, and P. Ridao, "SPARUS II, Design of a Lightweight Hovering AUV," *Fifth International Workshop in Marine Technology (MARTECH)*, 2013.

[2] D. Scaradozzi, G. Conte, G. de Capua, L. Sorbi, C. Luciani, P. de Cecco, and A. Sorci, "Innovative Technology for Studying Growth Areas of Posidonia Oceanica," in *Proceedings of the IEEE WorkShop on Environmental, Energy and Structural Monitoring Systems.*, pp. 71–75, 2009.

[3] M. Barisic, D. Nad, and A. Vasilijevic, "Texture Segmentation Applied to P. Oceanica Beds' Upper Boundary Tracking by ROVs," in *Proceedings of the 3rd IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles*, pp. 318–324, 2012.

[4] V. Pasqualini, C. Pergent-Martini, G. Pergent, M. Agreil, G. Skoufas, L. Sourbes, , and A. Tsirika, "Use of SPOT 5 for Mapping Seagrasses: An Application to Posidonia Oceanica," *Remote Sensing of Environment*, vol. 94, no. 1, pp. 39–45, 2005.

[5] R. Matarrese, M. Acquaro, A. Morea, K. Tijani, and M. Chiaradia, "Applications of Remote Sensing Techniques for Mapping Posidonia Oceanica Meadows," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 906–909, 2008.

[6] A. Burguera, F. Bonin-Font, and G. Oliver, "Trajectory-Based Visual Localization in Underwater Surveying Missions," *Sensors, MDPI*, vol. 15, no. 1, pp. 1708–1735, 2015.

[7] E. Rosten and T. Drummond, "Fusing Points and Lines for High Performance Tracking," in *In International Conference on Computer Vision*, pp. 1508–1515, 2005.

[8] M. Agrawal, K. Konolige, and M. Blas, "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching," *Lecture Notes in Computer Science (ECCV-2008)*, vol. 5305, pp. 102–115, 2008.

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

[10] H. Bay, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, pp. 346–359, 2008.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an Efficient Alternative to SIFT or SURF," in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 2564–2571, 2001.

[12] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo From Maximally Stable Extremal Regions," in *Proceedings of British Machine Vision Conference*, pp. 384–396, 2002.

[14] J. Shi and S. Tomasi, "Good Features to Track," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, (Seattle (USA)), pp. 593–600, 1994.

[15] C. Harris and M. Stephens, "Combined Corner and Edge Detector," in *Proceedings of the Fourth Alvey Vision Conference*, vol. 15, pp. 147–151, 1988.

[16] OpenCv, "Dense Feature Detector." http://docs.opencv.org, 2014.

[17] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *Proceedings of the 11th European Conference on Computer Vision: Part IV*, ECCV'10, pp. 778–792, 2010.

[18] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[19] X. Yang and T. Cheng, "Local Difference Binary for Ultra-fast and Distinctive Feature Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 188–194, 2014.

[20] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[21] K. Shertzer and M. Prager, "Least Median of Squares: A Suitable Objective Function for Stock Assessment Models," *Canadian Journal of Fisheres and Aquatic Sciences*, vol. 59, no. 9, pp. 1474–1481 Vol.2, 2002.

[22] M. Fischler and R. Bolles, "Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[23] E. García, A. Ortiz, F. Bonnín, and J. P. Company, "Vessel Visual Inspection: A Mosaicing Approach.," tech. rep., 2015.

[24] X. Yang, C. Huang, and T. Cheng, "libLDB: A Library for Extracting Ultrafast and Distinctive Binary Feature Description," in *Proceedings of the ACM International Conference on Multimedia (MM), Open Source Software Competition*, pp. 671–674, 2014.

[25] M. Massot-Campos, G. Oliver-Codina, L. Ruano-Amengual, and M. Miró-Julià, "Texture Analysis of Seabed Images: Quantifying the Presence of Posidonia Oceanica at Palma Bay," in *Proceedings of IEEE MTS/IEEE Oceans*, pp. 1–6, 2013.