Limit Characterization for Visual Place Recognition in Underwater Scenes

Ana Rita Gaspar(✉), Alexandra Nunes, and Aníbal Matos

FEUP, INESC TEC, Porto, Portugal
{up201402645,up201402644,anibal}@fe.up.pt

Abstract. The underwater environment has some structures that still need regular inspection. However, the nature of this environment presents a number of challenges in achieving accurate vehicle position and consequently successful image similarity detection. Although there are some factors - water turbidity or light attenuation - that degrade the quality of the captured images, visual sensors have shown a strong impact on mission scenarios - close range operations. Therefore, the purpose of this paper is to study whether these data are capable of addressing the aforementioned underwater challenges on their own. Considering the lack of available data in this context, a typical underwater scenario was recreated using the Stonefish simulator. Experiments were conducted on two predefined trajectories containing appearance scene changes. The loop closure situations provided by the bag-of-words (BoW) approach are correctly detected, but it is sensitive to some severe conditions.

Keywords: Appearance-based localization · Bag of binary words · Place recognition · Loop closure · Stonefish · Autonomous underwater vehicles

1 Introduction

In the still unknown underwater world, there are structures that need to be inspected regularly to detect damages or corrosion, for example. Therefore, autonomous underwater vehicles (AUV) are increasingly being used to assist humans in some of these dangerous situations. To perform these tasks, the vehicles must be able to navigate in a feasible way but the nature of the underwater environment presents several challenges to accurate vehicle positioning. In addition, these tasks require close range navigation, reducing the sensors that can be used due to perceptual limitations. Typically, these scenarios have higher levels of distortion and noise caused by factors such as light incidence, wind, suspended particles, currents, or physical factors related to vehicle control. These problems make the perception of the environment challenging, but the selection of very robust features is crucial for similarity detection. For robust close-range operations, vision-based systems are the most attractive solutions for sensing

FCT - Fundação para a Ciência e a Tecnologia.

D. Tardioli et al. (Eds.): ROBOT 2022, LNNS 589, pp. 65–77, 2023.
https://doi.org/10.1007/978-3-031-21065-5_6

the environment, since can operate at a range of less than 3 m, providing high resolution and simplicity [1]. *But are the visual approaches able to deal with the inherent underwater environment challenges?* To detect a location that the vehicle has already visited is an essential aspect to compensate the accumulated pose deviations [2] but, making this decision independent of the environment - unsupervised learning - is yet a issue because it requires effective identification of landmarks to produce a consistent map. Binary algorithms are increasingly used for place recognition because these features have a very compact representation [3]. On the other hand, the visual BoW algorithm is one of the most commonly used techniques to perform quickly and robustly appearance-based loop closure recognition [4], which provides results with lower computational requirements [5]. The BoW approaches have many advantages namely its efficience thanks to the use of an inverted index and potentially hierarchical structures. Thus, these models represent local features in a fixed-length vector, using a visual vocabulary where each image feature is associated with a word. Given the advantages of using binary features and the strengths of BoW techniques, a behavioral evaluation of a binary BoW technique was previously performed in [6], based on a traditional learning approach for unsupervised environments - based on outdoor public datasets - that include scene changes, perceptual aliasing conditions, or dynamic elements. This BoW technique based on ORB features shows a good balance to deal with such difficult conditions at low computational cost. Therefore, it is also important to understand its behavior in an underwater environment, where locating and creating maps is more difficult because the underwater world is still unknown and uncontrolled where the appearance of a place can change over time. According to the previous experiments in outdoor challenging scenarios and based on a state-of-the-art evaluation of several feature detectors/extractors on underwater images [6,7], the ORB features are also considered for underwater environment. Therefore, as main contribution this paper evaluates the behavior of the binary BoW technique in underwater environment, to analyze the feasibility of using a purely visual solution for similarity detection against strict visibility conditions. Given the lack of data in this context, a typical mission scenario including their common issues that degrade the quality of the acquired data were simulated.

This work is organized as follows: Sect. 2 provides a general review of basic work in this area. Section 3 gives an overview of a purely visual place recognition system. It also describes the used BoW-based approaches and performance metrics. Section 4 describes the Stonefish simulator used to collect underwater data, recreating real conditions and reports the evaluation of the detection of previously seen places for the different performed experiments. Finally, the main conclusions of this work are discussed in Sect. 5.

2 Related Works

Proper loop closure detection must be use to correct accumulated drifts during navigation and mapping. Therefore, simultaneous localization and mapping

(SLAM) systems rely on a robust place recognition method for a robot to perform successful autonomous navigation. In this context, the first vocabulary approach for recognizing previously visited locations in an image sequence based on efficient - binary - features was proposed in [8]. It uses FAST keypoints and BRIEF descriptors and the vocabulary is built offline based on a hierarchical BoW model. BRIEF descriptor has been shown to tolerate only small scale or rotation changes. In order to not limit the system to only rectilinear trajectories and loop events with a similar viewpoint, this work was later extended by using ORB features [9]. It achieved higher recall (strength of the algorithm to recognize known places) than BRIEF in outdoor scenarios with larger viewpoint differences. Usually, the BoW model is predefined based on features extracted from a training set, which is limiting: the model is environment-dependent and cannot handle changes. This is critical in cases where the robot operates in uncontrolled environments, such as in the underwater scenario. Therefore, an appearance-based navigation and mapping method where visual vocabularies are created as visual information becomes available during vehicle survey was presented in [10]. From experiments, this incremental vocabulary approach showed be a promising solution but the inherent difficulty of the underwater environment makes the performance of the method slightly inferior to that in urban environments. Thus, in addition to the severe visibility conditions, the lack of robust, stable, and matching features in some areas is a relevant challenge for performing location recognition in the underwater environment. Later, a stereo-SLAM to detect loop closure situations in an underwater coast area was proposed [11]. The method was validated in an indoor water tank (marine scenario without 3D structure) and in a real outdoor region without compromised visibility. Compared with other state-of-the-art odometry methods, the proposed approach detects all loop closure situations with less motion estimation error, computational time and system resources in both scenarios.

3 Visual Place Recognition

Place recognition puts image retrieval (IR) in the practical context of physical agents operating in a physical world, and it is used to search an image database for images with similar visual content to a query image. The matching between current and database images is based on a ”similarity measure”. This measure determines which images are considered most relevant to the query image, where using a suitable similarity measure a high accuracy can be achieved. Thus, the information about the scene is independent of pose and error estimation, making these methods a possible solution for fast and robust loop closure detection. Since databases can be very large, it is important to develop techniques for fast access and search to facilitate similarity retrieval. Thus, a data structure that stores data in an appropriately abstracted and compressed form - indexing - is required [12]. In the context of place recognition, the most appropriate approach is the inverse index, which maps each index term to a document list in which it occurs. The indexing techniques can be divided into hash and non-hash-based.

The hash-based techniques use hash functions with search keys as parameters to generate address of data record. They has been used as an efficient way to store, group, and search data, namely for a high amount of data. On the other hand, the non-hash indexing techniques are generally used to speed up access of data. In such case, there are two main strategies for clustering: hierarchical (or agglomerative) clustering and algorithms based on point assignments. For the intended context, the tree-based structures (hierarchical) are the most appropriate. They are not considered best for large databases, but they are good for small databases, easy to understand, and allow efficient data entry (ordered data) and search. Therefore, also based on a comparative study of diverse place recognition methods [13], the DBoW2[1] and DLoopDetector[2] libraries were used [8] using ORB features. DBoW2 implements a hierarchical tree for approximating nearest neighbors in the image feature space and creating a visual vocabulary based on K-Means++ clustering. Along the BoW, an inverted and direct files can be kept to allow fast queries and feature comparisons. On the other hand, DLoopDetector was used for detecting loops in an image sequence. It implements temporal and geometric constraints between arbitrary pairs of images of the loop candidates to achieve more consistent results in detecting similar places. More detailed, for each current image, their features are converted into a bag-of-words vector, v_t and the database is searched for v_t, resulting in a list of matched candidates based on the weights of the words and their normalized scores (L1 distance). Only matches whose score exceeds a minimum threshold α are considered. The geometrical checking is based on direct index, taking advantage of the BoW vocabulary. The main steps for image similarity detection are shown in Fig. 1.

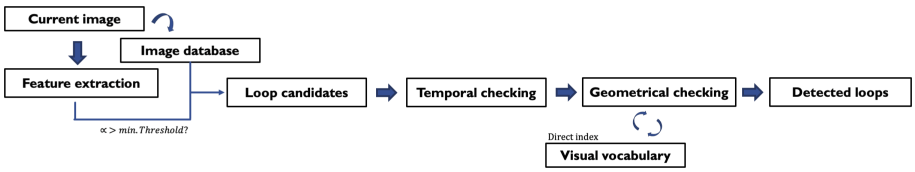

Fig. 1. Overview process of DLoopDetector for image similarity detection.

For evaluation against ground truth were counted the cases where the algorithm recognizes the queried image as a known place successfully (TP) or incorrectly (FP). The cases where the algorithm does not falsely recognize the queried image as a known place (FN) are also considered. Then, precision and recall metrics are computed: the first describes the robustness of the algorithm to detect a place without error; while recall determines the strength to detect known places without loss. To find an ideal combination of precision and recall, the F1-score - harmonic mean of both metrics - is also computed.

[1] https://github.com/dorian3d/DBoW2.
[2] https://github.com/dorian3d/DLoopDetector.

4 Experimental Results

In this section, an evaluation of the behavior of the proposed BoW scheme based on ORB features is provided for image similarity detection in underwater scenarios with challenging scenes. Experiments were conducted on several datasets created by the Stonefish simulator, including some common strict visibility conditions caused by external and inherent environment factors. First, in Sect. 4.1, the used Stonefish simulator tools and the predefined trajectories are described. Then, in Sect. 4.2, the performance of the binary BoW method for detecting previously seen locations under different operating conditions is demonstrated. All experiments were performed using an AMD Ryzen 7 2700X @ 3.7 Ghz with 16 GB RAM computer.

4.1 Stonefish Simulator

To evaluate the ability of the BoW technique in underwater environments, some datasets were created using the Stonefish simulator [14]. This is an open-source C++ library - version 1.1.0 - that aims to create realistic simulations of marine robots, taking into account the effects of light absorption and scattering. So, this library allows to simulate the ocean and atmosphere parameters. In terms of the ocean rendering it is possible to recreate currents, ocean optics - effects encountered in ocean waters, and to include suspended particles. Referring to atmosphere, it allows to create winds and to change sun's position on the sky. Therefore, to recreate a mission operation, port and archaeological seafloors were simulated, resorting to rendering process. Next, a camera was installed on the AUV and its parameters were configured to simulate an Allied Vision Mako G-125B/C GigE camera. Thus, a FOV of 44.2° is used and each image is composed of 300×200 pixels that corresponds an area of 2.7 m^2. In addition, positioning information was acquired by an odometry sensor - ground truth. Both sensors were configured with the same acquisition rate (10 Hz).

To collect data, the AUV was configured to autonomously perform close range predefined trajectories between waypoints. Thus, two different paths were defined: trajectory A is a simple route that ends at the starting position - just one situation where a loop closes. Trajectory B is a more complex path in the form of an "8". So, there are two different loop-closure situations where in the one of them, the robot revisits that area with a different point of view. Figure 2 shows the complete trajectories, highlighting the respective loop situations. Considering the intended context, each trajectory was performed with different parameters configurations to evaluate the impact of the scene appearance changes in visual place recognition. First, diverse water types were tested based on three Jerlov measurements to cover wide spectrum of the coastal water types. Next, to simulate the effects of different depths (that can indicate shallow and deep waters), three sun inclination angle were tested. Lastly, suspended particles were also included in the scenario to simulate some dinamism. Figure 3 shows an illustration of the effects of these environment parameters in the acquired images.

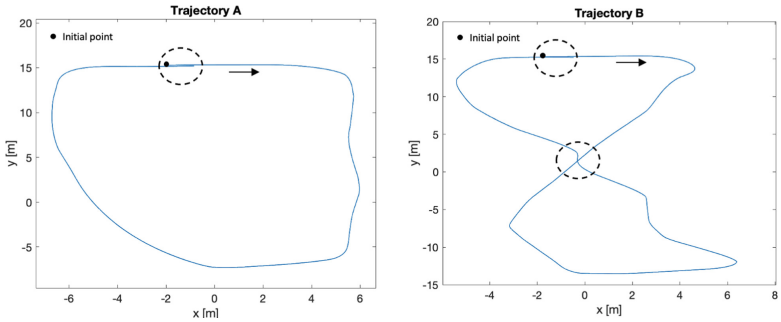

Fig. 2. Predefined trajectories performed by the AUV in the simulation environment.

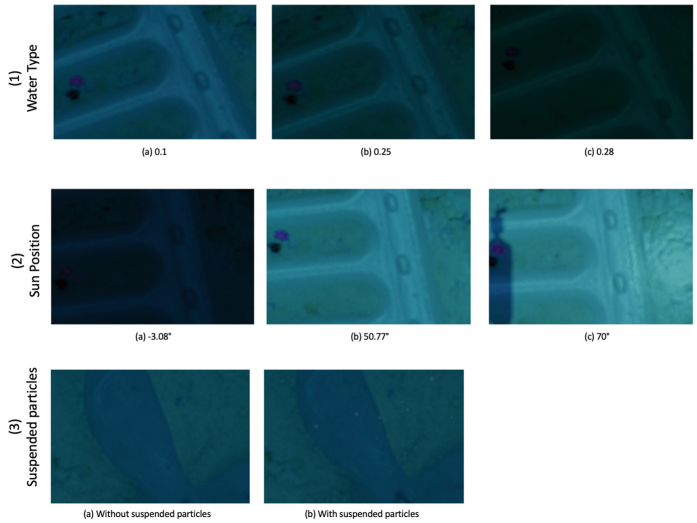

Fig. 3. Effect illustration of the three turbidity levels (1) and sun position (2) and the existence of suspended particles (3).

4.2 Loop Closure Detection

To measure the performance of the BoW scheme in the detection of the previously seen places, a ground truth for loop closure was created for each scenario. Each pose file is used to generate the ground truth, using approximately 1m as the threshold for placing two images at the same location based on the camera footprint. In addition, only the loops for which the difference between their image number and the image number of the previously detected loop is greater than the frequency F are considered as final loops, to prevent multiple loops from being detected in one second. The evaluation against the ground truth is done using the precision and recall metrics. Thus, for each sequence TP, FP and FN are counted as below described in Sect. 3. Full-indexing vocabularies - created

offline, but with images of the performed trajectory - are built for each scenario, to model the operating environment and evaluate the performance for detection of previously seen places similarly to online approaches. All experiments are based on the extraction of up to 1500 features per image.

The precision-recall curves for all scenarios in the Trajectory B, varying the threshold similarity parameter (α), between 0.2 and 0.35 are shown in Fig. 4.

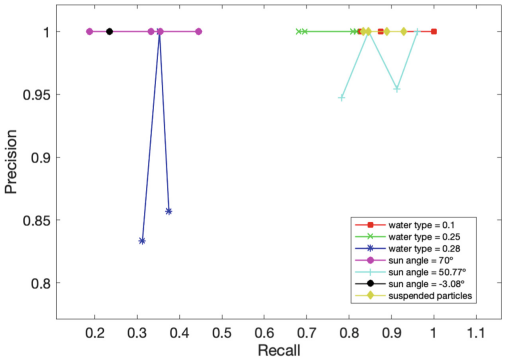

Fig. 4. Precision-recall curves in all scenarios in the Trajectory B for different values of the similarity threshold, α.

In water type = 0.28 and sun position = 50.77° conditions the behavior of the algorithm is not consistent: for some similarity values, a 100% of precision is not achieved (false loops are detected) which is not intended in a navigation context, since it may cause inappropriate trajectory adjustments and so, an incorrect pose estimation. In general, better precision-recall results are obtained with a similarity threshold of 0.2 or 0.3. Even so, with $\alpha = 0.2$ the similarity requirements are low, considering distant features as the same points and an overfitting result is obtained. Thus, the $\alpha = 0.3$ showed to be the most suitable and, so was used for all following experiments.

Thus, the three **levels of turbidity** - 0.1, 0.25 and 0.28 - were tested for both trajectories. As the Jerlov parameter increases, the image becomes darker, decreasing the image contrast which makes that the various elements there are in the image are no easily noticeable, as visible in Fig. 3 (1). For trajectory A with a Jerlov parameter configured to 0.1 the algorithm only fails to detect one loop closure (FN), as can be seen in Fig. 5. Therefore, a precision of 100% and a recall of 85.71% were obtained, which means that all loop situations were well detected.

In addition, combined with the wrongly not detected situation, resulted in an F1-score of 92.31%. In such scenario, the algorithm performance decreases only 8% as the Jerlov parameter increases, which was not expected given the strict visibility conditions visible to the naked eye. This can perhaps be explained by the fact that it is a simple trajectory with only one long-term loop acquired in same viewpoint in the different timestamps. Even so, looking at the number of the extracted features and the final matches between images, there is a decrease in this factor with increasing turbidity. So, it is then noticeable a higher difficulty in the detection of robust features, which can compromise a suitable navigation in this strict conditions.

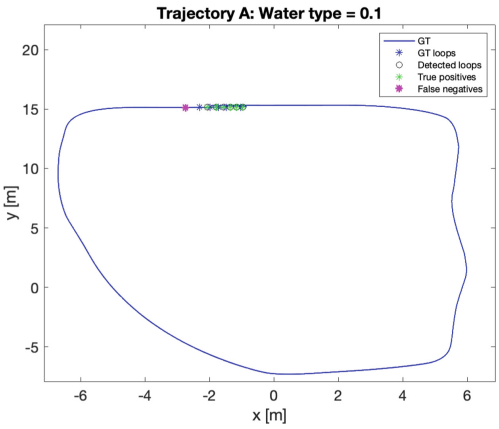

Fig. 5. Appearence-based loop closure for Trajectory A with water type=0.1.

For trajectory B the quality of the algorithm's performance deteriorated as the Jerlov parameter increased, as expected. Figure 6 shows the obtained results for each simulated water type (wt) value. In this trajectory, there are two loop-closure situations which makes harder to get a general good performance. As can be seen in this case, the decrease in image quality caused a delay in loop detection, and in the worst case, the algorithm is not able to detect the first loop situation.

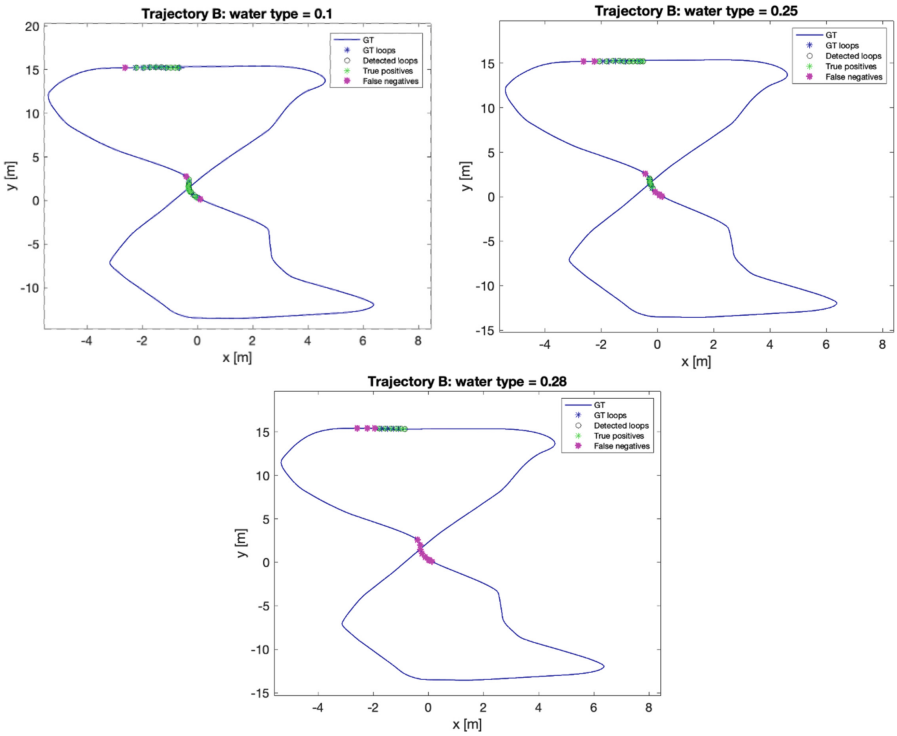

Fig. 6. Appearence-based loop closure for Trajectory B, varying the water type.

Table 1 shows the performance of the algorithm in detecting similar places, achieving an accuracy of 100% for all cases, which means that no loops were falsely detected, i.e. there are distinctive features.

Table 1. Results in the Trajectory B, varying the water type.

	wt = 0.1	wt = 0.25	wt = 0.28
Precision	100%		
Recall	87.50%	69.57%	35.29%
F1-score	93.33%	82.05%	52.17%

Contrary, recall decreased over the course of the experiments as environment conditions change, making the extraction of matching features more difficult. Thus, at wt = 0.25, the algorithm fails to detect 7 loop closure situations, implying a 20% decrease in recall. Even so, in both cases, it is able to deal with different viewpoints. In the last scenario, a larger reduction in recall is observed since none loop in the crossing area was detected. Thus, the algorithm does not deal with some challenges at the same time: dark environment, short loop situation (little

overlap between images and small similarity area), and viewpoint changes. This behavior is not reliable in a navigation context, since a false negative result can produces wrong trajectory adjustments and so, an unreliable pose estimation.

Next, the **sun position** was changed between the three inclination angles –3.08°, 50.77° and 70° - to evaluate the effects of illumination variation on the similarity detection of between images. This test is crucial to simulate the different common scenarios in underwater operations, namely deep areas (–3.08°) or shallow waters (70°). Figure 3 (2) shows the effect of each sun angle inclination in the acquired images. Table 2 illustrates the obtained quantitative results for trajectory A. For an angle of –3.08°, the behavior of the algorithm is similar to wt = 0.28, since the visibility is also very low (darker scenes). However, in this sun experiment there are more two false detections (false negatives), which corresponds to a 27% decrease in recall. At an angle of 50.77°, the algorithm achieves higher performance with a 12.5% increase in recall. In the last scenario, the obtained recall is the same compared to the previous result (sun angle = 50.77°), but there is a decrease in the number of extracted features as well as identifiable matches. Although the image quality is higher for humans, the algorithm is more sensitive to this last configuration - 70° of sun inclination - compared to the worst visibility condition in terms of water type (wt = 0.28) and shows a lower recall - 77% versus 62.5%. In fact, bright scenes have lower image contrast, which makes the feature extraction and so, a similarity detection more difficult.

Table 2. Results in the Trajectory A, varying the sun position.

	angle = –3.08°	angle = 50.77°	angle = 70°
Precision	100%		
Recall	50.00%	62.50%	62.50%
F1-score	66.67%	76.92%	76.92%

Figure 7 shows the results of the same experiment but, now for trajectory B. It can be seen that increasing the illumination level, the algorithm leads to better loop-closure detections but, when the sun inclination is high (clear and bright scenes), the performance decreases and some loops are not detected. Once again, this can be explained by the fact that higher sun inclination produces more unstructured images that make feature extraction more difficult. Another problem caused by a higher sun inclination can be the appearance changes caused by the vehicle shadow. These changes are dynamic and therefore can affect the similarity detection between images acquired in different timestamps. Thus, in Table 3, the significant drop in place recognition performance is demonstrated for extreme sun inclinations. In the darker scenario (–3.08°), the algorithm has a delay in starting similarity detection - one more false negative. Nevertheless, both cases are not suitable for real-world use, since there are many unrecognized loop closures. In the best cases (sun inclination = 50.77° and wt = 0.1), the algorithm behaves similarly and fails to detect 3 loop closures.

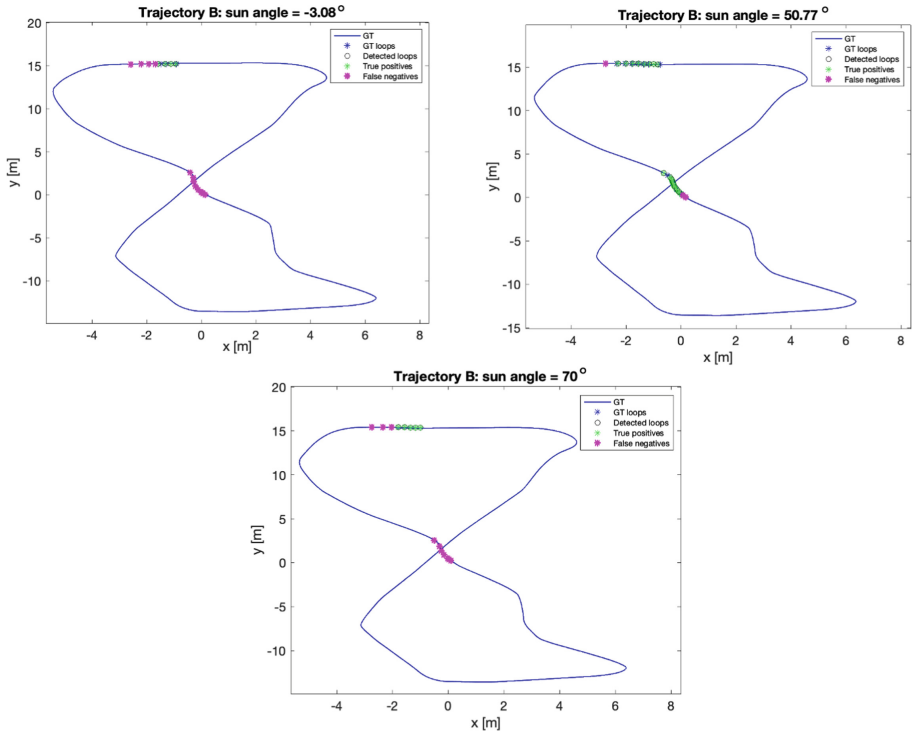

Fig. 7. Appearence-based loop closure for Trajectory B, varying the sun position.

Table 3. Results in the Trajectory B, varying the sun position.

	angle = –3.08°	angle = 50.77°	angle = 70°
Precision	100%		
Recall	23.53%	84.62%	33.33%
F1-score	38.10%	91.67%	50.00%

Lastly, the **suspended particles** effects are therefore shown in the Fig. 8. It was expected that the algorithm would misdetect loop-closure situations (precision<100%) and fails to detect others (FN), as the dynamics change the appearance of the scene. But, only the last situation occurs, perhaps because the amount of particles possible to simulate is small and sparse. Compared to the experiment without particles (wt = 0.1), the recall decreases only by 3%. Moreover, the number of final matches are also lower which can indicates that the particles added no value in the feature extraction - no matchable features.

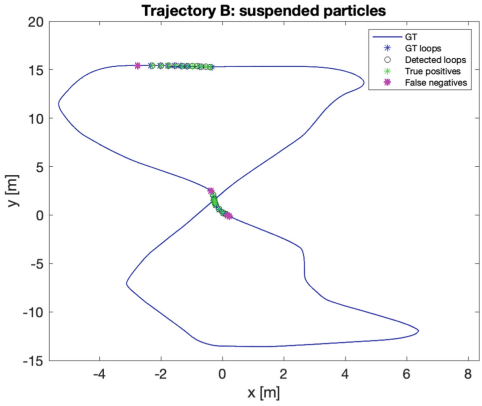

Fig. 8. Appearence-based loop closure for Trajectory B with suspended particles.

5 Conclusions

In this paper, a behavioral evaluation of a BoW technique based on ORB features for an underwater scenario was presented, including some common problems of a typical operation mission that affect the quality of the acquired images. The Stonefish simulator was used, relying on its configuration features to render the seafloor and some general inherent conditions for the intended context. A simple trajectory with an one-time and long-term loop closure situation was performed by the AUV. All loops detected by the algorithm were correctly identified. In this data, the scene changes did not affect too much the overall performance, since the loop closure situation was always detected, albeit with a delay of one detection in the worst case of water type and sun position, achieving a recall of 77.78% and 50%, respectively. For trajectory B, the algorithm performance decreases as the Jerlov parameter increases. With wt = 0.25, the recall rate decreases by 20% as the algorithm fails to perform 7 detections. With poorer visibility (wt = 0.28), the algorithm fails to detect the first loop closure situation - can be dangerous in the navigation context since there are no trajectory adjustments and consequently an incorrect pose estimate is obtained. Sun inclination experiments showed poor place recognition performance for extreme values (–3.08° and 70°). The first case illustrates a depth operation where a darker environment produces scenes without large differences in intensity, i.e. recognisable features. Contrary, shallow waters (70°) can produce unstructured images and shadows, which also makes feature extraction a harder task. Thus, it has been showed that visual similarity detection is nevertheless sensitive to some visibility conditions, which are common in uncontrolled and challenging scenarios. So, it could be interesting to combine cameras with other sensors to evaluate a possible hybrid solution for place recognition. In the near future, the parameters will be dynamically changed during the simulation and a performance evaluation will be accomplished in a scenario with repetitive patterns and few features.

Acknowledgements. This work is financed by FCT - Fundação para a Ciência e a Tecnologia - and by FSE - Fundo Social Europeu through of the Norte 2020 – Programa Operacional Regional do Norte - through of the doctoral scholarship SFRH/BD/146460/2019. This work is also financed by K2D Project - Knowledge and Data from the Deep to the Space (POCI-01-0247-FEDER-045941) funded within the scope of MIT Portugal.

References

1. Lu, H., Li, Y., Zhang, Y., Chen, M., Serikawa, S., Kim, H.: Underwater optical image processing: a comprehensive review. Mob. Netw. Appl. **22**(6), 1204–1211 (2017). https://doi.org/10.1007/s11036-017-0863-4
2. Melo, J., Matos, A.: Survey on advances on terrain based navigation for autonomous underwater vehicles. Ocean Eng. **139**, 250–264 (2017). http://dx.doi.org/10.1016/j.oceaneng.2017.04.047
3. Tareen, S.A.K., Saleem, Z.: A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In: International Conference on Computing, Mathematics and Engineering Technologies, pp. 1–10 (2014)
4. Fuentes-Pacheco, J., Ruiz-Ascencio, J., Rendón-Mancha, J.M.: Visual simultaneous localization and mapping: a survey. Artif. Intell. Rev. **43**, 55–81 (2015)
5. Kejriwal, N., Kumar, S., Shibata, T.: High performance loop closure detection using bag of word pairs. Robot. Auton. Syst. **77**, 55–65 (2016)
6. Gaspar, A.R., Nunes, A., Matos, A.: Evaluation of bags of binary words for place recognition in challenging scenarios. In: 2021 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2021, pp. 19–24 (2021)
7. Hidalgo, F., Bräunl, T.: Evaluation of several feature detectors/extractors on underwater images towards vSLAM. Sensors. **08**, 20 (2020)
8. Gálvez-López, D., Tardós, J.D.: Real-time loop detection with bags of binary words. In: International Conference on Intelligent Robots and Systems, pp. 51–58 (2011)
9. Mur-Artal, R., Tardós, J.D.: Fast relocalisation and Loop closing in Keyframe-based SLAM. In: IEEE International Conference on Robotics and Automation, pp. 846–853, June 2014
10. Nicosevici, T., Garcia, R.: Automatic visual bag-of-words for online robot navigation and mapping. IEEE Trans. Rob. **28**(4), 886–898 (2012)
11. Negre, P.L., Bonin-Font, F., Oliver, G.: Cluster-based loop closing detection for underwater SLAM in feature-poor regions. In: IEEE International Conference on Robotics and Automation. IEEE, pp. 2589–2595 (2016)
12. Sharma, S., Gupta, V., Juneja, M.: A survey of image data indexing techniques. Artif. Intell. Rev. **08**, 52 (2019)
13. Wang, H., Wang, C., Xie, L.: Online visual place recognition via saliency re-identification. In: IEEE RSJ International Conference on Intelligent Robots and Systems, pp. 5030–5036 (2020)
14. Cieślak, P.: Stonefish: an advanced open-source simulation tool designed for marine robotics, with a ROS interface. In: OCEANS 2019 - Marseille, pp. 1–6 (2019)