Martech 2021.
Marine Technology Workshop
16,18th June. Vigo, Spain

Image Hashing for Loop Closing in Underwater Visual SLAM

Francisco Bonin-Font[1], Antoni Burguera Burguera[1], Gabriel Oliver Codina

[1] *Systems, Robotics and Vision Group (University of the Balearic Islands, ctra de Valldemossa, km 7,5, 07122 Palma de Mallorca, francisco.bonin@uib.es, antoni.burguera@uib.es)*

***Abstract* – This article presents an experimental assessment of a hash-based loop closure detection methodology specially addressed to Multi-robot underwater visual Simultaneous Localization and Mapping (SLAM). This methodology uses two different top quality image global descriptors, one learned (NetVLAD) and one handcrafted (HALOC). Complete tests were done to compare the performance of both hashing techniques applied in an extensive set of real underwater imagery.**

***Keywords* –** *Visual Loop Closing Detection, Underwater Robotics, SLAM, Convolution Neural Networks.*

I. INTRODUCTION AND OVERVIEW

Loop closing, which consists in detecting whether the robot is observing a previously visited area [1], is one of the most important tasks in a visual Simultaneous Localization and Mapping (SLAM) module for *Autonomous Underwater Vehicles* (AUV). This observation is reflected in pairs of images that overlap, regardless the differences in orientation, scale or viewpoint. Closing a loop is more than a simple visual place recognition, but also an image registration part, that is, the extraction of a transform, in translation and rotation, between the pair of images that close the loop. Large-scale or long-term missions projected with a single AUV generate large maps with huge amounts of visual data. A common approach to mitigate this issue is to join, in a common origin, different trajectories (so called *sessions)* estimated by one or multiple robots, in different times or simultaneously [2]. Due to the lack of geometric constraints between different sessions, one way to detect inter-session loop-closings is to apply brute-force visual feature matching techniques between an image of one session and all images of another session. But this process is extremely costly in computational resources and time, and requires, specially in centralised *Multi-Robot systems*, the delivery of huge amounts of visual data to the master robot [2]. Transferring images between underwater robots is particularly problematic since underwater communications bandwidth is extremely reduced if using acoustic baselines and blue light modems applied underwater require short distances and a certain directionality between the transmitter and the receiver. One way to solve this problem is reducing the images to a short vector of numbers, so called a global image signature, or a hash for short. Traditionally, a hash is usually used as a digital signature to authenticate dispatched messages. Conventional hashes are extremely sensitive; a change in 1 bit of the input message changes the output dramatically [3]. However in applications of scene recognition or visual loop closing, it is accepted that similar or overlapping images produce similar or close hashes while distinct images produce clearly distinctive hashes [4]. Hashes are very fast to calculate and to compare, reduce drastically the data storage and exchange, and do not compromise the global localization process since, hashes are obtained from visual features, and, in visual-based navigation architectures, visual features have to be obtained for all images, to calculate the visual odometry and to register images that close loops.

In the context of the ongoing national project TWINBOT (TWIN roBOTs for Cooperative Underwater Intervention Missions) [5], diverse missions of cooperative intervention must be run using one or several AUVs in underwater areas with different benthic habitats. Sharing visual data between different robots is necessary to estimate a common map. This paper presents an essentially applied work: the assessment of a hash-based loop closing detection methodology that uses, alternatively, two top image hashes that showed excellent performance: 1) NetVLAD [7], a well known *Convolutional Neural Network* (CNN) architecture specially trained for weakly supervised place recognition in urban environments, and that, once trained returns a global descriptor in the form of a vector with 512 floats, and b) HALOC [6], a handcrafted global image descriptor formed by a vector of 384 floats; HALOC has already demonstrated a great performance in limited underwater scenarios [8].

In order to save time and resources, the proposed methodology consists in: a) All images of all sessions are hashed. Find a set of N candidates to close a loop with a current query, as those images among all captured in other sessions that present the lowest difference, in terms of L1-norm, between their hash and the query hash, b) Candidates to close a loop with the query are confirmed by means of a RANSAC-based process described in [9], c) Candidates confirmed by RANSAC will be True or False Positives (TP/FP) if they really do close or not a loop (determined visually); candidates rejected by RANSAC will be either True or False Negatives (TN/FN), also determined by visual inspection .

II. EXPERIMENTAL ASSESSMENT AND RESULTS

A new dataset of 642 underwater images obtained by us in the Mediterranean has been created recently to perform the assessment presented here. The dataset has been organised in order to minimize/avoid the overfitting in the NetVLAD training/testing. Images were separated in 75 queries and another database of 567 candidates. Images were taken with 3 different bottom looking cameras (a GoPro, a Point Grey CM3-U3-31S4 and a Manta G283), by divers and using an AUV,

model SPARUS II, in 5 different points of the coast of Mallorca, in different environments with different benthic habitats, forming different types of sea bottoms. Three different datasets were formed with all these images, named DS1, DS2 and DS3. Each one had 25 query images, and 183, 177 and 207 database images, respectively, in such a way that, every query has, at least, one loop closing in the corresponding database. The three datasets together with the Matlab files that contain their structure and contents, and the NetVLAD configuration files are available and accessible for the community in a corporative github repository [10]. NetVLAD was trained using 6 different configurations: two different networks, AlexNet and VGG-16, and for each network, the last convolutional layer was cropped with 3 different options: a max pooling layer, an average layer and the NetVLAD layer. The 6 configurations were applied on the 3 different datasets, in total 18 tests. According to [7], the trained models that presented the highest *Recall* metrics were selected for the final loop closure detection assessment: a CNN using Alexnet with the NetVLAD layer, called *Caffe_VLAD,* and another using VGG16 with the average pooling layer, called *Vd16_AVG. Caffe_VLAD* and *Vd16_AVG* were used to hash all images of DS1/DS2, and DS2/DS3, respectively. All images of all datasets were also hashed using HALOC. Afterwards, each query of each dataset was associated with 5 top loop closing candidates taken from the corresponding database. The 5 candidates were those that presented the lowest difference (L1-norm) between their hash and the hash of the query, and all were confirmed/rejected using RANSAC. The number of TP, TN, FP and FN were obtained by means of visual inspection. Table I shows a summary of the tests performed with HALOC and NetVLAD, and the obtained results in terms of Accuracy (A), Recall (R) and Fall-out (FO).

Dataset	**A**	**R**	**FO**		**Training Dataset**	**Testing Dataset**	**Model**	**A**	**R**	**FO**
DS1	0,92	0,86	0,08		DS3	DS1	*Caffe_VLAD*	0,76	0,79	0,28
DS2	0,9	0,88	0,09		DS3	DS2	*Caffe_VLAD*	0,82	0,93	0,22
DS3	0,86	0,71	0,12		DS1	DS2	*Vd16_AVG*	0,82	0,87	0,22
Mean	0,89	0,82	0,09		DS1	DS3	*Vd16_AVG*	0,87	1	0,20
Mean	--	--	--		--	--	--	0,82	0,89	0,23

Table 1. *Testing results using HALOC (left) and NetVLAD (right)*

III. CONCLUSIONS

The results of the assessment of the hash-based loop closing detection methodology using HALOC and NetVLAD, revealed a certain better performance of HALOC. HALOC gives, in average, a slightly higher Accuracy and Recall, and a clearly lower Fall-out. However, the NetVLAD results should not be underestimated because differences in Accuracies and Recalls are very small. The problem is the Fall-out: higher Fall-outs imply more FP. The merit of the NetVLAD results is in that they were obtained training a maximum of 642 images, in front of the thousands used to train the urban models described in [7]. Ongoing work includes trying to reduce the NetVLAD Fall-out with more training images since the objective would be using learned descriptors in order to increase the adaptability of the loop-closing detection processes in different environments.

ACKNOWLEDGMENTS

This work is partially supported by Ministry of Economy and Competitiveness under contract DPI2017-86372-C3-3-R.

REFERENCES

[1] Angeli, Adrien & Doncieux, Stéphane & Meyer, J.-A & Filliat, David. (2008). Real-Time Visual Loop-Closure Detection. Proceedings - IEEE International Conference on Robotics and Automation. 1842 - 1847. 10.1109/ROBOT.2008.4543475.
[2] M. Labbé and F. Michaud. 2018. Long-term Online Multi-session Graph-based SPLAM with Memory Management. Autonomous Robots 42, 6 (Aug. 2018).
[3] V. Monga and B. L. Evans. Perceptual Image Hashing Via Feature Points: Performance Evaluation and Tradeoffs. IEEE Transactions on Image Processing, 15(11):3452–3465, 2006.
[4] U Jain, V.P. Namboodiri, and G. Pandey. Compact Environment-Invariant Codes for Robust Visual Place Recognition. In 14th Conference on Computer and Robot Vision (CRV), pages 40–47, 2017.
[5] Laboratory UJI-IRSLab (Interactive, Robotic Systems), University of Girona (CIRS), and Balearic Islands University (UIB-SRV Group). TWIN roBOTs for Cooperative Underwater Intervention Missions, 2019. http://www.irs.uji.es/twinbot/twinbot.html.
[6] P.Ll. Negre Carrasco, F. Bonin-Font, and G. Oliver-Codina. 2016. Global Image Signature for Visual Loop-Closure Detection. Autonomous Robots 40 (2016), 1403–1417.
[7] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. 2018. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 40, 6 (June 2018), 1437–1451.
[8] P.Ll. Negre Carrasco, F. Bonin-Font, and G. Oliver Codina. 2016. Cluster-Based Loop Closing Detection for Underwater SLAM in Feature-Poor Regions. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). 2589–2595.
[9] A. Burguera, F. Bonin-Font, and G. Oliver. 2015. Trajectory-based Visual Localization in Underwater Surveying Missions. Sensors (Switzerland) 15, 1 (2015), 1708–1735.
[10] F. Bonin-Font. Underwater Dataset for CNN-based Loop Closing Detection, 2019. https://github.com/srv/Underwater-Dataset.