The Aqualoc Dataset: Towards Real-Time Underwater Localization from a Visual-Inertial-Pressure Acquisition System

Maxime Ferrera[1,2], Julien Moras[1], Pauline Trouvé-Peloux[1], Vincent Creuze[2], Denis Dégez[2]

I. INTRODUCTION

Robots localization is a challenging task which is essential for any autonomous robot or remotely operated one. In outdoor, GNSS signals can be used to get a quite accurate estimate of the position. However, in GNSS denied environment, such as indoor or underwater, localization needs to be estimated from other sensors. In aerial and terrestrial robotics, monocular VO (*Visual Odometry*), VSLAM (*Visual Simultaneous Localization And Mapping*) and, more recently, VI-SLAM (*Visual-Inertial SLAM*) have shown great results [1], [2], [3], [4]. Filtered SLAM based on Extended Kalman Filters (EKF) have been set aside to the profit of Bundle Adjustment (BA) based SLAM [5]. These SLAM methods, hugely relying on nonlinear optimization, have been successfully used to estimate localization from low-cost sensors such as cameras and MEMS-IMU (Micro-Electro-Mechanical System - Inertial Measurement Unit). The availability of many public terrestrial or aerial datasets have helped a lot in the quick development of these methods. For example, KITTI [6], EuRoC [7] and Mono-VI TUM [8] are famous datasets dedicated to the study of VSLAM or VI-SLAM algorithms.

In this paper, we focus on vision-based SLAM methods in underwater environments. The use of visual information is challenging underwater as the medium creates many visual degradations such as turbidity, back-scattering and light absorption. These difficulties led most works to turn to sonar based SLAM systems [9], [10]. Nevertheless, some works have also demonstrated the potential of cameras for underwater localization [11], [12]. However, most of these SLAM methods rely on the integration of expensive navigational sensors (Doppler Velocity Logs, Fibber Optic Gyroscopes or high-end IMUs) to provide accurate enough localization information, using only sonars or cameras to bound the drift [13], [14].

We believe that underwater localization from low-cost sensors processed through BA based SLAM could open the way to new localization techniques. However, the lack of underwater datasets is a limitation to the development of such algorithms. The only underwater dataset with visual information allowing the use of VSLAM methods we are aware of is [15] but it consists only of simulated images. Hence, in order to open the way to deeper studies of these SLAM techniques, we propose a new underwater dataset that we make publicly available online[1]. This dataset has been recorded in an harbor and provides several sequences with synchronized measurements from a monocular camera, a MEMS-IMU and a pressure sensor. To the best of our knowledge, this is the first underwater dataset dedicated to the study of underwater localization methods from low-cost sensors.

The rest of this paper is organized as follow. First, we present the payload designed for this acquisition. Then, we give details about the recorded dataset. Finally, we present results of an evaluation of state-of-the-art open-source monocular VO and VSLAM which can be used as a benchmark on this dataset and highlights the potential of such vision based localization methods.

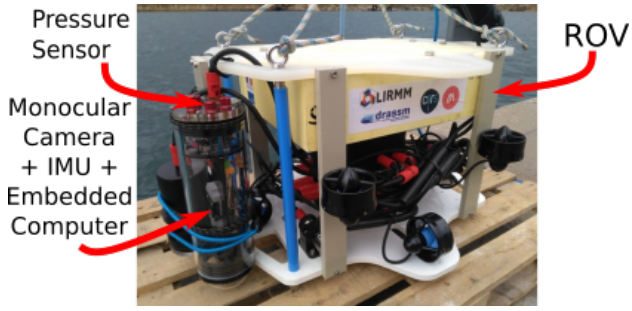


Fig. 1. The Remotely Operated Vehicle (ROV) and acquisition system used to record the dataset.

II. ACQUISITION SYSTEM

The acquisition system that we designed to record the dataset can be seen in figure 2. It consists of a watertight enclosure containing a monochromatic camera, a pressure sensor and an Nvidia Tegra Jeston TX2 module mounted on Auvidea J120-IMU carrier board. Details are given in table I. The monochromatic camera is equipped with a wide-angle lens and records images with a resolution of 640x512 pixels at 20 Hz. The camera is set behind a dome end cap in order to reduce distortion from the passing of light rays through different media. The IMU provides linear acceleration and angular velocity measurements at 200 Hz, as well as compass measurements at 80 Hz and the pressure sensor between 5 and 10 Hz. The computing module is running Ubuntu 16.04

This work was partially supported by the CNRS (Mission pour l'interdisciplinarité - Instrumentation aux limites 2018 - Aqualoc project) and Région Occitanie (ARPE Pilotplus project).

[1]M. Ferrera, J. Moras and P. Trouvé-Peloux are with DTIS, ONERA, Université Paris Saclay F-91123 Palaiseau - France.

[2]M. Ferrera and V. Creuze are with LIRMM, Université de Montpellier, CNRS, Montpellier, France.

[3]D. Dégez is with the DRASSM, Ministry of Culture, Marseilles, France

[1]http://www.lirmm.fr/aqualoc/

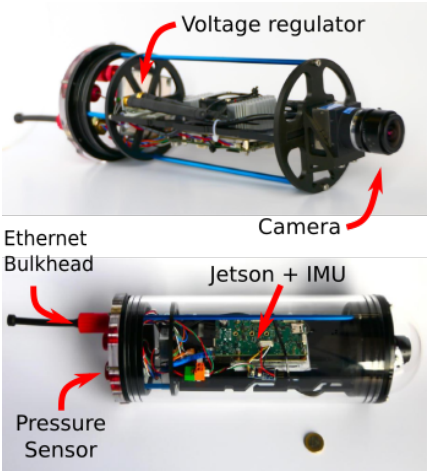


Fig. 2. The acquisition system integrating a monocular monochromatic camera, a pressure sensor and an IMU.

	Specifications	Values	Description
Vehicle	Size	65x60x55 cm	
	Mass	28 Kg	
	Thrusters	8	Blue Robotics T200
Payload	**Camera sensor**		uEye - UI-1240SE
	Resolution	640x512 px	
	Sensor	Monochromatic	
	Frame per second	20 Hz	
	Lens		Kowa LM4NCL C-Mount
	Focal Length	3.5mm	
	Field of View	131deg	
	Inertial Measurement Unit		MEMS - MPU-9250
	Gyroscope frequency	200 Hz	
	Accelerometer frequency	200 Hz	
	Compass	80 Hz	
	Pressure Sensor		MS5837 - 30BA
	Depth Range	0 - 300m	
	Resolution	0.2 mbar	
	Frequency	5-10 Hz	
	Embedded Computer		
	Computing Unit		Nvidia - Tegra Jetson TX2
	Carrier board		Auvidea J120 - IMU
	Housing		
	Enclosure	33.4 x 11.4 cm	4" Blue Robotics Watertight Enclosure
	Enclosure End Cap	Dome	4" Blue Robotics Dome End Cap

TABLE I

SUMMARY OF THE FEATURES OF THE ROV AND THE PAYLOAD.

and records synchronously the different sensors measurements thanks to the ROS middleware. An advantage of our payload is that it is independent of any robotic architecture and can thus be embedded on any kind of Remotely Operated Vehicle (ROV) or Autonomous Underwater Vehicle (AUV). This contrasts with classical underwater robots instrumentation load which are usually integrated in the design of the robots [16].

III. DATASET

In order to record the dataset, the acquisition system is set to face downward on the ROV as shown in figure 1. An aprilgrid is used both for the calibration of the system and as a marker for navigation. In fact, in each sequence, the ROV starts from the aprilgrid (visible), navigates away from it (no-more visible), and finally comes back to it (visible again). Intrinsic calibration of the camera is done in situ while calibration of the extrinsic parameters between the IMU and the camera is performed in-air in order to make fast motions, required to estimate these parameters. These calibration steps are computed using Kalibr [17]. As the camera is equipped with a wide-angle camera, we used the equidistant distortion

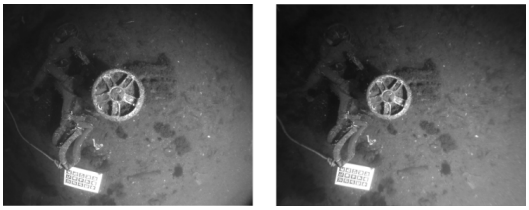


Fig. 3. Equidistant distortion effects and removal from Kalibr calibration. Left: raw image. Right: undistorted image.

model of Kalibr. Results of the calibration can be seen figure 3.

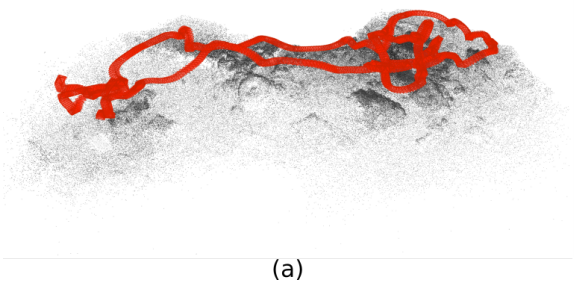
All the measurements have been recorded with the ROS middleware and they are hence all synchronized in ROS bags. In the online repository of the dataset, we provide both the data in a ROS bag format and as plain files. For the plain files, the IMU, pressure and compass measurements are given in independent files with a timestamp linked to every measure. The distorted and undistorted version of the images are also given as *png* images and the timestamp of each image is written in a separate file.

A total of seven sequences have been recorded using this setup. The sequences were recorded in an harbor in collaboration with the DRASSM[2]. The sequences expose different levels of difficulty with sometimes parts where vision becomes unusable. Details about each sequence are given in Fig.4b. Obtaining a ground-truth is very difficult in underwater environments and often requires the use of external infrastructures. However, Structure-from-Motion methods are able to compute very accurate 3D reconstruction from sequence of images by means of extensive full batch Bundle Adjustment run offline. In order to compute such a reconstruction we have used the state-of-the-art library Colmap [18]. Setting very low the features detection threshold, Colmap has been able to produce very accurate reconstruction (Fig.4a). As the poses of every image is computed by Colmap in order to produce the 3D reconstructions, we have extracted the estimated cameras' trajectories to use it as a ground-truth. We further scaled these trajectories using depth measurements to get metric trajectories.

IV. BENCHMARK

As a benchmark, we have run experiments using state-of-the-art monocular VO and VSLAM algorithms on each sequences. We compare ORB-SLAM [1], SVO-2 [2] and DSO [19]. ORB-SLAM is a feature-based SLAM algorithm extracting ORB [20] features for tracking and for loop closure detections. It makes use of keyframes, connected to each others from their feature matches, and efficiently optimize the built graph through Bundle Adjustment. SVO-2 is a semi-direct sparse odometry method. Direct methods do not rely on matched features as a mean of tracking and pose estimation but instead directly track photometric patches in the images. SVO-2 is semi-direct in the sense that FAST features [21] are detected in each new keyframe and then image-alignment is performed using photometric patches around the features to estimate the pose. DSO is a pure

[2]DRASSM: French Department of Underwater Archaeological Research



(a)

Sequence	Duration	Length	Specifications	Absolute Translation Error (m)		
				ORB-SLAM	DSO	SVO-2
#1	3'49"	52.6m	Some back-scattering	0.52	X	**0.49**
#2	6'47"	74.7m	Some back-scattering	**0.50**	0.63	0.56
#3	4'17"	23.6m	-	0.45	***0.25***[1]	*0.26*[1]
#4	3'26"	51.1m	Low vision parts	X	X	X
#5	2'52"	28.5m	Some back-scattering and low vision parts	**0.24**	0.67	X
#6	2'06"	19.5m	-	0.51	0.24	**0.02**
#7	1'53"	32.9m	Low vision parts	X	X	X

TABLE II

(b)

Fig. 4. (a) Sequence #3 Colmap 3D reconstruction. (b) Evaluation of state-of-the-arts monocular VSLAM methods on the recorded dataset *(1: Results obtained starting the sequence 15s later otherwise the methods failed to initialize).*

direct odometry algorithm. It performs pose estimation from the minimization of photometric errors across several images by tracking photometric patches of areas with high intensity gradients. As the tracking is direct, the algorithm takes into account illumination changes in the energy minimization step, thus relaxing the constant brightness assumption of many direct methods.

The results for each method are given in Fig.4b. As any monocular systems, trajectories are estimated up to scale. In order to compute an absolute translation error for every sequence, the monocular trajectories are first scaled by applying the similarity transformation that best fit the ground-truth. ORB-SLAM is the most stable algorithm on this dataset as it manages to run on five out of the seven sequences while DSO and SVO-2 run only on four sequences each. Interestingly, ORB-SLAM does not manage to detect any loop closure in the sequences, thus asking the question of the relevance of its loop detection mechanism for underwater environments.

These results highlight the potential of vision based localization methods for underwater environments. While not all the sequences can be processed, the result are promising. Both the direct and feature-based paradigms seem to work but their seem to be a compromise between stability and accuracy. A feature-based SLAM algorithm combined to a direct tracking of the features could be a good balance for underwater localization.

V. Conclusion

In this paper, we have presented a new underwater dataset focusing on vision-based localization methods and including IMU and pressure measurements. We made publicly available this dataset to the benefit of the community. The designed acquisition system as well as the recorded sequences were then presented in detail. Finally, we gave the results in terms of localization drift for VO and VSLAM state-of-the-art methods on each sequence. These results gave a short overview of what can be achieved using pure visual information and could be used as a benchmark on this dataset. There is room for improvement as the evaluated methods were originally developed for terrestrial and aerial applications. Therefore, methods specifically dedicated to underwater visual degradations and the integration of the IMU and pressure measurements are expected to give very interesting results. In future work, we will investigate the use of such SLAM methods.

References

[1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE T-RO*, 2015.

[2] C. Forster, Z. Zhang, M. Gassner, and et al., "SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems," *IEEE T-RO*, April 2017.

[3] S. Leutenegger, S. Lynen, M. Bosse, and et al., "Keyframe-Based Visual-Inertial Odometry using Nonlinear Optimization," *IJRR*, 2015.

[4] Y. Lin, F. Gao, T. Qin, W. Gao, and et al., "Autonomous aerial navigation using monocular visual-inertial fusion," *JFR*, 2018.

[5] H. Strasdat, J. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Image and Vision Computing*, 2012.

[6] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *CVPR*, 2012.

[7] M. Burri, J. Nikolic, P. Gohl, T. Schneider, and et al., "The EuRoC micro aerial vehicle datasets," *IJRR*, 2016.

[8] D. Schubert, T. Goll, N. Demmel, V. Usenko, and et al., "The TUM VI Benchmark for Evaluating Visual-Inertial Odometry," in *IROS*, 2018.

[9] A. Mallios, P. Ridao, D. Ribas, F. Maurelli, and Y. Petillot, "EKF-SLAM for AUV navigation under probabilistic sonar scan-matching," in *IROS*, 2010.

[10] D. Ribas, P. Ridao, J. Tardos, and J. Neira, "Underwater SLAM in man-made structured environments," *JFR*, 2008.

[11] P. Carrasco, F. Bonin-Font, M. M. Campos, and G. Codina, "Stereo-Vision Graph-SLAM for Robust Navigation of the AUV SPARUS II," *IFAC-PapersOnLine*, 2015.

[12] V. Creuze, "Monocular Odometry for Underwater Vehicles with Online Estimation of the Scale Factor," in *IFAC World Congress*, 2017.

[13] R. M. Eustice, O. Pizarro, and H. Singh, "Visually Augmented Navigation for Autonomous Underwater Vehicles," *IEEE Journal of Oceanic Engineering*, 2008.

[14] N. Palomeras, N. Hurts, M. Carreras, and P. Ridao, "Autonomous Mapping of Underwater 3-D Structures: From View Planning To Execution," *IEEE RAL*, 2018.

[15] A. C. Duarte, G. B. Zaffari, R. T. S. da Rosa, L. M. Longaray, and et al., "Towards comparison of underwater SLAM methods: An open dataset collection," in *MTS/IEEE OCEANS - Monterey*, 2016.

[16] M. Carreras, J. D. Hernandez, E. Vidal, N. Palomeras, D. Ribas, and P. Ridao, "Sparus II AUV - A Hovering Vehicle for Seabed Inspection," *IEEE Journal of Oceanic Engineering*, 2018.

[17] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *IROS*, 2013.

[18] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[19] J. Engel, V. Koltun, and D. Cremers, "Direct Sparse Odometry," *IEEE T-PAMI*, 2018.

[20] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *ICCV*, 2011.

[21] E. Rosten, R. Porter, and T. Drummond, "Faster and Better: A Machine Learning Approach to Corner Detection," *IEEE T-PAMI*, 2010.