Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-31 (2022) 369–374

Visual-based Navigation Strategy for Autonomous Underwater Vehicles in Monitoring Scenarios

F. Ruscio *,** **S. Tani** *,** **M. Bresciani** *,** **A. Caiti** *,**,*** **R. Costanzi** *,**,***

* *Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Pisa, Italy*

** *Interuniversity Center of Integrated Systems for the Marine Environment (ISME), Italy*

*** *Centro di Ricerca "E. Piaggio", Università di Pisa, Italy*

Abstract: Autonomous Underwater Vehicles (AUVs) performing visual surveys aimed at the preservation of marine environments are equipped with optical sensors for image acquisition. In addition, an altitude sensor is usually installed on-board to control the distance from the seabed and avoid possible collisions. Within this context, this work proposes a navigation strategy for underwater monitoring scenarios, which fuses a single bottom-looking camera and altitude information for linear velocity estimation. This allows to exploit the payload already required by monitoring activities also for navigation purposes, thus reducing the number of sensors on-board the AUV. The linear velocity is provided by a monocular Visual Odometry (VO) technique that switches between homography and epipolar models for motion estimation and leverages altitude measurements to overcome the scale ambiguity issue. The navigation framework relies on an Extended Kalman Filter (EKF) that combines visual-based linear velocity with attitude and depth measurements for trajectory estimation. The proposed strategy has been tested on real data acquired by using Zeno AUV, equipped with bottom-looking camera, DVL, Attitude and Heading Reference System (AHRS), and depth sensor. The performance has been assessed comparing the estimated linear velocities with the DVL readings, and the VO-based estimated trajectory with that provided by a DVL-based dead-reckoning approach, yielding to a maximum absolute error of 2.16m for a reference trajectory of 166m. Given the promising results, this strategy could represent an affordable solution for underwater navigation where visibility conditions allow the use of optical sensors.

Copyright © 2022 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Autonomous underwater vehicles, Visual odometry, Estimation and filtering, Robot Navigation, Programming and Vision, Perception and sensing

1. INTRODUCTION

Autonomous Underwater Vehicles (AUVs) are expected to play a crucial role in the preservation of marine environments (Di Ciaccio and Troisi (2021)). Thanks to the presence of on-board cameras, AUVs are capable of collecting data that can be useful for monitoring and inspection applications (Ruscio et al. (2021)). The successful outcome of the mission task mostly relies on the accuracy of underwater navigation. Traditionally, this is achieved using dead-reckoning approaches based on expensive Doppler Velocity Log (DVL) sonars combined with Inertial Navigation Systems (INS) and depth sensors (Panish and Taylor (2011)).

As a result of the benefits provided by visual sensors in terms of compact size, low cost, and abundant data information, visual-based techniques have been investigated for estimating the motion of robots and autonomous vehicles (Bonin-Font et al. (2008)). However, underwater environment poses severe limitations to imaging conditions, mainly due to rapid light attenuation, haze, turbidity, and backscattering (Lu et al. (2017)). This is the reason why many state-of-the-art visual-based navigation algorithms commonly used in aerial and terrestrial domains do not perform satisfactory underwater (Joshi et al. (2019)).

Most approaches tailored to the underwater domain belong to the Visual Simultaneous Localisation And Mapping (VSLAM) paradigm, whose goal is not only to estimate the motion of the vehicle, but also to reconstruct a map of the surrounding environment (Salvi et al. (2008), Kim and Eustice (2015), Ferrera et al. (2019)). On the contrary, Visual Odometry (VO) techniques focus solely on frame-to-frame motion estimation, without any map generation (Scaramuzza and Fraundorfer (2011)). Although this does not guarantee a global consistency of the motion, it heavily reduces complexity and computational resources. Monocular VO strategies allow to achieve motion estimation using a single camera.However, they suffer from the scale ambiguity problem, meaning that motion can only be recovered up to a scale factor. In the under-

⋆ Work partially supported by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project.

2405-8963 Copyright © 2022 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2022.10.456

water domain, the problem is usually tackled leveraging additional information coming from extra sensors, such as depth measurements (Creuze (2017)) or acoustic ranges (Roznere and Li (2020)). To support visual-based ego-motion estimation, many underwater solutions investigate the use of VO approaches aided by other sensors. Shkurti et al. (2011) proposed a navigation algorithm based on the Multi-State Constrained Kalman Filter combining information from monocular camera, inertial, and pressure sensors. In Bucci et al. (2022) an Unscented Kalman Filter fuses compass, inertial and fiber optic gyroscope measurements with visual and altitude information for vehicle navigation. Within the context of underwater monitoring scenarios involving the visual inspection of marine environments, this paper proposes a visual-based navigation strategy combining a single bottom-looking camera and altitude information for AUV linear velocity estimation. In particular, compared to the other presented solutions, the main contributions of this work are: (i) the development of a monocular VO approach, available online [1], which alternately exploits homography and epipolar models for motion estimation in dynamic environments; (ii) the proposal of a cost effective navigation strategy leveraging sensors typically installed on-board AUVs in underwater monitoring contexts; (iii) the quantitative evaluation of the proposed strategy in real experiments using DVL measurements as benchmark.

The monocular VO approach exploits feature detection and matching in consecutive image pairs and makes use of a 2D-to-2D method for relative motion estimation. This is computed switching between homography and epipolar geometry-based model so as to effectively provide a motion estimate in different conditions. The estimated relative motion is used to triangulate the matched features, thus obtaining a 3D point cloud. The scene distance information available from an altitude sensor is then compared to the average depth of the triangulated points to solve the scale ambiguity issue. Finally, an estimate of the AUV trajectory is obtained by fusing the resulting linear velocity with attitude and depth measurements within an Extended Kalman Filter (EKF). The solution allows to reduce the cost as well as the complexity of the system since the sensors already installed on-board the vehicle to carry out the specific monitoring activities are also used for navigation purposes. Indeed, such activities require good visibility conditions that justify the employment of optical sensors. Moreover, the AUV is usually equipped with an altitude sensor which is used to control the distance from the seabed, thus ensuring the collection of meaningful images while avoiding possible damaging impacts.

The proposed strategy has been tested on a real dataset collected in March 2021 in Rapallo (Italy), during a monitoring mission over an area characterised by the presence of Posidonia oceanica (Po) meadows. The dataset was acquired using Zeno AUV, which is part of the equipment of the CrossLab of the Dipartimento di Ingegneria dell'Informazione - Università di Pisa. The AUV, shown in Fig. 1, was equipped with a DVL, an Attitude and Heading Reference System (AHRS), and a depth sensor to perform the underwater navigation. Moreover, a bottom-looking camera was mounted on-board for image acquisition. For the specific dataset, the altitude information needed to solve the scale ambiguity issue is provided by the DVL, which also gives distance measurements from the seabed. Taking DVL as benchmark for accurate underwater navigation, the performance of the proposed approach is evaluated by comparing the estimated linear velocity and trajectory with the DVL readings and the DVL-based dead-reckoning trajectory, respectively.

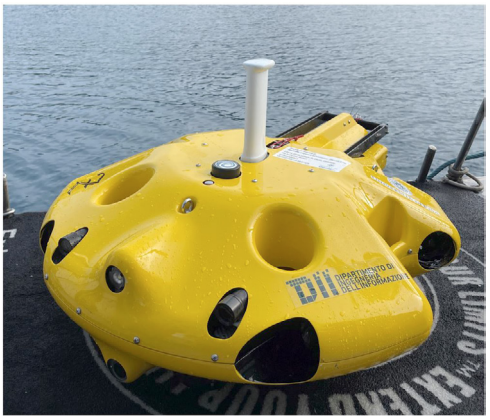

Fig. 1. Zeno AUV.

The rest of the paper is organised as follows: Sec. 2 describes the proposed visual-based underwater navigation approach. In particular, Subsec. 2.1 details the monocular VO algorithm exploited to compute the linear velocity from consecutive image pairs, while Subsec. 2.2 describes the navigation algorithm adopted to estimate the trajectory of the vehicle. The experimental activities and the corresponding results used to assess the performance of the system are reported in Sec. 3. Finally, conclusions and possible future works are summarised in Sec. 4.

2. VISUAL-BASED NAVIGATION SYSTEM

2.1 Linear velocity estimation

In typical underwater navigation, dead-reckoning strategies take DVL readings to determine the AUV position in time. Conversely, the proposed strategy replaces DVL measurements with a monocular VO algorithm, which retrieves the AUV linear velocity by estimating the motion between consecutive image pairs I_i, I_{i+1} grabbed by a single camera. The VO algorithm relies on a feature-based approach, making use of feature correspondences in 2D image plane for motion estimation. The VO workflow is schematically represented in green in Fig. 2 and is composed of the following phases:

1. *Preprocessing.* Each image is initially processed to account for both refraction effects and lens distortion. This is accomplished using distortion parameters estimated during a camera calibration process. Moreover, since underwater images have low contrast and are low textured, an histogram equalisation has been applied to the image sequence to improve the detection of salient points. In particular, each undistorted RGB image is converted to grayscale and then processed by using the Contrast Limited Adaptive Histogram Equalisation (CLAHE) method (Reza (2004)), which

[1] https://github.com/team-ergo-unipi/ergo_uvo

has been shown to be effective in enhancing underwater images (Hasibuan et al. (2021)).

2. *Feature Detection & Matching.* The identification of repeatable and distinguishable salient points in image pairs is of utmost importance for the correct estimation of motion. This is even more crucial underwater, due to the high variability of scenarios and the degradation of images. To address this problem, Speeded Up Robust Feature (SURF) algorithm (Bay et al. (2008)) has been chosen for its robustness with respect to image rotation, translation, and scaling. Moreover it is partially robust to affine 3D projections and illumination changes, while ensuring a limited computational cost for real-time applications. SURF is applied independently to each pair of images to extract robust features, which are then compared with each other to find potential matches. During this phase, false matches can occur which may negatively affect motion estimation: therefore, they must be removed to achieve good results. For this reason, the distance ratio test proposed by Lowe (2004) is applied to the matched feature pairs to filter ambiguous correspondences. Therefore, only those that satisfy the Lowe's ratio test are retained for the motion estimation phase.

3. *Motion Estimation.* The relative camera motion between two consecutive images comes from a 2D-to-2D model fitting framework. In particular, the geometric model employed within the VO strategy exploits either essential or homography matrix, depending on the specific geometry conditions. Essential matrix derives from epipolar constraints (Longuet-Higgins (1981)), while homography matrix results from planar assumption (Hartley and Zisserman (2003)). Both geometric models have been adopted within monocular VO approaches to estimate the motion of a camera. In details, homography is more suitable for planar scenes and when the point of view remains unchanged, as in pure rotational motion. This is due to the fact that co-planar points and low parallax views do not provide enough constraints to determine the epipolar geometry. On the contrary, in the presence of a structured scene, homography is prone to fail. In the proposed approach, the algorithm switches between the two models with the aim of adapting to different scenarios and camera motions. The validity of the selected geometry model is assessed at each iteration by means of a cheirality check, which verifies if the triangulated points are located in front of both camera poses. In the event that the percentage of 3D points satisfying the cheirality check is less than a predefined threshold, the selected model is discarded and the other one is used thereafter. This strategy assumes that as soon as one model fails, the scenario is changing and the other approach better encodes the geometric transformation between image pairs. Despite the use of Lowe's ratio test, false matches can still be present. Consequently, both models are estimated using RANSAC regression algorithm (Fischler and Bolles (1981)), which allows to cope with motion estimation in presence of outliers. However, it may happen that neither model satisfies the check. In such circumstances, the algorithm is unable to provide an adequate estimate of the relative motion between the two current images and the process starts again as soon as a new image is captured by the camera.

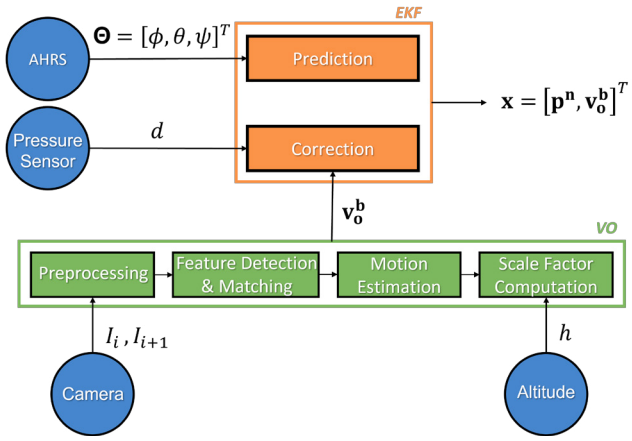

Fig. 2. Navigation filter flowchart.

4. *Scale Factor Computation.* Since the approach exploits a monocular strategy, camera motion can only be recovered up to a scale factor. This means that at this level only the direction of motion is known and relative translation between image pairs has to be appropriately scaled to be on a metric scale. In the proposed approach, scale factor is retrieved integrating additional information about the distance of the camera with respect to the scene. More specifically, scale ambiguity is solved comparing altitude readings with the 3D scene reconstructed by triangulating corresponding features from two successive views. This solution is based on the assumption that the portion of the area captured by the camera coincides with that relative to the range measurement. This way, the mean value of the triangulated points along the optical axis Z_{mean} is used to represent the up-to-scale distance between the camera and the scene. Therefore, the ratio between range measurement h, in meters, and the distance perceived by the camera corresponds to the unknown scale factor λ, which is finally used to restore the metric scale of relative translation.

$$\lambda = h/Z_{mean} \tag{1}$$

2.2 Navigation algorithm

The linear velocity provided by the monocular VO algorithm is exploited within an Extended Kalman Filter (EKF) for navigation purposes. The filter is in charge of refining the linear velocity updates coming from the VO system as well as providing an estimate of the AUV trajectory. To do so, the filter combines the linear velocity data with attitude and depth information provided by an AHRS and a pressure sensor, respectively. Following the *SNAME* notation (Fossen (2002)), the filter state has been represented as $\mathbf{x} = [\mathbf{p^n}, \mathbf{v_o^b}]^T \in \mathbb{R}^6$, where $\mathbf{p^n} = [n, e, d]^T \in \mathbb{R}^3$ contains the position of the AUV along the North, East and Down axes of the NED reference frame and $\mathbf{v_o^b} = [u, v, w]^T \in \mathbb{R}^3$ represents the body-fixed linear velocities with respect to the surge, sway and heave axes.

Fig. 3. Images from the reference mission.

The prediction phase of the EKF exploits a kinematic model, according to which the evolution of the AUV's position in discrete time can be formalized as:

$$\mathbf{p^n}_{k+1} = \mathbf{p^n}_k + dT\mathbf{R^n_b}(\boldsymbol{\Theta}_k)\mathbf{v^b_o}_k \tag{2}$$

where $\boldsymbol{\Theta} = [\phi, \theta, \psi]^T \in \mathbb{R}^3$ contains the Euler angles representing the attitude of the vehicle, which are used within the rotation matrix $\mathbf{R^n_b}(\boldsymbol{\Theta}_k)$ to decompose the body-fixed velocities in the NED reference frame, whilst dT represents the discretisation time. On the other hand, the evolution of the body-fixed velocities has been modeled as white Gaussian process noises:

$$\mathbf{v^b_o}_{k+1} = \mathbf{v^b_o}_k + dT\mathbf{W}_k \tag{3}$$

where $\mathbf{W}_k \in \mathbb{R}^3$ represents the white Gaussian noise vector.

For the correction step, the EKF makes use of the depth measurements provided by the pressure sensor and the body-fixed linear velocity updates coming from the monocular VO algorithm. It is important to point out that the linear velocity vector provided by the VO approach is expressed in camera-fixed frame and hence needs to be decomposed in the body-fixed reference frame prior to being used within the EKF. The overall visual-based navigation approach is depicted in Fig. 2.

3. EXPERIMENTAL RESULTS

The developed visual-based underwater navigation strategy has been tested on a real dataset collected during at-sea experiments by using Zeno AUV. The proposed approach has been executed in post-processing by reproducing the recorded data as if they were elaborated in real-time. In this way it has been possible to evaluate the feasibility of the algorithm for future online applications. For the implementation of the VO algorithm, the C++/OpenCV libraries (OpenCV (2015)) have been adopted, using a value of 0.7 for the Lowe's ratio test and 0.8 as threshold for the cheirality check.

3.1 System setup

The vehicle was equipped with a DVL, an AHRS, and a pressure sensor providing linear velocity, attitude, and depth measurements, respectively. Such information was used by the AUV within a dead-reckoning framework to perform the underwater navigation task during the experimental activities. The vehicle was also provided with a bottom looking camera to collect images of the seabed. Table 1 reports the main specifications of the DVL sensor together with the main features of the camera as they were configured during the experiments. Along with linear velocity, the DVL was also providing altitude measurements. Since DVL and camera sensors are mounted downward in the lower part of the vehicle, altitude readings are used as range information to solve the scale factor ambiguity in the VO strategy.

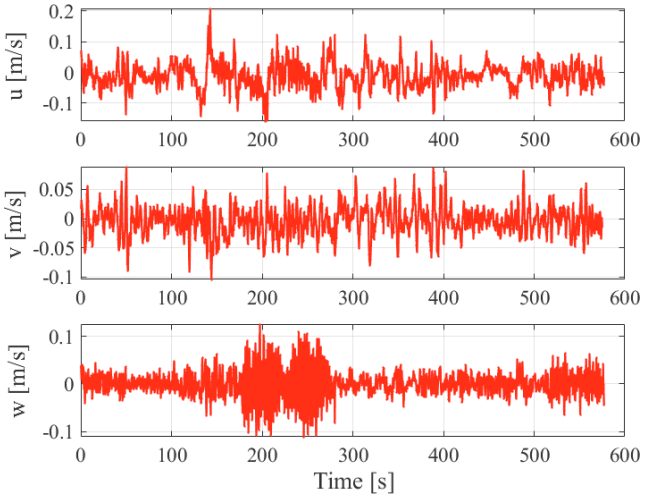

Fig. 4. Trend over time of EKF linear velocity errors using DVL reading as reference values.

Table 1. DVL and camera main specifications.

Nortek-DVL 1000		*Camera Basler acA2500-20gc*	
Spec.	Value	Spec.	Value
Accuracy	±0.1 [cm/s]	Sensor	CMOS
Altitude	0.2 - 75 [m]	Frame Rate	3 [fps]
Resolution	0.01 [mm/s]	Resolution	820x648 [px]

3.2 Experiment description

The dataset refers to an underwater survey carried out in Rapallo, Italy, in March 2021 in collaboration with the Agenzia Regionale per la Protezione dell'Ambiente Ligure (ARPAL). The survey was performed during environmental monitoring activities aimed at acquiring visual information of the coastal area colonised with Po. The desired trajectory was accomplished autonomously by the vehicle following a lawn mower pattern with a reference velocity of 0.4m/s and a reference altitude of 2.5m. Fig. 3 depicts 3 representative images of the dataset collected throughout the underwater mission. As can be seen from these images, the selected distance from the seabed and the good visibility conditions make the dataset suitable for the application and the evaluation of the proposed strategy. Furthermore, the diversity of the seabed, characterised by the coexistence of rocks, Po, and sand, allows to assess the effectiveness and robustness of the strategy to different underwater scenarios.

3.3 Results

The presented results refer to a portion of the entire mission, where the vehicle was navigating at the desired distance from the seabed. At this stage, the AUV performed a trajectory approximately 166m long with a duration of about 10 minutes. Fig. 4 and Fig. 5 report the performance of the visual-based navigation strategy in terms of estimated linear velocity considering the errors with regard to DVL measurements. In particular, Figure 4 shows the trend over time of linear velocity errors along

the three axes of the body-fixed frame, whilst Figure 5 depicts their Probability Density Function (PDF). The errors exhibit an oscillating trend characterised by mean value and distribution that are comparable along the 3 axes. These outcomes are summarised in Table 2, which reports Root Mean Square Error (RMSE), mean, and standard deviation of the EKF linear velocity errors. Table 2 also contains the error metrics computed for VO linear velocity to evaluate the improvement given by the filtering process. The surge axis is the most significant one, as it corresponds to the main direction of motion. For both methods, the RMSE and mean errors are in the order of centimetres per second.

Table 2. Linear velocity errors.

		u	v	w
RMSE $[m/s]$	VO	0.057	0.043	0.021
	EKF	0.042	0.026	0.027
Mean $[m/s]$	VO	-0.007	-0.004	0.006
	EKF	-0.011	-0.004	$-9.4e^{-5}$
Std. Dev. $[m/s]$	VO	0.059	0.043	0.029
	EKF	0.036	0.025	0.033

Fig. 6 shows the trajectory comparison in the North-East plane. The green line refers to the dead-reckoning strategy based on DVL measurements, which is taken as ground truth for the positions occupied by the AUV during the underwater survey. On the other hand, the red line represents the trajectory computed by the EKF according to the linear velocity estimates provided by the monocular VO algorithm. Such a comparison in the horizontal plane is significant because the result depends closely on the velocity estimates, whereas depth information is not reported since both methods make use of the same pressure sensor.

Fig. 7 shows the trend over time of the position errors in the North-East plane. Moreover, distance errors have been included to better understand the overall difference between the two trajectories. Table 3 encloses the performance of the proposed visual-based underwater navigation strategy, reporting RMSE and Maximum Absolute Error (MAE) of the estimated positions. The MAE is approximately 2.16m. Given that the reference survey is 166m long, the error corresponds to about 1.3% of the total length of the trajectory.

Table 3. Position errors.

	RMSE $[m]$	MAE $[m]$
North	0.717	1.263
East	0.348	1.256
Distance	1.013	2.163

4. CONCLUSIONS & FUTURE WORKS

This work proposes a visual-based underwater navigation strategy for AUVs for environmental inspection and monitoring contexts where good visibility conditions justify the use of optical sensors. A monocular VO strategy was developed to retrieve the AUV linear velocity using image sequence integrated with altitude measurements to solve the scale ambiguity. The visual-based linear velocity was then employed within an EKF fusing attitude and depth measurements to estimate the AUV trajectory. The strategy was tested on real data collected by Zeno AUV

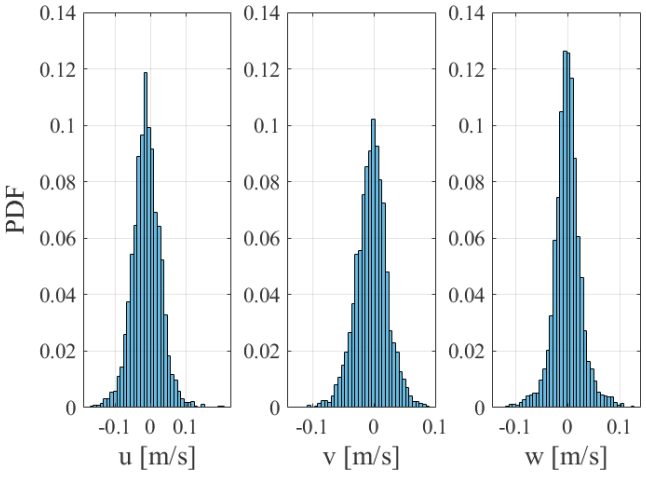

Fig. 5. Probability Density Function (PDF) of EKF linear velocity errors.

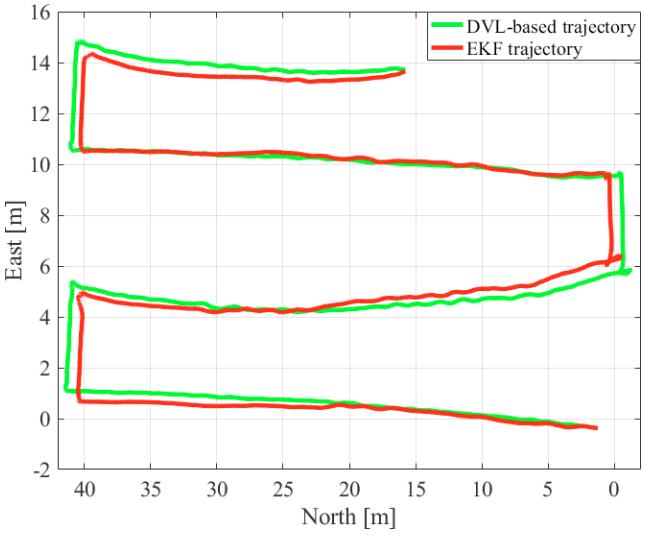

Fig. 6. North-East plane trajectories: in green the DVL-based trajectory, used as benchmark, and in red the trajectory provided by the EKF.

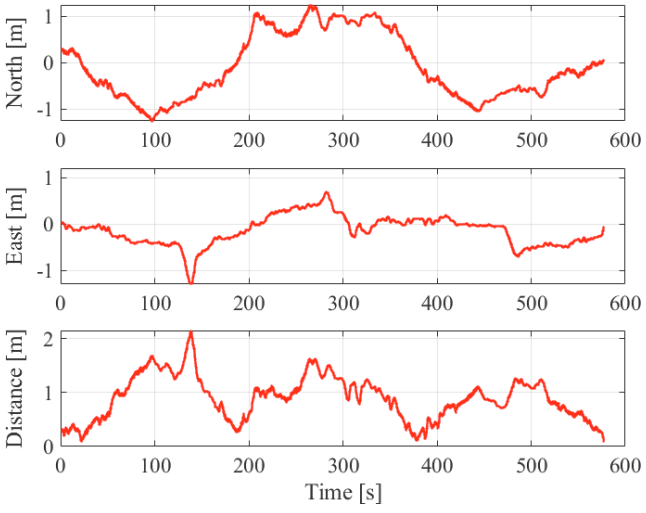

Fig. 7. Trend over time of estimated position errors along the North and East axes and the overall distance values.

during underwater monitoring activities carried out over an area colonised with Po. The results suggest that a filtering approach is necessary to reject outliers that can occur as output from the VO algorithm, thus bringing benefits to the velocity estimation. The filtering process allows to achieve velocity estimates that are consistent with the DVL readings. Moreover, the comparison of the estimated trajectory with that provided by the DVL-based dead-reckoning approach, used as benchmark, show that an accurate visual-based navigation system could be achieved. The obtained results are promising considering the specific test scenario. Indeed, the presence of Po makes VO applications particularly challenging, as the Po leaves tend to move under the force of sea currents. This goes against the assumption that the majority of the points in the environment are rigid, thus negatively affecting visual-based motion estimation. Nonetheless, the results show that the presented visual-based strategy is robust enough to deal with such an issue.

Possible future developments will investigate alternative strategies to solve the scale ambiguity. Moreover, the approach will be tested using different datasets to assess its robustness with respect to the high variability of underwater environments. This will also allow to evaluate the approach effectiveness in featureless sea bottom environment, which may jeopardise the accomplishment of the mission. Finally, optimisation techniques will be integrated in the VO framework to further improve the overall performance of motion estimation.

ACKNOWLEDGEMENTS

The authors would like to thank the members of ARPAL for their support during the missions at sea.

REFERENCES

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3), 346–359.

Bonin-Font, F., Ortiz, A., and Oliver, G. (2008). Visual navigation for mobile robots: A survey. *Journal of intelligent and robotic systems*, 53(3), 263–296.

Bucci, A., Zacchini, L., Franchi, M., Ridolfi, A., and Allotta, B. (2022). Comparison of feature detection and outlier removal strategies in a mono visual odometry algorithm for underwater navigation. *Applied Ocean Research*, 118, 102961.

Creuze, V. (2017). Monocular odometry for underwater vehicles with online estimation of the scale factor. In *IFAC 2017 World Congress*.

Di Ciaccio, F. and Troisi, S. (2021). Monitoring marine environments with autonomous underwater vehicles: a bibliometric analysis. *Results in Engineering*, 9, 100205.

Ferrera, M., Moras, J., Trouvé-Peloux, P., and Creuze, V. (2019). Real-time monocular visual odometry for turbid and dynamic underwater environments. *Sensors*, 19(3), 687.

Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Fossen, T.I. (2002). Marine control systems–guidance. navigation, and control of ships, rigs and underwater vehicles. *Marine Cybernetics, Trondheim, Norway, Org. Number NO 985 195 005 MVA, www. marinecybernetics. com, ISBN: 82 92356 00 2*.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

Hasibuan, Z., Andono, P., Pujiono, D., Setiadi, R., et al. (2021). Contrast limited adaptive histogram equalization for underwater image matching optimization use surf. In *Journal of Physics: Conference Series*, volume 1803, 012008. IOP Publishing.

Joshi, B., Rahman, S., Kalaitzakis, M., Cain, B., Johnson, J., Xanthidis, M., Karapetyan, N., Hernandez, A., Li, A.Q., Vitzilaios, N., et al. (2019). Experimental comparison of open source visual-inertial-based state estimation algorithms in the underwater domain. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7227–7233. IEEE.

Kim, A. and Eustice, R.M. (2015). Active visual slam for robotic area coverage: Theory and experiment. *The International Journal of Robotics Research*, 34(4-5), 457–475.

Longuet-Higgins, H.C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828), 133–135.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110.

Lu, H., Li, Y., and Serikawa, S. (2017). Computer vision for ocean observing. In *Artificial Intelligence and Computer Vision*, 1–16. Springer.

OpenCV (2015). Open source computer vision library.

Panish, R. and Taylor, M. (2011). Achieving high navigation accuracy using inertial navigation systems in autonomous underwater vehicles. In *OCEANS 2011 IEEE-Spain*, 1–7. IEEE.

Reza, A.M. (2004). Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1), 35–44.

Roznere, M. and Li, A.Q. (2020). Underwater monocular image depth estimation using single-beam echosounder. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1785–1790. IEEE.

Ruscio, F., Peralta, G., Pollini, L., and Costanzi, R. (2021). Information communication technology (ict) tools for preservation of underwater environment: A vision-based posidonia oceanica monitoring. *Marine Technology Society Journal*, 55(4), 11–23.

Salvi, J., Petillo, Y., Thomas, S., and Aulinas, J. (2008). Visual slam for underwater vehicles using video velocity log and natural landmarks. In *OCEANS 2008*, 1–6. IEEE.

Scaramuzza, D. and Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE robotics & automation magazine*, 18(4), 80–92.

Shkurti, F., Rekleitis, I., Scaccia, M., and Dudek, G. (2011). State estimation of an underwater robot using visual and inertial information. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5054–5060. IEEE.