Vision-Based Control for an AUV in a Multi-robot Undersea Intervention Task

Emilio Garcia-Fidalgo(✉), Alberto Ortiz, and Miquel Massot-Campos

Department of Mathematics and Computer Science,
University of the Balearic Islands, Cra. Valldemossa km 7.5, 07122 Palma, Spain
{emilio.garcia,alberto.ortiz,miquel.massot}@uib.es

Abstract. This paper presents a novel vision-based framework for controlling an Autonomous Underwater Vehicle (AUV). In our application, this AUV is in charge of providing an alternative point of view of a predefined target during a multi-robot intervention mission, where two vehicles cooperate in order to perform the required task. Given this scenario, our framework is based on two main modules: on the one hand, a target detection and tracking module is used to determine the position of the target in the scene; on the other hand, a visual servoing module generates the required velocities for controlling the platform according to the estimated position of the target in the image plane. Results for a set of experiments in different environments are reported and discussed.

Keywords: Target detection · Tracking · Visual servoing · Underwater robotics

1 Introduction

In the last decades, robots have been widely used to explore areas which are usually hard to reach for humans. Underwater environments fall into this category, since their operating conditions make even simpler operations risky for human divers, especially when they have to be performed at high depth. A possible approach to overcome this problem is to use a Remotely Operated Vehicle (ROV), although this easily becomes a difficult and expensive solution because it usually requires a sophisticated support infrastructure and specialized staff. In this regard, the Spanish project MERBOTS proposes a new robot-based methodology to make intervention tasks safer, simpler and at a lower cost. The proposed system involves two vehicles, being one of them a Hybrid ROV (H-ROV) equipped with an arm and a manipulator, as well as the necessary perception devices, which altogether implement the supervised intervention task, while the other is an Autonomous Underwater Vehicle (AUV) endowed with cameras to provide alternative points of view of the target for the operator in charge of the H-ROV, enabling thus a more robust and reliable operation. In this paper,

This work was mainly supported by the Spanish project MERBOTS, subproject SUPERION (MINECO DPI2014-57746-C3-2-R).

A. Ollero et al. (eds.), *ROBOT 2017: Third Iberian Robotics Conference*, Advances in Intelligent Systems and Computing 693, https://doi.org/10.1007/978-3-319-70833-1_4

we focus on the task of controlling this AUV to provide these secondary views, which are part of the goals addressed by the MERBOTS project.

To this end, the object to manipulate must be ensured to appear continuously in the field of view (FOV) of the camera placed in the AUV. This naturally leads to the implementation of a visual servoing task, whose input is the image stream coming from the AUV camera and its outputs are the velocity commands to be sent to the vehicle controller so as to keep the target in the camera FOV at all times. Due to its well-known robustness and simpler implementation, in this application, we choose an Image-Based Visual Servo-control (IBVS) approach [3]. From a global point of view, the solution comprises two interacting processes, target detection and tracking, which provide input to the visual servo control strategy. Therefore, the main contribution of this paper is a generic and robust vision-based strategy for controlling an AUV which we experimentally validate in a real underwater platform under different operating conditions. The following sections describe all aspects of our solution: Sect. 2 discusses related work, Sect. 3 details our vision-based framework, Sect. 4 reports on the results of the different experiments performed and Sect. 5 draws some conclusions and discusses about future research.

2 Related Work

Undersea has always resulted a challenging environment for vision-based approaches due to the quality of the optical data which are finally available from the on-board imaging devices. Briefly speaking, absorption and scattering phenomena affecting light propagation in the significantly participative underwater media tend to produce blurred and low-contrasted images, which introduce additional difficulties into image-based inferring processes. Nonetheless, nowadays one can find a number of works reporting successful results for vision-based solutions operating underwater, some times taking into account those difficulties [13,23], while other approaches manage them minimizing the aforementioned optical effects, e.g. because image capture takes place at close distance to the scene, or by means of robust vision techniques [14,20]. In order to cope with real-time restrictions, our solution belongs to the second category, so no special steps are taken in order to enhance the images, but we deal with them through a robust software solution.

In any case, leaving the aforementioned difficulties aside and focusing on the use of vision for platform control, nowadays, given the computational capabilities of current state-of-the-art processors, vision has become a good choice for platform control because they are able to provide high resolution data with high speed acquisition at low cost. Moreover, vision-based control becomes specially useful in underwater intervention applications, where the robot usually needs to hover over the mission area [19,24]. Within this context, the FP7 project TRIDENT [24] proposed a new methodology elaborated around the concept of an Intervention Autonomous Underwater Vehicle (I-AUV), comprising two main elements: the vehicle itself, which, for the case of the TRIDENT project, consisted in the Girona-500 platform [22], and a dexterous hand-arm system attached to the robot. In order to determine the position of the object to manipulate, a pure feature-based detection technique was employed, i.e. the object of interest was located by means of a descriptor based on a constellation of image features stemming from its appearance underwater, and this process was repeated frame after frame, without taking into account the inherent temporal correlation between frames. A similar concept was employed during the Spanish project TRITON [19], which focused on improving the maintenance of submerged observatories through the use of robots instead of human divers. Our current solution adopts a tracking-based approach which takes into account the above-mentioned temporal correlation to reduce false positive detections of the target as well as to lower the computational times.

Several other works on visual control for underwater environments can be found in the literature: e.g. Lots *et al.* [15] make use of an IBVS approach in combination with a proportional-integral-derivative (PID) controller for solving a station-keeping problem, while Sattar *et al.* [25] proposes a system for controlling the underwater legged robotic system AQUA, using an object detection method based on a colour blob tracker and a proportional-derivative (PD) controller to generate the required pitch and yaw commands; finally, Gao *et al.* [8], in a more recent approach, introduce a hierarchical scheme for controlling underwater vehicles based on an adaptive neural network.

Some object detection methods based on a combination of a detection and a tracking stage have served us as inspiration while developing our approach. Dayoub *et al.* [5] have recently introduced a vision-based method for the identification and tracking of Crown-Of-Thorns starfishes in coral reefs. This method makes use of a machine learning approach for detection (based on a random forest classifier) and a particle filter for tracking. In [17], Martinez *et al.* describe a method for detecting and tracking electric towers using an aerial platform. The detection stage also utilizes machine learning through a two-class multilayer perceptron for target detection, while the tracking stage is based on a hierarchical strategy.

3 Visual Control Approach

Figure 1 outlines our approach. As can be seen, the system comprises a module to detect and track the object in the image (DAT) and a module to generate the corresponding control velocities for the vehicle (IBVS). Initially, the target is selected in the current image by defining a Region of Interest (ROI). The DAT module computes then a set of SIFT keypoints [16], which are used as a visual target model to search and track the target in the image stream. The coordinates of the ROI where the target has been found are accordingly updated and sent to the IBVS module, which generates the necessary control commands that are to make the target get centred in the image. Both modules, DAT and IBVS, are detailed next.

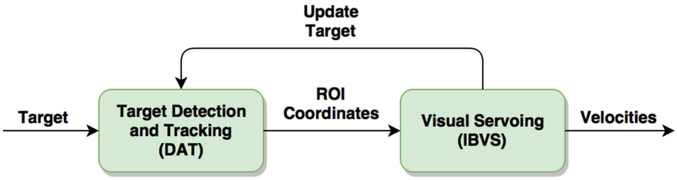

Fig. 1. Outline of the visual control approach. The DAT module is in charge of detecting the target in the current image and feed the IBVS module, which generates the AUV control commands.

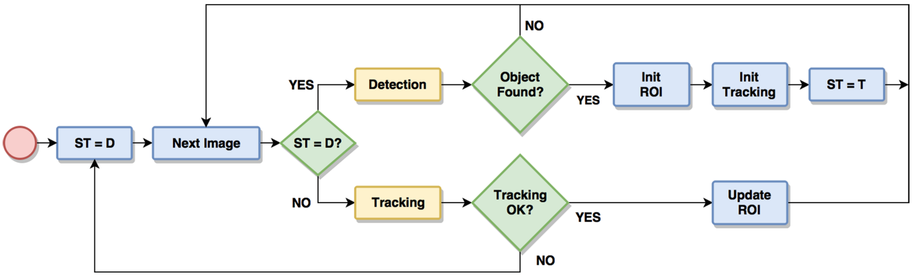

Fig. 2. Target detection and tracking strategy, based on the interaction between the detection and tracking stages. ST flags the current operation mode.

3.1 Target Detection and Tracking

As shown in Fig. 2, our strategy to estimate the position of the target in the image plane is based on two different stages, *detection* and *tracking*, which interact with one another. The *detection* stage is computationally expensive but robust to appearance changes. Conversely, the *tracking* stage is a more efficient process, but tends to lose the target from time to time, depending on the operating conditions. Taking into account these considerations, our strategy employs the tracking stage as much as possible and the detection stage is only used when the tracking system needs to be reinitialized. The system starts executing the detection stage of the DAT module. If the target is found in the current image, the corresponding bounding box is set as the ROI and used to initialize the tracking process. This stage keeps estimating the position of the target until it considers that it has lost track of it. In such a case, the detection process activates again and operates until the target is relocated.

The detection stage begins computing a set of SIFT keypoints in the current image. SIFT has been selected due to its well-known tolerance to appearance changes [16], what usually results convenient in underwater environments. In order to determine the presence of the target in the scene, a collection of putative matches are found between the current image SIFT descriptors and the target model, also consisting of a set of SIFT descriptors. For efficiency reasons, this step is implemented using a set of randomized kd-trees, which index the SIFT descriptors of the target model by means of several hierarchical structures. Using these kd-trees, for each SIFT descriptor found in the current image, the nearest and the second-nearest neighbours are searched and then the nearest neighbour distance ratio test [16] is applied to discard incorrect matches. Next, the surviving matches are employed to compute a homography between both sets of descriptors. The estimation of this homography relies on Random Sample Consensus (RANSAC) [7] as a robust estimation algorithm to minimize the error and discard outliers. After that, if the resulting number of inliers is high enough, we consider that the target has been found and the resulting homography is used to determine the coordinates of the target corners in the current image. The minimal up-right bounding box is calculated using these coordinates, and the corresponding corners used as input to the IBVS module.

For the tracking stage, we have considered two well-known visual tracking algorithms, Struck [11] and KCF [12], which have correspondingly been adapted to our purposes, so that the system can make use of any of them. Nonetheless, we have empirically noted that KCF performs better in computational terms. Irrespective of the option chosen, during tracking, we determine whether the target has been lost in order to correctly feed the IBVS module during the full intervention mission. We have opted for computing the χ^2 distance between two Pyramid of Histograms of Orientation Gradients (PHOG) descriptors [1], one for the target and one for the current ROI, provided by the tracker. This global descriptor, originally developed for image classification, has been configured for this application for 60 bins and 3 levels, generating a real-valued vector of 1260 components. The detection stage becomes active again when the χ^2 distance is high enough. This threshold clearly affects the number of times that the detection stage is executed.

3.2 Image-Based Visual Servoing

IBVS control operates in terms of image positions. In one of the many possible approaches, the goal is to make a set of image points (features) s attain a set of desired positions s^*, which implicitly moves the involved platform. To this end, IBVS defines a model that relates the camera velocities $\xi_c(t)$ to the velocities of the selected features over the image plane $\dot{s}(t) = [\dot{s}_{1,x}(t), \dot{s}_{1,y}(t), \ldots, \dot{s}_{n,x}(t), \dot{s}_{n,y}(t)]^T$ through the so-called *interaction matrix* L [4]. In our case, we conveniently include the transformation from robot to camera ${}^cT_r$, to obtain velocity commands in the robot frame (ξ_r):

$$\dot{s}(t) = L\,\xi_c(t) = L\,({}^cT_r\,\xi_r(t)) = L'\,\xi_r(t)\,. \tag{1}$$

Robot motion needed to move the image features to the desired image positions is then derived from (1) in the form of (2):

$$\xi_r = (L')^+\,\dot{s}(t)\,, \tag{2}$$

where $(L')^+$ is the pseudoinverse of L' resulting from a least squares framework. For our application, the corners of the ROI detected by the DAT module are used as the features s, while, to set s^*, those corners are required to get centred in the image.

In general terms, IBVS is designed to make the current feature positions $s(t)$ coincide with the set of desired positions s^*, i.e. minimize the corresponding error function $e(t) = s(t) - s^*$. In our approach, similarly to [15], we adopt a PID-like control scheme to this end, so that the final control law results to be:

$$\xi_r(t) = -(L')^+ \left(\lambda_\mathrm{p}\, e(t) + \lambda_\mathrm{i} \int_0^t e(\tau)\, d\tau + \lambda_\mathrm{d} \frac{d\, e(t)}{dt} \right) \tag{3}$$

being λ_p, λ_i and λ_d the, respectively, proportional, integral and derivative gains of the controller. This control scheme is replicated for each degree of freedom (d.o.f) of the AUV that needs to be controlled, adopting an uncoupled control solution, so that different gain values result for each d.o.f.

As previously said, in this work, we make use of the ROI corners as image features, which have to be properly tracked to correctly compute the error function $e(t)$ required by (3). Additionally, the appearance of the target is updated during the intervention to improve the performance of the tracking module; the update takes place whenever the norm of $e(t)$ is low enough.

4 Experimental Results

The visual control approach described in this paper has been implemented under the AUV control framework COLA2 [18] as a set of ROS/C++ nodes [9,10]. Our implementation has been first tested under simulation within the UWSim simulator [21], and lastly on a real AUV.

Both approaches make use of the same software as for the DAT and the IBVS modules, running on an Intel i5 @ 2.4 GHz/8 Gb RAM computer. The input for the whole system is a rectified 512 × 384 colour image stream at 10 frames per second, altogether with the camera calibration information (intrinsics and extrinsics) and the vehicle altitude (e.g. to compute the distance to the target), whilst the output is a body-centered velocity request which is sent to the vehicle controller. This request contains surge, sway and yaw velocity commands. Heave is governed by the altitude controller available in the AUV control architecture.

The robot is a 1.6 m long torpedo-shaped AUV, named SPARUS II [2], designed and built by the University of Girona, Spain. The vehicle is equipped with three propellers, one in heave and two in surge, a Doppler Velocity Log (DVL), a Global Positioning System (GPS), an Inertial Measurement Unit (IMU) and a pressure sensor. These navigation sensors are fused in an Extended Kalman Filter (EKF) whose output is the position and orientation of the vehicle. The nose of the vehicle holds a modular payload area that can be equipped with the needed sensors for a particular intervention. In the context of the project MERBOTS, the vehicle has been equipped with an additional propeller that allows the vehicle to move in sway, and a Bumblebee2 stereo camera oriented at a 45-degree angle. This camera configuration permits us to keep the platform at a certain distance from the target, avoiding interferences during the manipulation task performed by the H-ROV (as described in Sect. 1). Therefore the vehicle is governable in surge, sway, heave and yaw. Pitch and roll are passively stabilized by means of floats and loads in the vehicle.

The next sections report on the results obtained during some of the simulation and field experiments which have been carried out. In order to show the visual control approach working on the real AUV, the videos for the following water tank and open sea experiments have been made publicly available at http://srv.uib.es/superion.

4.1 Simulation Results

The dynamic model of SPARUS II has been incorporated into UWSim to evaluate the performance of the visual servoing under three different external perturbations: (1) surge, (2) sway and (3) yaw rotation, trying to emulate possible effects of underwater currents.

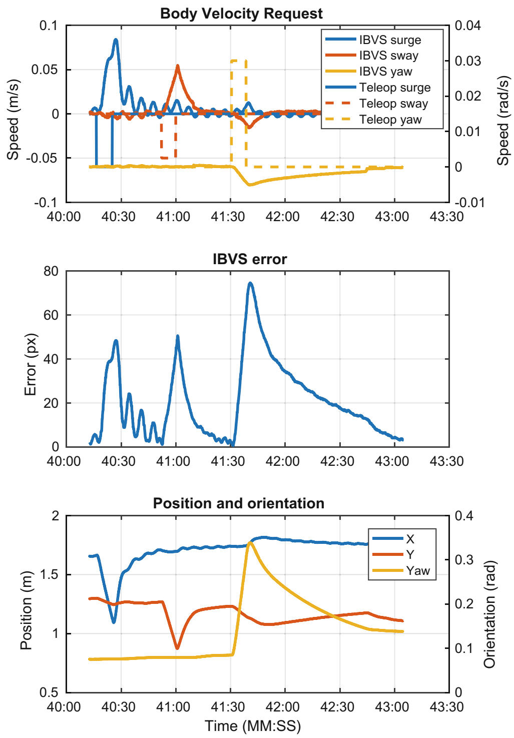

Fig. 3. Simulation results in UWSim. The vehicle is requested to move in three different degrees of freedom using a remote controller (through the *Teleop* module), while the *visual servoing* module (IBVS) responds to these perturbations. The IBVS commands succeed in decreasing the error by moving the vehicle to the desired position.

Figure 3 shows the evolution of the position and orientation of the vehicle under velocity requests, coming from the *Image-based Visual Servoing* (IBVS) module, while external perturbations, introduced by means of a remote controller through the *Teleop* module, are in place. In this figure, the error is expressed as the average Euclidean distance in pixels between the corners of the desired and the detected target ROIs. Starting with a centred and tracked object, a negative surge speed is requested at time 40:20. As soon as the target is no longer in the centre of the image, the IBVS module sends velocity commands to the AUV. This same behaviour can be seen for sway commands (time 40:50) and for a

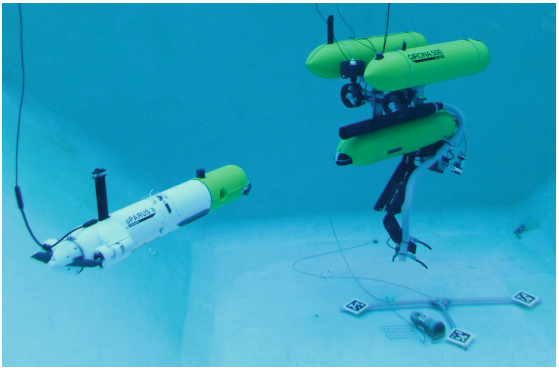

Fig. 4. CIRS water tank experiment (I). SPARUS II AUV and Girona-500 H-ROV during a water tank trial intervention. SPARUS II is running the DAT and the IBVS modules to provide the H-ROV pilot with a second point of view of the area.

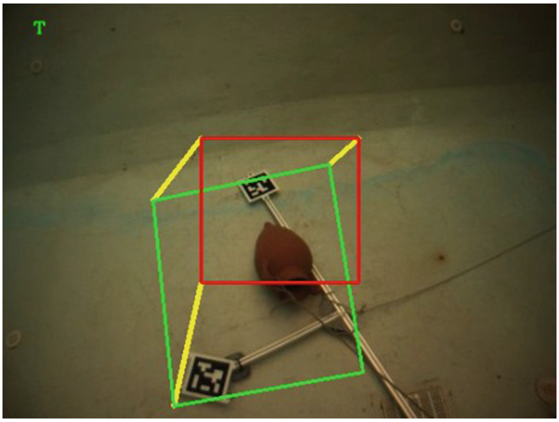

Fig. 5. CIRS water tank experiment (II). SPARUS II camera view under visual servoing. The detected ROI is shown in green, while the desired position is indicated in red. In order to minimize the error, the IBVS module has to command the vehicle to match the four corners as close as possible.

rotation in yaw (time 41:30). In these three cases, the vehicle is requested by IBVS to move to a position where the target is again centred in the image, thus decreasing the IBVS error.

The transient response strongly depends on the value of the PID gains, affecting the speed, settling time and overshoot of the system. These values have been tuned to achieve a movement free of motion blur in the camera image. The oscillatory surge response which can be observed is mostly due to the low hydrodynamic drag of the simulated vehicle.

4.2 Water Tank Experiments

A number of water tank tests have been carried out at the facilities of the Girona Underwater Vision and Robotics Centre (CIRS), in order to validate the target detection/tracking and platform servoing functionalities prior to the field trials at sea, similarly to the simulation experiments. Figure 4 shows the complete intervention scenario employed as the experimental set-up. As can be observed,

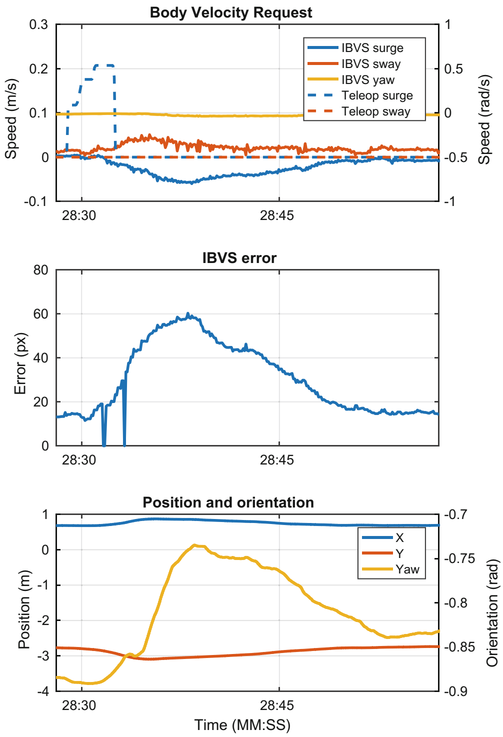

Fig. 6. CIRS water tank experiment (III). The vehicle is requested to move in surge by means of the remote controller (through the *Teleop* module) and the *visual servoing* (IBVS) counteracts the displacement, diminishing the error between the desired and the detected ROI corners.

this scenario comprised an amphora, to be grasped during the intervention, and a set of AR markers to share a common reference frame between the two submarines involved. These markers are not used for detection nor tracking.

The viewport of SPARUS II performing visual servoing of the amphora can be found in Fig. 5, where the detected ROI is drawn in green and the desired positions of these same corners appear in red. As explained before, the velocity commands to the vehicle depend on the positions of the two sets of corners.

In Fig. 6, a perturbation in surge is shown at time 28:30, starting with a centred and tracked object. As before, IBVS starts sending velocity commands as soon as the target gets out of its centred position, moving the AUV to a position that diminishes the error, e.g. with the object in the centre of the image. In this scenario, the error jumps to zero twice when the tracking gets lost. On the next few frames, the target is quickly re-detected and IBVS moves the vehicle accordingly.

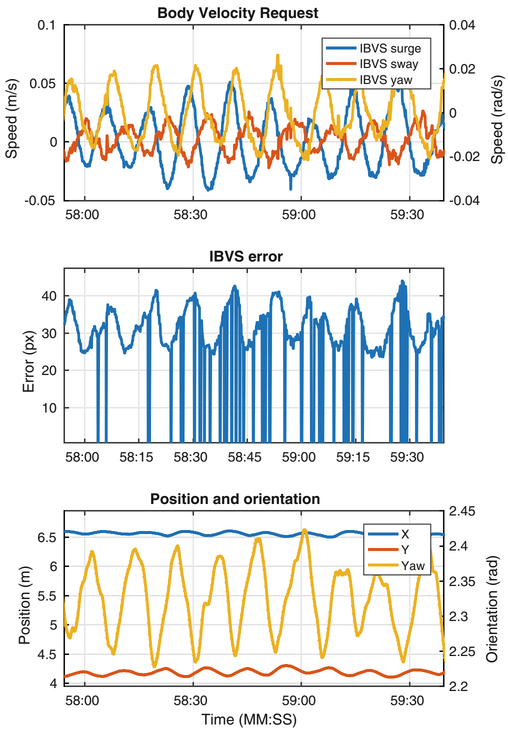

Fig. 7. Sea trials at Sant Feliu de Guixols (II). IBVS reacts to the ocean currents demanding surge, sway and yaw velocity requests, accomplishing (1) bounded error performance [7–12% of half the image diagonal length] and (2) confinement of the target in the camera FOV at all times.

4.3 Open Sea Experiments

Sea trials were carried out at Sant Feliu de Guixols (Girona), where our complete approach was tested under real ocean currents. The two vehicles were deployed and were working at a depth of six to ten meters. The seabed was covered in mud, poor in features, and the lack of light made the intervention conditions not quite similar to the previous water tank tests. Velocity requests to keep the target in the camera FOV during another experiment are reported in Fig. 7.

This challenging scenario shows that not only is the IBVS capable of holding such currents, but we can also see that the error is bounded, even when tracking gets lost (in Fig. 7 [middle] IBVS jumps to zero error when tracking is lost) and new detections have to be performed. Note that the orientation is stretched in the figure, thus the minima to maxima difference is about nine degrees.

5 Conclusions

This paper has described a complete and robust vision-based framework for controlling an AUV. Our approach is devised to keep a constant, although alternative, point of view of a predefined target during an intervention mission. Due to this reason, the solution consists of two main components: a *detection and tracking* module, which is in charge of estimating the position of the involved target in the image stream, and an *image-based visual servoing* module, which generates the velocities required to maintain the target centred in the image plane. The solution has been validated by means of several experiments using a real platform and under different operating conditions.

Referring to future work, we believe that recent results on deep learning methods for object detection [6] is a potential research line to investigate to further improve our solution. We will also investigate other tracking methods that does not restrict the target to a rectangular ROI (as required by Struck [11] and KCF [12]). Finally, we plan to consider the implementation of other visual servoing methods and their suitability for underwater environments.

References

1. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel, pp. 401–408 (2007)
2. Carreras, M., Candela, C., Ribas, D., Mallios, A., Magí, L., Vidal, E., Palomeras, N., Ridao, P.: Sparus II, design of a lightweight hovering AUV. In: Proceedings of the International Workshop on Marine Technology, vol. 911, pp. 163–164 (2013)
3. Chaumette, F., Hutchinson, S.: Visual servo control. Part I: basic approaches. IEEE Robot. Autom. Mag. **13**(4), 82–90 (2006)
4. Corke, P.: Robotics, Vision and Control: Fundamental Algorithms in MATLAB. Springer, New York (2011)
5. Dayoub, F., Dunbabin, M., Corke, P.: Robotic detection and tracking of crown-of-thorns starfish. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1921–1928 (2015)

6. Druzhkov, P.N., Kustikova, V.D.: A survey of deep learning methods and software tools for image classification and object detection. Pattern Recogn. Image Anal. **26**(1), 9–15 (2016)
7. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
8. Gao, J., Proctor, A.A., Shi, Y., Bradley, C.: Hierarchical model predictive image-based visual servoing of underwater vehicles with adaptive neural network dynamic control. IEEE Trans. Cybern. **46**(10), 2323–2334 (2016)
9. Garcia-Fidalgo, E., Massot-Campos, M.: srv/merbots_ibvs: v0.1, May 2017. doi:10.5281/zenodo.576367
10. Garcia-Fidalgo, E., Massot-Campos, M.: srv/merbots_tracking: v0.1, May 2017. doi:10.5281/zenodo.576370
11. Hare, S., Golodetz, S., Saffari, A., Vineet, V., Cheng, M.M., Hicks, S.L., Torr, P.H.: Struck: structured output tracking with kernels. IEEE Trans. Pattern Anal. Mach. Intell. **38**(10), 2096–2109 (2016)
12. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
13. Hildebrandt, M., Kerdels, J., Albiez, J., Kirchner, F.: Robust vision-based semi-autonomous underwater manipulation. In: Proceedings of the Intelligent Autonomous Systems (2008)
14. Krupinski, S., Desouche, R., Palomeras, N., Allibert, G., Hua, M.D.: Pool testing of auv visual servoing for autonomous inspection. In: Proceedings of the IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles (2015)
15. Lots, J.F., Lane, D.M., Trucco, E., Chaumette, F.: A 2D visual servoing for underwater vehicle station keeping. In: Proceedings of the IEEE International Conference on Robotics and Automation, vol. 3, pp. 2767–2772 (2001)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
17. Martinez, C., Sampedro, C., Chauhan, A., Campoy, P.: Towards autonomous detection and tracking of electric towers for aerial power line inspection. In: Proceedings of the International Conference on Unmanned Aircraft Systems, pp. 284–295 (2014)
18. Palomeras, N., El-Fakdi, A., Carreras, M., Ridao, P.: COLA2: a control architecture for AUVs. IEEE J. Oceanic Eng. **37**(4), 695–716 (2012)
19. Palomeras, N., Penalver, A., Massot-Campos, M., Vallicrosa, G., Negre, P.L., Fernández, J.J., Ridao, P., Sanz, P.J., Oliver-Codina, G., Palomer, A.: I-AUV docking and intervention in a subsea panel. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2279–2285 (2014)
20. Perez-Alcocer, R., Torres-Mendez, L.A., Olguin-Diaz, E., Maldonado-Ramirez, A.A.: Vision-based autonomous underwater vehicle navigation in poor visibility conditions using a model-free robust control. J. Sens. **2016** (article ID 8594096) (2016)
21. Prats, M., Perez, J., Fernandez, J.J., Sanz, P.J.: An open source tool for simulation and supervision of underwater intervention missions. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2577–2582 (2012)
22. Ribas, D., Palomeras, N., Ridao, P., Carreras, M., Mallios, A.: Girona 500 AUV: from survey to intervention. IEEE/ASME Trans. Mechatron. **17**(1), 46–53 (2012)

23. Roser, M., Dunbabin, M., Geiger, A.: Simultaneous underwater visibility assessment, enhancement and improved stereo. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 3840–3847 (2014)
24. Sanz, P.J., Ridao, P., Oliver, G., Casalino, G., Insaurralde, C., Silvestre, C., Melchiorri, C., Turetta, A.: TRIDENT: recent improvements about autonomous underwater intervention missions. IFAC Proc. Volumes **45**(5), 355–360 (2012)
25. Sattar, J., Giguere, P., Dudek, G., Prahacs, C.: A visual servoing system for an aquatic swimming robot. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1483–1488 (2005)