Jornadas Nacionales de Robótica
Spanish Robotics Conference
8-9 Junio 2017 http://jnr2017.ai2.upv.es

Multifunctional Cooperative Marine Robots for Intervention Domains: Target detection, tracking and recognition issues

Emilio García-Fidalgo, Joan Pep Company-Córcoles, Alberto Ortiz*

Miquel Massot-Campos, Pep Lluis Negre-Carrasco, Gabriel Oliver-Codina

Departament of Mathematics and Computer Science, Universitat de les Illes Balears, Cra. Valldemossa, km 7.5, 07122, Palma de Mallorca, Spain

Abstract

This paper is one of a series of three describing the MERBOTS project, a three-year research initiative funded under the well-known DPI Spanish research program. In brief, MERBOTS aims at progressing in the field of underwater intervention operations. Nowadays, when the mission area is too deep and risky to be carried out by divers, the alternative consists in using remotely operated vehicles (ROV). This is a difficult and expensive solution requiring sophisticated support infrastructure and specialized personnel. Consequently, the use of robotic technology is normally limited to strategic or high added value operations, e.g. rescue, offshore industry or security and defense. The MERBOTS project proposes a robot-based intervention system, and the underlying methodology, that will permit safer intervention tasks, at a lower cost, and operationally simpler, thanks to multi-robot cooperation and extensive use of multimodal perception systems. As a result, new application areas, such as marine archaeology at high depth, turn to be affordable, with important consequences not only from the economic point of view, but also scientifically, socially and culturally speaking. As a final note, MERBOTS is a coordinated project involving three different Spanish research groups at University Jaume I of Castellon (UJI), University of Girona (UdG) and University of Balearic Islands (UIB), where each group is assuming specific goals under three different subprojects and the coordination of UJI. It is the intention of this consortium to describe MERBOTS through three different papers, corresponding each to one of the subprojects. This paper provides thus the UIB point of view in the form of the SUPERION subproject. SUPERION mainly focuses on the perception aspects of the intervention operation. In particular, in this paper we address the target detection, tracking and recognition tasks, and present first experimental results for the archeological application. *Copyright* © XXXX CEA.

Keywords:
Marine robotics; Underwater intervention; Multi-robot system; Visual tracking; Visual servoing; Target detection.
Project data:
Title: Multifunctional coopERative marine roBOTs for Intervention DomainS - MERBOTS
Reference: DPI2014-57746-C3-2-R (Superion)
Main researcher: Gabriel Oliver (IP1) & Alberto Ortiz (IP2) [UIB]
Type (international, national, local, technology transfer): National
Funding entity: MINECO
From/to dates: 1/01/2015 – 31/12/2017

1. Introduction

In the last decades, robots have been used to explore areas hard to reach for humans. Underwater environments fall into this category, since their operating conditions make even simpler operations risky to be carried out by 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 project MERBOTS proposes a new robot-based methodology to make intervention tasks safer, simpler and at a lower cost. On the one hand, our methodology considers a semi-supervised operation, i.e. an operator is within the main control loop, assisted by the system during the operation. On the other hand, the proposed system comprises 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, we focus on the visual target detection and tracking tasks to be performed to provide this secondary view, as well as on the target recognition task that provides input to the manipulator from the point clouds regularly stemming from the perception devices the H-ROV is equipped with. These are part of the goals addressed by the SUPERION subproject (together with visual mapping, 3D

* Corresponding author. *e-mail*: emilio.garcia@uib.es (Emilio García), joanpep.company@uib.es (Joan Pep Company), alberto.ortiz@uib.es (Alberto Ortiz), miquel.massot@uib.cat (Miquel Massot), pl.negre@uib.es (Pep Lluis Negre), goliver@uib.es (Gabriel Oliver)

reconstruction and multi-modal sensor fusion), for which we also report preliminary experimental results.

To provide the alternative view, the object to manipulate must be ensured to appear continuously in the field of view 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 field of view 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 (Chaumette & Hutchinson 2006). 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.

The intervention operation is performed in parallel with the generation of these alternative views. As mentioned before, an operator is involved during the intervention, making decisions with the assistance of the system, which, among others, recognizes the target in the perception data stream and provides input to the subsystem guiding the arm and the manipulator. The perception stream consists, in this case, in a sequence of point clouds of the scene which are generated by a suitable device, e.g. a stereo camera or an underwater laser scanner, carried by the H-ROV. A 3D model of the object to recognize, from a library of possible targets, is employed through a sequence of matching operations which aim at providing the manipulator controller with the target pose, i.e. position and orientation, to readily perform the grasping step.

The rest of the paper is organized as follows: Section 2 describes the alternative view generation approach; Section 3 details the target recognition algorithm; to finish, conclusions are summarized in Section 4.

2. Alternative view generation approach

Figure 1 outlines our approach, comprising the *target detection and tracking* (DAT) module and the *visual servoing* (IBVS) module, which generates 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 (Lowe 2004) as the target model. This model is used 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 centered in the image. Both modules, DAT and IBVS, are detailed next.

2.1 Target detection and tracking

As shown in Figure 2, our strategy to estimate the position of the target in the image plane is based on two different stages, *detection* and *tracking*, each interacting 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. 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 retrained.

The system starts executing 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.

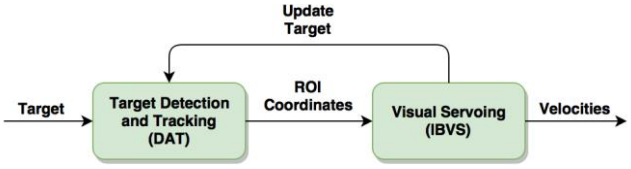

Figure 1: Outline of the alternative view generation approach. The DAT module is in charge of detecting the target in the current image. The detected position is then used by the IBVS module to generate the AUV control commands.

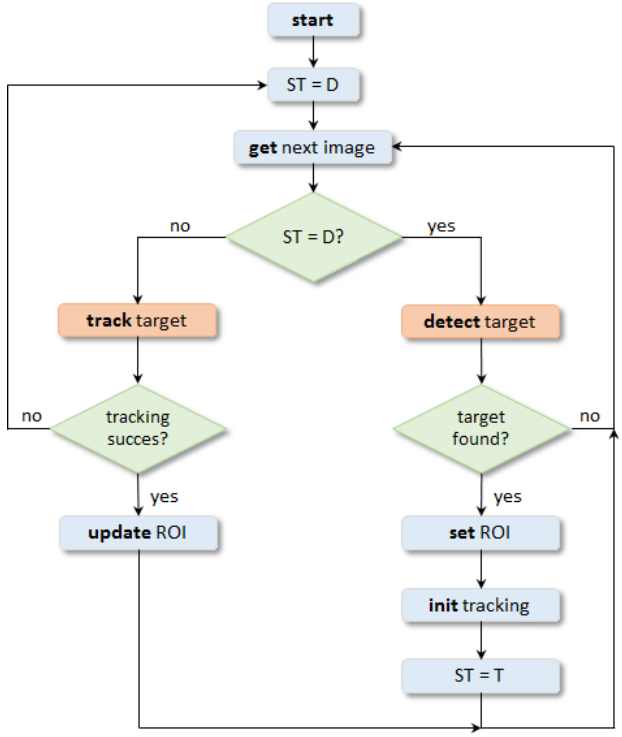

Figure 2: Target detection and tracking. As can be seen, the strategy is based on the interaction between the *detection* and *tracking* stages. ST (*status*) flags the current operation mode: D – *detection*, T – *tracking*.

The detection stage begins computing a set of SIFT keypoints in the current image. A collection of putative matches are found between the current image SIFT features and the target model, also consisting of a set of SIFT features. For efficiency reasons, this task is implemented using a set of randomized kd-trees and applying the nearest neighbour distance ratio test to discard incorrect matches (Lowe 2004). The surviving matches are then employed to compute a homography between both descriptors. 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 estimate the coordinates of the target ROI corners in the current image. The minimal up-right bounding box is calculated using these coordinates, and the corresponding corners used as input by the IBVS module.

For the tracking process, we have considered two well-known visual tracking algorithms, Struck (Hare et al. 2016) and KCF (Henriques et al. 2015), 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. In any case, during tracking, we compute the distance between global PHOG descriptors (Bosch et al. 2007) for the target and the current ROI to determine whether the target has been lost. The detection stage becomes active again if this distance is higher than a threshold.

2.2 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 $\dot{s}(t) = [\dot{s}_{1,x}(t), \dot{s}_{1,y}(t), \cdots, \dot{s}_{n,x}(t), \dot{s}_{n,y}(t)]^T$ through the so-called *interaction matrix L* (Corke 2011). In our case, we conveniently include the transformation from robot to camera T_R^C, to obtain velocity commands in the robot frame (ξ_R):

$$\dot{s}(t) = L\xi_C(t) = L\left(T_R^C \xi_R(t)\right) = L'\xi_R(t) \quad (1)$$

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

$$\xi_R(t) = (L')^+ \dot{s}(t) \quad (2)$$

where $(L')^+$ is the pseudoinverse of L'. 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 centered in the image.

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

$$\xi_R(t) = (L')^+ \left(\lambda_p e(t) + \lambda_d \frac{de(t)}{dt} + \lambda_i \int_0^t e(t)dt\right) \quad (3)$$

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, 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 equation (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 (see Figure 1).

2.3 Experimental results

Figure 3 illustrates some experiments involving the Girona-500 platform (Ribas et al. 2012) as H-ROV and the Sparus II platform (Carreras et al. 2013) as the AUV, being the latter fitted with a lateral thruster for sway motion. These pictures come from a first series of field trials performed by the MERBOTS consortium in a water tank at the Research Center in Underwater Robotics (CIRS, UdG) and in the sea at Sant Feliu de Guíxols (Girona). Videos of trials in both the water tank and at sea are also available at http://srv.uib.es/superion.

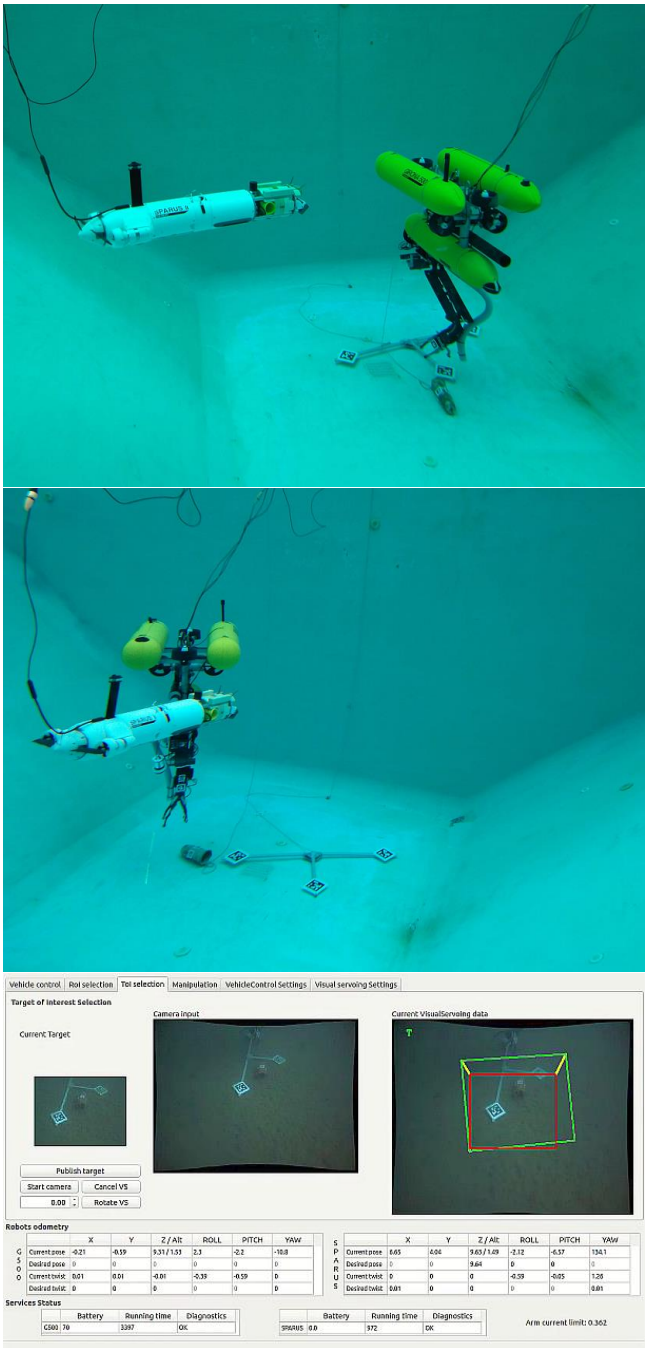

Figure 3: Illustration of a multi-robot operation: [rows 1 & 2] pictures from two intervention operations in a water tank involving the H-ROV and the UAV, the latter providing an alternative view of the scene; [row 3] MERBOTS GUI during an intervention in the sea: the right subimage illustrates the state of the IBVS during the operation, where the red and the green boxes correspond to, respectively, the detected/tracked corners (s) and the desired corners (s^*).

3. Target recognition approach for grasping

In this section, we describe our approach for recognizing a specific object in a point cloud, which has been used not only

for detecting the presence of the target in the scene during the intervention operation, but also to determine its pose over the sea floor in order to use it as input during the grasping operation. To this end, the algorithm assumes the availability of a 3D model of the target, which is registered to the incoming point clouds.

After filtering the current point cloud to discard outliers (arising from time to time depending on the operating conditions and the sensor used, e.g. a stereo pair or a laser scanner), the object recognition pipeline comprises two stages, *detection* and *tracking*, which alternate according to the quality of the results. The *detection* stage combines two target detection methods which aim at being able to locate the target over generic backgrounds, being either a flat sandy seabed or a non-regular rocky surface. The *tracker* essentially makes use of a fast registration strategy to update the target pose from point cloud to point cloud, assuming slow or no motion in-between. Our strategy employs the tracking stage as much as possible for efficiency reasons. Figure 4 outlines the full process.

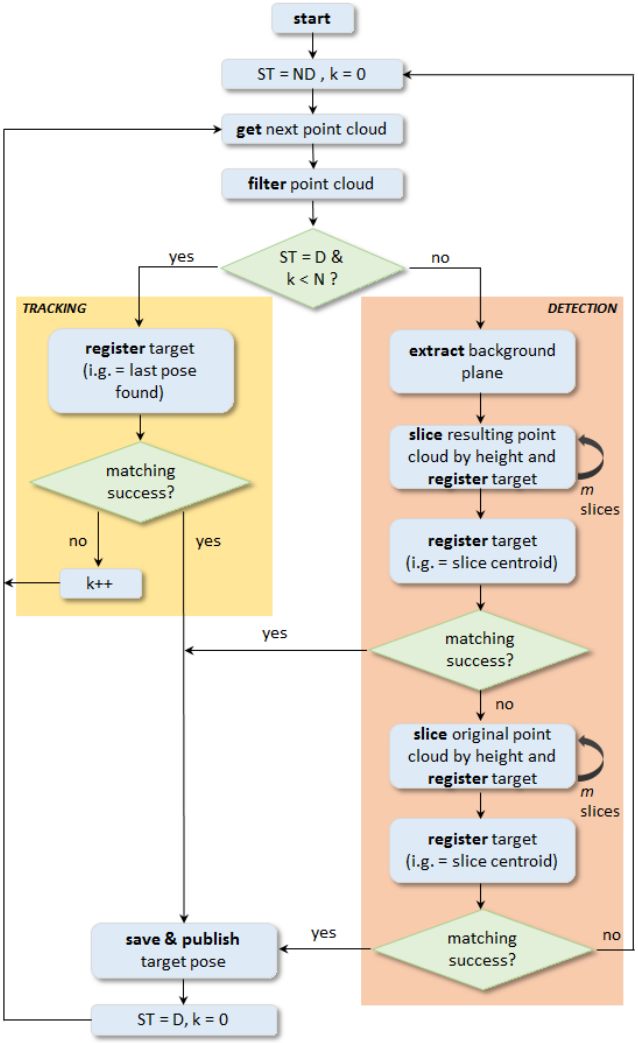

Figure 4: Object detection and tracking pipeline: i.g. stands for *initial guess*, while ST (*status*) flags the current operation mode: ND – *target not detected*, D – *target detected*.

Both processes make use, one way or another, of the Iterative Closest Point (ICP) algorithm (Pomerleau et al. 2015). ICP is commonly used to obtain the transformation that aligns two point sets, where one point set, the reference, is kept fixed, while the other, the source, is transformed to best match the reference. The ICP algorithm mainly works searching, for each point in the source point set, the closest point in the reference point set. In our case, the ICP is used to estimate the spatial transformation that aligns the target and the scene, and to obtain a fitting score as the average distance between the points of one set to the nearest points of the other set. The two sets are deemed to match if this score is low enough.

3.1 Target detection

The purpose of this stage is to detect the target in the scene point cloud and provide a first approximate location. As a first step, this stage starts by rectifying the point cloud using the known, constant transformation from sensor to robot, so that the scene point cloud frame is "parallel" to the robot frame (and also independent of the sensor, e.g. a stereo camera or a laser scanner, the point cloud comes from).

A RANSAC-based plane model segmentation (Fischler & Bolles 1981) follows next to the point cloud rectification, in order to search for all the points within the point cloud that support a plane model. This is to detect the ground plane, and the objects lying over it, within those scenes where the seabed is horizontal, typically sandy areas. The next step subtracts the ground plane from the point cloud.

The resulting point cloud is next sliced into vertical layers along the direction of the Z axis using Δz increments, while the target model is sequentially registered to the respective point sets. When the target model is determined to register with the current slice with low error, the target is considered to have been found and the slice centroid is used as initial guess to register the target model to the full point cloud.

This segmentation-by-height-and-match method has been shown to both reduce the execution time and improve the registration when there are multiple objects lying over the floor.

In case the floor is not planar, the previous steps based on detecting the seafloor plane are prone to fail, what activates a second strategy which repeats the process without background plane subtraction.

When the target is detected, its pose is saved to be used as initial guess by the tracking stage and transformed to the world frame to be published for the grasping process.

3.2 Target tracking

Once the object has been detected, the tracking module evaluates the next point cloud using the previous pose as the initial guess for the registration between the input cloud and the target. Unlike the detection process, this stage does not require a rectified point cloud, because no height segmentation is required, but the target has to be registered against the input scene.

If the registration succeeds, i.e. the registration error is low enough, then the object is considered correctly tracked, the detected pose saved for the next tracking cycle and also transformed to the world frame, to be finally published to guide the grasping.

If the registration fails for more than N consecutive times, the system is considered to have lost track of the target and the process is reinitialized to the target detection mode. In our experiments, N was set to 5.

3.3 Experimental results

Figures 5 and 6 show preliminary successful target recognition results for two experiments from the series of field trials mentioned in Section 2.3, finishing both in a correct grasping.

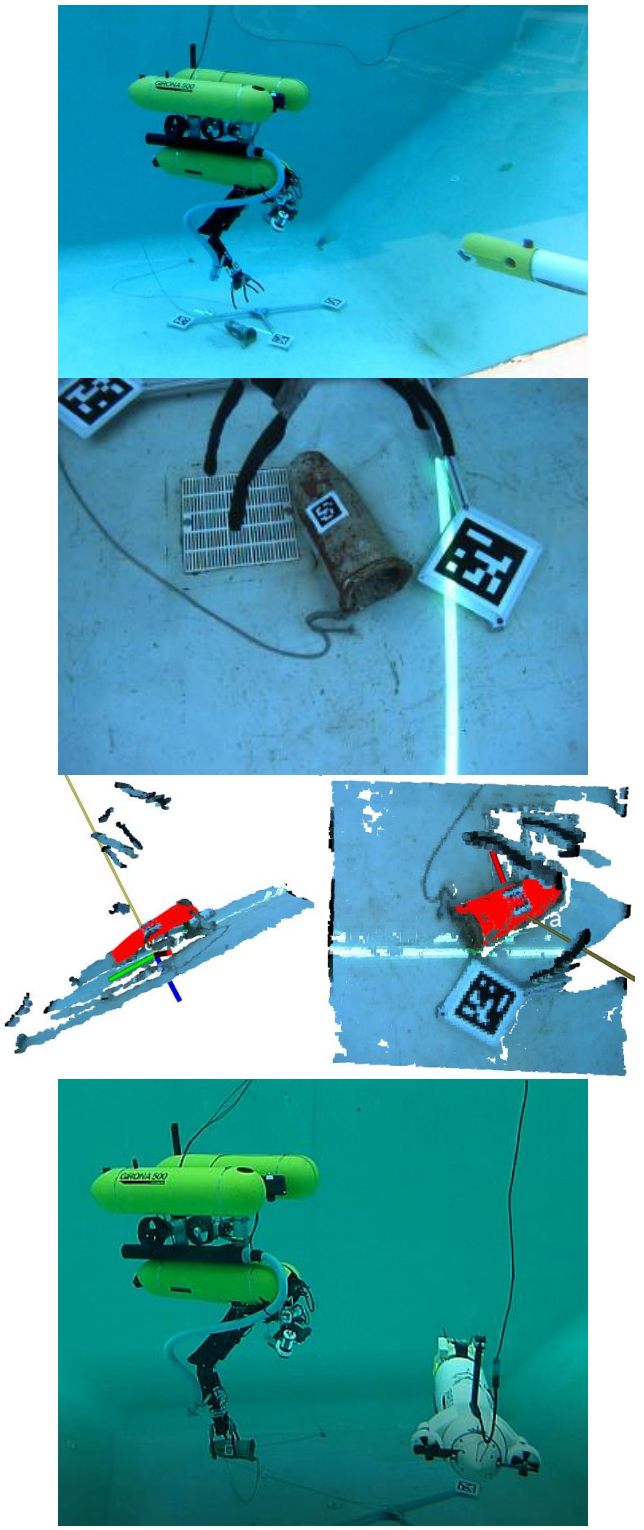

Figure 5: Illustration of an object recognition operation at the CIRS water tank: [row 1] Girona-500 and Sparus II platforms and global view of the scene; [row 2] example of image captured by the H-ROV; [row 3] different views of the recovered point cloud together with the target detected (red points) and its pose (coloured frame at [row 3, left]); [row 4] successful grasping.

As can be seen, one figure reports on an experiment at the water tank, while the other corresponds to a sea trial. Both figures show the target which is to be grasped by the Girona-500 manipulator, a sort of waterwheel bucket. In both cases, the point clouds come from a stereo camera carried by the H-ROV. Videos at http://srv.uib.es/superion also report on the target recognition/grasping experiments.

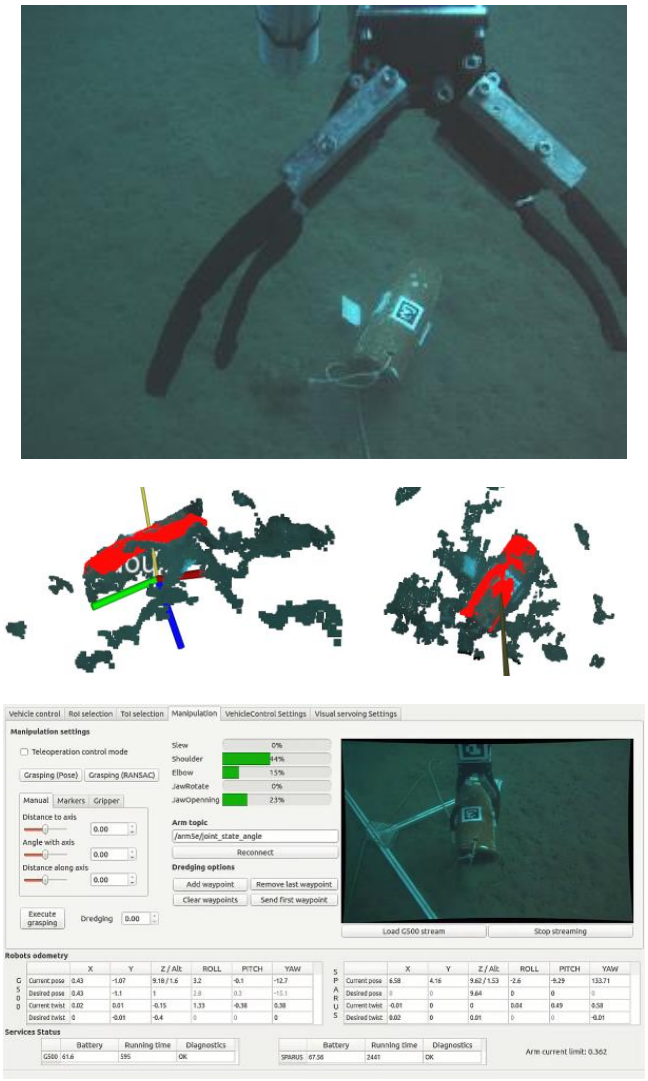

Figure 6: Illustration of an object recognition operation at sea: [row 1] global view of the scene; [row 2] different views of the recovered point cloud together with the target detected (red points) and the estimated pose (coloured frame at [row 2, left]); [row 3] successful grasping.

4. Conclusions

In this paper, we have addressed target detection, tracking and recognition issues related to the MERBOTS/SUPERION projects. For validation purposes, a first series of field trials in a water tank at the Research Center in Underwater Robotics (CIRS, UdG) and in the sea at Sant Feliu de Guíxols (Girona) have been recently performed by the MERBOTS consortium. Successful results for some of these trials, involving the Girona-500 and the Sparus II platforms in an archeological underwater application, have been reported.

Acknowledgements

This work has been partially supported by scholarship BES-2015-071804 and by the SUPERION project (MINECO DPI2014-57746-C3-2-R).

References

Bosch, A., Zisserman, A. & Muñoz, X. 2007. Representing Shape with a Spatial Pyramid Kernel, in IET Image Processing, 5(2), pp. 401-408.

Carreras, M., Candela, C., Ribas, D., Mallios, A., Magí, L., Vidal, E., Palomeras, N. and Ridao, P. 2013. Sparus II, design of a lightweght hovering AUV, in Instrumentation Viewpoint (Proc. Maritime Technology Conference), vol. 911, pp. 163-164.

Chaumette, F. & Hutchinson, S. 2006. Visual Servo Control. Part I: Basic Approaches, in IEEE Robotics & Automation Magazine, 13(4), pp. 82-90.

Corke, P. 2011. Robotics, Vision and Control. Springer.

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

Hare, S., Golodetz, S., Saffari, A., Vineet, V., Cheng, M., Hicks, S. & Torr, P. 2016. Struck: Structured Output Tracking with Kernels, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(10), pp. 2096-2109.

Henriques, J., Caseiro, R., Martins, P. & Batista, J. 2015. High-speed Tracking with Kernelized Correlation Filters, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), pp. 583-596.

Lowe, D. 2004. Disctinctive Image Features from Scale-Invariant Keypoints, in Int. J. Comput. Vision, 60(2), pp. 91-110.

Pomerleau, F., Colas, F. & Siegwart, R. 2015. A Review of Point Cloud Registration Algorithms for Mobile Robotics, in Foundations and Trends in Robotics. 4 (1), pp. 1–104.

Ribas, D., Palomeras, N., Ridao, P., Carreras, M. & Mallios, A. 2012. Girona 500 AUV: From Survey to Intervention, in IEEE/ASME Transactions on Mechatronics, 17(1), pp. 46-53.