Integration of a stereo vision system into an autonomous underwater vehicle for pipe manipulation tasks

Integration of a Stereo Vision System into an Autonomous Underwater Vehicle for Pipe Manipulation Tasks

Dario Lodi Rizzini[a], Fabjan Kallasi[a], Jacopo Aleotti[a], Fabio Oleari[a], Stefano Caselli[a]

[a]*Robotics and Intelligent Machines Laboratory, Department of Information Engineering, University of Parma, 43124, Italy*

Abstract

Underwater object detection and recognition using computer vision are challenging tasks due to the poor light condition of submerged environments. For intervention missions requiring grasping and manipulation of submerged objects, a vision system must provide an Autonomous Underwater Vehicles (AUV) with object detection, localization and tracking capabilities. In this paper, we describe the integration of a vision system in the MARIS intervention AUV and its configuration for detecting cylindrical pipes, a typical artifact of interest in underwater operations. Pipe edges are tracked using an alpha-beta filter to achieve robustness and return a reliable pose estimation even in case of partial pipe visibility. Experiments in an outdoor water pool in different light conditions show that the adopted algorithmic approach allows detection of target pipes and provides a sufficiently accurate estimation of their pose even when they become partially visible, thereby supporting the AUV in several successful pipe grasping operations.

Keywords: underwater computer vision, robot manipulation, stereo processing, object detection

1. Introduction

Autonomous robotic intervention in underwater environments is a growing research field with potential applications in inspection and maintenance tasks. Although several steps towards autonomous manipulation in harsh environments have been made, robot manipulation in underwater scenarios like oceans is still a very challenging problem for real-time perception and cognition. Water influences the mechanical and electrical design of the robot sub-systems, interferes with sensors by limiting their capabilities, heavily impacts on data transmissions, and it generally requires processing units with low power consumption to ensure an adequate mission duration. Among the various difficulties faced by underwater vision, water may range from clear to turbid even in water pools,

ambient light scattering varies with daylight hours and depth, strong color distortion may affect the perceived features of the object.

In this paper we present the stereo vision system that was designed in the context of the Italian national Project MARIS [1] for an autonomous underwater vehicle. The goal of the AUV, equipped with a robot arm with a three-fingered hand, is to perform autonomous manipulation tasks. The main contribution of the paper is the development and the experimental validation of a vision system capable to detect partially visible objects of known shape and color. Previous works regarding computer vision for underwater manipulation rested on stronger assumptions, such as full visibility of the target object and the use of markers or even simpler objects like boxes. In particular, we tackle the problem of designing an underwater vision system able to detect and track cylindrical objects like pipes. For this purpose, we exploit redundancy in actuation and adopt an alpha-beta filter to improve object detection. Pipes are relevant objects for grasping and transportation tasks on the underwater off-shore industry. Also, transportation of long pipes requires cooperative floating manipulation by two AUVs, which is the ultimate goal of the MARIS project. Cylindrical objects are often only partially visible during the task, since they are likely to exit cameras field of view (FoV) or be occluded by the manipulator itself. Hence, they pose additional challenges compared with small objects. The proposed system has been validated in an outdoor water pool manipulating cylindrical pipes with different color and sizes in both daylight and night-time conditions.

The paper is organized as follows. Section 2 reviews the state of the art in underwater perception and autonomous robot intervention using computer vision. Section 3 describes the AUV, the underwater vision system as well as the system setup and calibration phases. In Section 4 the proposed algorithmic solution for pipe detection and pose estimation is presented. Section 5 reports the experimental results. Section 6 concludes the work.

2. Related Work

Autonomous robots that perform manipulation tasks require more comprehensive and accurate models of the environment than other perception systems. For manipulation purposes, object models must include position, shape and other relevant information such as the way objects can be grasped. To achieve successful underwater manipulation tasks, problems like tracking and range estimation must be solved together with integration of perception with control and navigation.

Although computer vision is a standard perceptual modality in robotics for object detection, in underwater environments vision is not as widely used due to the problems arising with light transmission in water. Water turbidity, color aberrations, light reflections, and backscattering phenomena represent the major problems with underwater computer vision applications. Common applications of computer vision include detection and tracking of submerged artifacts [2] and underwater SLAM [3]. Stereo vision systems have been only recently introduced

in underwater applications due to the difficulty of calibration and the computational demand of stereo processing [4]. In [4] a robust underwater stereo vision system was developed for large-scale 3D reconstruction and visualization using texture-mapped surfaces at different scales.

Several approaches [5, 6] exploit color segmentation to find regions of interest where to perform object detection and recognition. Feature-based techniques have also been applied to salient color regions of interest or in scenarios where prior knowledge about the object shape is available [7]. Lodi Rizzini et al. [8] presented comparative experiments of object detection based on salient color uniformity and sharp contours using datasets acquired in heterogeneous underwater environments.

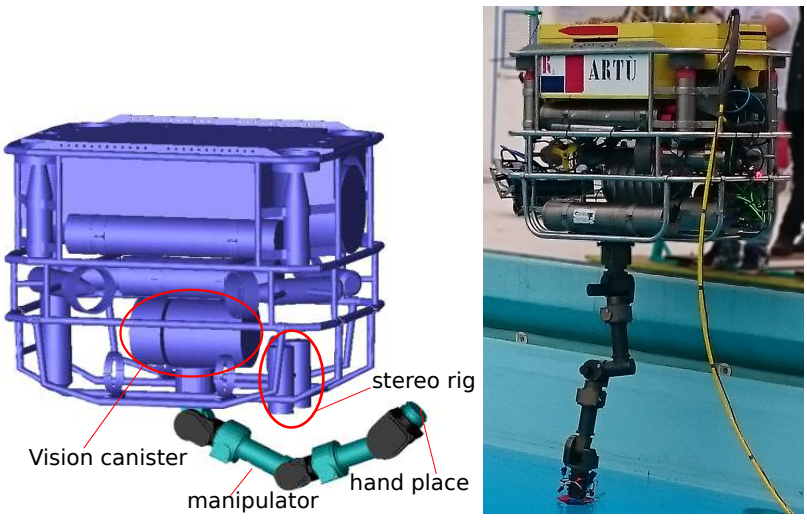
Given the complexity of underwater perception, there are few experimental works on autonomous underwater floating vehicles with manipulation capabilities. An extended survey on the -rather slow- evolution of these systems is presented in [9]. The main experiments span from pioneering endeavors like UNION [10] and AMADEUS [11] to more recent experiments like TRIDENT [12] and PANDORA [13].

TRIDENT and PANDORA projects are among the few attempts to perform autonomous underwater grasping and manipulation guided by computer vision. In [5] the underwater stereo vision system used in the TRIDENT European project is described. Object detection is performed by constructing a color histogram in HSV space of the target object, which is a patterned box that is always entirely visible even in the final grasping phase of object approaching. In the performed experiments, images of the underwater site are used to estimate the target object appearance and to train the object recognition algorithm. In the valve turning experiment of PANDORA project, a stereo camera is used to detect the position of a panel with a valve by using ORB features and by exploiting knowledge of the panel geometry [7].

The analysis of previous works highlights how strong assumptions are always made in underwater perception for robot manipulation, like full visibility of the target object during the task and the use of markers. Hence, underwater manipulation of complex and partially observable objects is still a challenging task. The main contribution of this paper is the reporting of a stereo vision system, capable of detecting partially visible pipes of known color, and its integration into an autonomous underwater vehicle.

3. Underwater System Overview: Setup and Calibration

In this section, we present the underwater robotic platform for interventions developed with the MARIS project [1] and the role of the underwater computer vision system. Communication among the different system modules (vision, vehicle, arm and robot hand) is a critical issue for the execution of manipulation tasks. In particular, the system modules communicate in real-time their state and the relative poses of their reference frames. For example, the vision subsystem requires the poses of the robotic arm joints, inertial measurements and the velocity of the AUV, and it provides the target object frame w.r.t. the



(a) (b)

Figure 1: (a) CAD model of AUV *Artù* with robot arm; (b) the actual vehicle before the start of a mission.

camera and the vehicle frames. Thus, the extrinsic calibration of the stereo camera is an essential step for proper system integration and correct execution of a manipulation task.

3.1. Underwater System Overview

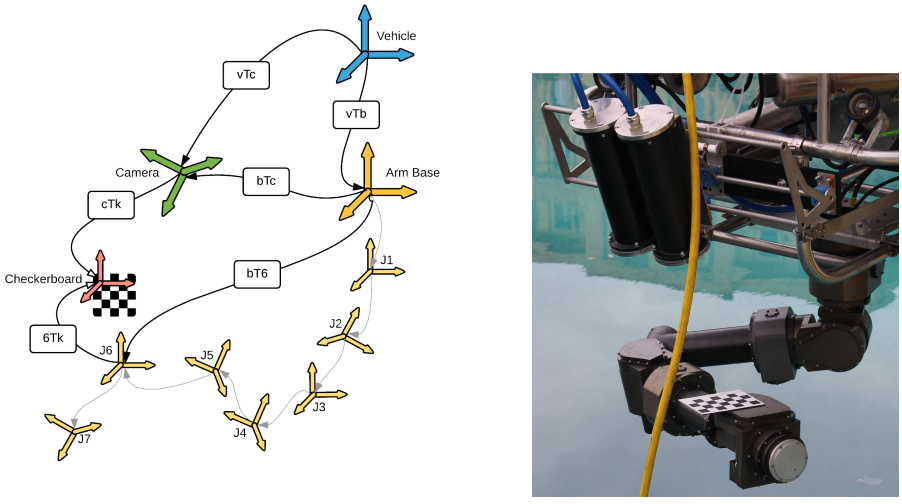
The MARIS AUV, named *Artù*, has been designed by CNR-ISSIA. The main part of the vehicle includes the buoyancy system, eight thrusters, batteries and a canister incorporating vehicle low-level control units, power distribution, sensors and network connections. The AUV has a modular design that simplifies integration by decoupling the main vehicle from the payload frame that contains all additional components, whose presence depends on the particular experiment the vehicle is involved in. Logistics and maintenance activities are also simplified thanks to the possibility of decoupling the vehicle from its payload. Six small headlights are mounted on the front of the AUV, although only two of them are pointed along the camera direction. When the scene is lighted up by natural light, the contribution of headlights is negligible.

The robotic arm provided by University of Genova has 7 degrees of freedom (d.o.f.) and is mounted at the bottom of the payload frame (Figure 1(a)). A gripper –designed and assembled by University of Bologna as an improvement of the one developed for the TRIDENT project [14]– is attached to the manipulator wrist. The robotic hand has three fingers and its kinematic configuration allows the execution of both power and precision grasps on objects with diameter up to 200 mm. The underactuated gripper has 8 degrees of freedom driven by three motors based on cable transmission. The gripper and the robot arm share the same control unit held in a very small canister.

The vision system consists of the vision canister and the stereo rig with the two camera canisters [15]. The two cameras are AVT Mako G125C GigE mounting a Sony ICX445 1/3" high resolution color sensor, capable of acquiring frames at 30 fps in full resolution (1292 $\times$ 964 pixels). For higher flexibility, varifocal lenses with a focal length between 4.4 mm and 11.0 mm have been chosen. The two cameras of the stereo vision system must be able to observe most of the workspace of the robotic arm, possibly from a point of view in which occlusions projected by the arm itself are reduced. Hence, the stereo rig was placed so that it sticks out one side of the vehicle, looking down with a small tilt angle. In robot manipulation experiments, the vision system is expected to work at distance of about $2m$ from the target; hence, the two cameras have been kept as close as possible with a resulting baseline of about $15cm$. Figure 1(b) shows the complete system, whereas Figure 2(b) illustrates the stereo camera rig.

During manipulation, the MARIS system is highly redundant (6 d.o.f. for the free-floating vehicle and 7 d.o.f. for the arm) and must be suitably controlled to achieve object grasping. Redundancy makes it possible to take into account visibility constraints (e.g. the target object terminals should be kept within FoV) and to coordinate different tasks [16, 17]. Thus, the vision system provides the reference frames of the target object, of its terminal part and of the grasping point.

3.2. Calibration



(a) (b)

Figure 2: Extrinsic calibration of the stereo camera w.r.t. the AUV system by means of a checkerboard: (a) reference frames; (b) picture of the checkerboard placed on arm joint 6.

Calibration of the vision system is a fundamental step of the integration

phase that enables reliable execution of manipulation and grasping tasks. The calibration procedure relies on the accuracy of the arm kinematics. The output of the calibration phase includes different sets of parameters: the *intrinsic* parameters, the *relative extrinsics* and the *system extrinsics*. The estimation of intrinsic and relative extrinsic parameters is a standard procedure for stereo vision cameras, usually performed by observing a checkerboard at slightly different poses. Due to light refraction through multiple materials, an underwater camera should be modeled as an axial camera [18]. However, an axial camera model is difficult to use with stereo processing (even for the geometric procedure illustrated in section 4.3). Hence, it is common practice in underwater computer vision to adopt a pinhole model with the parameters obtained by an in-water calibration using a checkerboard. The calibration parameters adopted in the experiments reported in this paper have been estimated according to this difficult and time-demanding in-water procedure. A less accurate alternative is to calibrate the stereo system before immersion and to optimize parameter values (in particular radial distortion) using ad hoc correction techniques [15] or to find a local approximation with the pinhole model [19].

The position and orientation of the stereo camera w.r.t. the vehicle and the robotic arm frames is given by six system extrinsic parameters. This transformation is required to plan the robot motion and to grasp the target object. Figure 2(a) illustrates all the reference frames. The transformation of a reference frame d w.r.t. a frame s, corresponding to homogenous matrix ${}^{s}_{d}T$, is labeled as sTd. The pose of a checkerboard k w.r.t. the vision reference frame c, namely ${}^{c}_{k}T$, can be obtained from the camera. Indeed, the checkerboard was rigidly fixed on joint 6 of the manipulator (with known fixed transformation ${}^{6}_{k}T$) and the pose of the checkerboard was estimated w.r.t. the manipulator base b using the arm kinematics. Figure 2(b) illustrates how the checkerboard is fixed on the arm during calibration. Given the transformations ${}^{c}_{k}T$ and ${}^{b}_{k}T$, the transformation ${}^{b}_{c}T$ of the camera w.r.t. the base is determined. This calibration procedure is sufficiently accurate and has enabled the manipulation experiments reported in section 5.

4. Object Detection and Pose Estimation

In this section, we illustrate the complete algorithm for the detection of the target object and for the estimation of its pose. As mentioned earlier, manipulation tasks require higher accuracy than localization, mapping, terrain classification, and even object detection and recognition. Moreover, computer vision in underwater environment is more challenging than in standard settings for robotic applications. Hence, to achieve successful underwater manipulation tasks the proposed system assumes that shape and color of the target object are known in advance, focusing on pipes with cylindrical shape and uniform color.

4.1. Pre-processing

The goal of image pre-processing is to remove disturbances due to the environment or the underwater robot system itself, and to enhance the quality

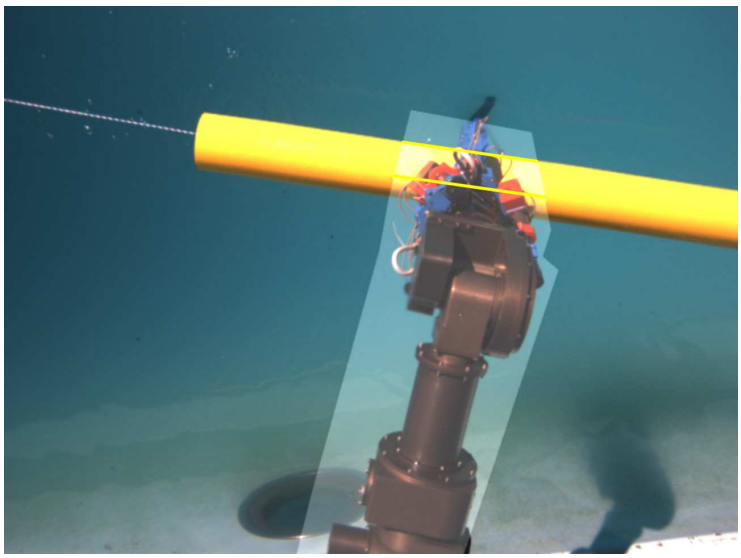


Figure 3: An example of robotic arm projection in the image.

of the image. The occlusion caused by the robot arm during manipulation is a predictable disturbance. While in the initial configuration the manipulator does not interfere with the two camera FoVs, the occlusion is almost unavoidable in the final phase of object grasping. To prevent wrong detection, a 3D model of the robot arm is projected in the image planes of left and right cameras using the values of arm joints and the arm base pose (Figure 2(a)). To this purpose a hierarchy of approximating bounding boxes is used to model the robot arm shape. The approximate shape of robot arm is projected in the image plane of the left camera to generate a mask as shown in Figure 3.

The object detection algorithm presented in Section 4.2 looks for regions with uniform color and regular contour. However, the water medium modifies colors due to light attenuation [6]. Instead of using an accurate light propagation model, color restoration is obtained according to *grey-world hypothesis* [20], which assumes that the average surface reflectance in a scene is achromatic. In particular, we assume that the maximum response in the channels is caused by a white patch. The mean values of RGB channels in the image, say $\bar{\mathbf{c}} = [\bar{r}, \bar{g}, \bar{b}]$, is computed. Given the maximum mean value $\bar{m} = \max\{\bar{r}, \bar{g}, \bar{b}\}$, the channels of i-th pixel $\mathbf{c}_i = [r_i, g_i, b_i]$ are corrected as

$$\hat{r}_i = \frac{r_i \bar{m}}{\bar{r}} \qquad \hat{g}_i = \frac{g_i \bar{m}}{\bar{g}} \qquad \hat{b}_i = \frac{b_i \bar{m}}{\bar{b}} \tag{1}$$

This approach allows an approximate but sufficient restoration of colors (Figure 4(a)-(b)) as required for ROI detection. Alternative approaches for image enhancement compensate attenuation discrepancy along the propagation path [21] accounting also for scattering and artificial lighting spatial distribution [22]. However, sophisticated techniques often rely on the assessment of

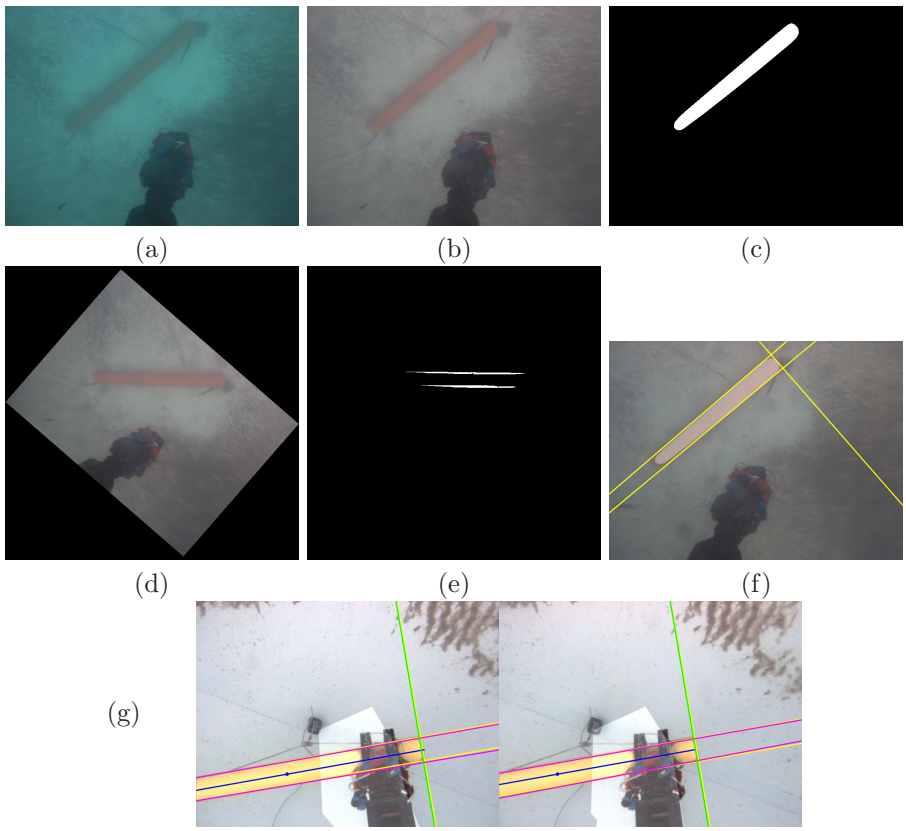


Figure 4: Outline of pipe detection algorithm: (a) input image of target object; (b) color restoration with grey-world hypothesis; (c) color ROI detection; (d) image rotation according to the main orientation given by the contours of the ROI; (e) horizontal line extraction in rotated image; (f) re-rotated lines in the original image. (g) An example of cylinder edge detection in left and right frames: yellow lines are used for the current measurement, whereas the purple and green lines are estimated by the edge trackers.

parameters depending on the relative position of lights and camera, and on the camera physical model. On-the-field calibration of such parameters is a demanding and time-consuming operation. Moreover, many advanced color correction techniques are unsuitable for online execution.

4.2. Object Detection

The object detection algorithm exploits the *a priori* (approximate) information about the color and the geometry of the target. Several approaches to object detection like popular keypoint feature matching or pattern identification fail in underwater conditions [8]. Although less general and requiring a proper compensation, color information enables the identification of the region-of-interest (ROI). Given the ROI, the object contour lines are searched by operating in a

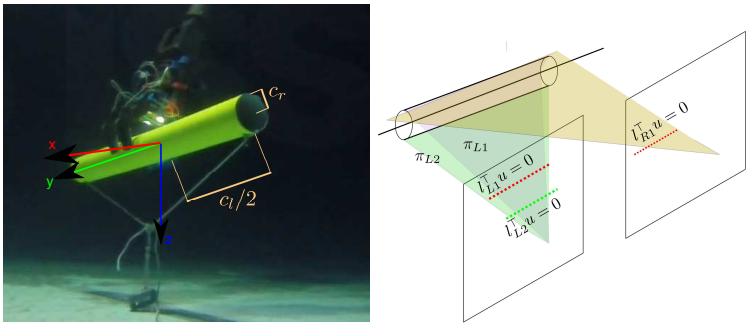
limited portion of the image. Since the cylindrical shape of the object is known, the contour of its projection in the image are contour lines. These two criteria are exploited by the method illustrated in the following.

The algorithm operates on the unmasked portion of the image and finds ROIs which satisfy the desired HSV channel constraints (Figure 4(c)). Each ROI has a main orientation given by its moments. A dominant orientation in the whole image is searched by extracting edges, by fitting segments using ED-Lines algorithm [23], and by building an orientation histogram with segments. The histogram maxima represent the candidate angles, which are matched with the ROI orientations. The selected dominant orientation corresponds to the histogram maximum that is closest to a ROI orientation. The image is rotated according to such dominant orientation 4(d). The rotated image simplifies the extraction of horizontal and vertical lines. The horizontal and vertical lines are extracted using respectively a vertical and a horizontal Sobel gradient operator. The horizontal lines correspond to the long border of the cylinder and the vertical ones to the terminal edges. Finally, the lines are re-rotated to the original image format (Figure 4(f)).

The extracted lines are not directly used for pose estimation. The accurate detection of line contours is critical for the pose estimation algorithm discussed in Section 4.3. As observed during preliminary tests, wrong detection of cylinder contours causes wrong estimation, which in turn prevents successful execution of grasps. To make detection robust, each edge is tracked frame-to-frame by an alpha-beta filter, which accepts the lines as input. Line state is represented using Hessian polar parameters (θ, ρ), and the update rates of state vector and state velocity are respectively $\alpha = 0.5$ and $\beta = 0.1$. Edge trackers improve the robustness of the estimation by preventing missing or wrong edge detection. The previously described procedure is performed for both left and right frames of the stereo camera. Figure 4(g) illustrates an example of the final output of object detection.

4.3. Pose Estimation

Robot manipulation requires accurate determination of the pose of the target object to be grasped w.r.t. the robot system. Dense stereo processing does not enable sufficiently robust and accurate estimation of object position and orientation in challenging underwater environments (e.g. see [24, 8]). Moreover, 3D point cloud processing is computationally demanding and the vision system must provide frequent evaluation of the target object pose to the control system (about 10 Hz). To satisfy these requirements in the underwater environment, the *a priori* information about the object shape is exploited. The reference frame of the object (see Figure 5(a)) can be conveniently placed provided that its position and orientation can be obtained from the recognizable parts in the images. In the proposed application, the target object has a cylindrical shape, whose main features in the image are its straight contour lines. Due to the symmetry of the cylinder, there are two d.o.f.'s that must be specified. The first d.o.f. to be specified is the position of the object frame origin on the cylinder axis, which can be set if the cylinder length is known and at least one cylinder



(a) (b)

Figure 5: (a) Target object frame: its origin is in the center of the cylinder, $\hat{y}$ is aligned with the cylinder axis, $\hat{z}$ is oriented with the gravity direction and $\hat{x}$ is consistent with right-handed convention. (b) Example of cylinder axis estimation from its edges in the left and right image planes.

end is visible. The second d.o.f. to be specified is due to radial symmetry of the object. Hence, the inertial sensor of the AUV is used to set the orientation angle of the object frame around the cylinder axis.

Pose computation of the cylinder can be performed from a single frame or from two frames acquired by the stereo vision system. In the first case, the radius of the cylinder must be known in order to correctly find the cylinder symmetry axis [25]. If the length of the cylinder c_l is also known, then the cylinder position is know completely. In particular, a cylinder is defined given the radius c_r and its axis, a line in the form $\mathbf{c}(t) = \mathbf{c}_p + \mathbf{c}_d\ t$. In the image plane, the cylinder contour is delimited by two lines with equations $\mathbf{l}_i^\top \mathbf{u} = 0$, where $\mathbf{u} = [u_x, u_y, 1]^\top$ is the pixel coordinate vector and $\mathbf{l}_1$, $\mathbf{l}_2$ are the line coefficients. The sign of each line coefficient vector l_i is conventionally chosen s.t. the cylinder lies in the positive half-plane $\mathbf{l}_i^\top \mathbf{u} > 0$. Figure 5(b) shows the projection of cylinder contour on left and right image planes. Let $\mathbf{P}_c$ be the projection matrix which projects 3D points in the camera coordinate frame to 2D pixel coordinates using intrinsic and distortion camera parameters. Each contour line is projected in the 3D space as a plane π_i with equation $\mathbf{n}_i^\top \mathbf{p} + c_i = 0$ where $[\mathbf{n}_i^\top | c_i]^\top = \mathbf{P}_c^\top \mathbf{l}_i$ and $\mathbf{p}$ a generic point in camera reference frame coordinates. Since $\mathbf{n}_i$ points orthogonally through the center of the cylinder, the axis line can be obtained by intersecting both translated planes of the radius c_r in $\hat{\mathbf{n}}_i$ direction, where $\hat{\mathbf{n}}_i = \mathbf{n}_i / \|\mathbf{n}_i\|$. Thus, the direction of the cylinder axis is given by direction vector $\mathbf{c}_d = \hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2$. Let π_{ep} be the plane $\mathbf{n}_{ep}^\top \mathbf{p} + c_{ep} = 0$ passing through one of the cylinder end-point with normal vector $\mathbf{n}_{ep}$ parallel to $\mathbf{c}_d$. The cylinder center point $\mathbf{c}_c$ is computed as

$$\begin{bmatrix} \mathbf{n}_1^\top \\ \mathbf{n}_2^\top \\ \mathbf{n}_{ep}^\top \end{bmatrix} \cdot \mathbf{p}_{ep} = \begin{bmatrix} -c_1 \\ -c_2 \\ -c_{ep} \end{bmatrix} \tag{2}$$

$$\mathbf{c}_c = \mathbf{p}_{ep} + \hat{\mathbf{n}}_{ep} \frac{c_l}{2} \tag{3}$$

The accuracy of this approach depends on the detection of object edges in the image and on the camera calibration parameters.

Stereo processing can be exploited in different ways. For example, a dense point cloud can be obtained by computing the disparity image from the two frames and the cylinder could be found through shape fitting techniques. Although this approach is rather general and can be applied to arbitrary shapes, it has several drawbacks. First, the point cloud density depends on the availability of reliable homologous points and, thus, on the color and pattern of the scene. Color uniformity facilitates the detection of the target object even in presence of blurred and poor light conditions of underwater environments [6], but it may result into empty regions in the disparity image. Second, the accuracy of stereo 3D estimation depends on the accuracy of camera calibration. To overstep these drawbacks, an approach similar to the single-frame method can be performed on a stereo image pair. The same cylindrical contour is projected in two different image planes resulting in four major contour lines. Each line is then reprojected in the 3D space resulting in four planes tangent to the cylinder. Let $\mathbf{P}_L$ and $\mathbf{P}_R$ be the projection matrices of the left and the right cameras respectively, π_{Li} and π_{Ri} the planes tangent to the cylinder obtained projecting in the 3D space the lines contour computed in both stereo images. Matrices $\mathbf{P}_L$ and $\mathbf{P}_R$ are referred to the same reference frame. The direction of the cylinder axis vector $\mathbf{c}_d$ must be orthogonal to each plane normal vectors $\hat{\mathbf{n}}_{Li}$ and $\hat{\mathbf{n}}_{Ri}$ respectively. Let $\mathbf{N}$ be the matrix whose rows are the normal vectors of planes through the cylinder axis

$$\mathbf{N} = \begin{bmatrix} \mathbf{n}_{L1}^\top \\ \mathbf{n}_{L2}^\top \\ \mathbf{n}_{R1}^\top \\ \mathbf{n}_{R2}^\top \end{bmatrix} \tag{4}$$

The cylinder axis direction $\mathbf{c}_d$ is estimated as the normalized vector that satisfies the homogeneous linear equation:

$$\mathbf{N} \cdot \mathbf{c}_d = 0 \tag{5}$$

Clearly, this method is consistent with the single frame case which requires just a cross product to compute the cylinder axis vector. Moreover, solving (5) is more robust in presence of noisy data as can be the projection of lines in 3D space. This method does not require all the projected planes. If there are only two planes with normals $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$, the cylinder axis is estimated like in the single frame case without loss of generality. Similarly to the single frame method, let π_{ep} be the plane passing through one of the cylinder end-point detected in one

of the stereo images. Then, the cylinder center point is obtained as

$$\begin{bmatrix} \mathbf{n}_{L1}^{\top} \\ \mathbf{n}_{L2}^{\top} \\ \mathbf{n}_{R1}^{\top} \\ \mathbf{n}_{R2}^{\top} \\ \mathbf{n}_{ep}^{\top} \end{bmatrix} \cdot \mathbf{p}_{ep} = \begin{bmatrix} -c_{L1} \\ -c_{L2} \\ -c_{R1} \\ -c_{R2} \\ -c_{ep} \end{bmatrix} \tag{6}$$

$$\mathbf{c}_c = \mathbf{p}_{ep} + \hat{\mathbf{n}}_{ep} \frac{c_l}{2} \tag{7}$$

The cylinder reference frame cannot be estimated without ambiguity due to the intrinsic symmetry of the cylinder. Several conventions can be exploited to attach a frame on the cylinder axis.

When the terminal edge of the cylinder does not lie in camera FoVs or is occluded by the robot arm, the method illustrated before is unable to estimate the pose, but it returns the cylinder axis. Moreover, sporadic errors in the edge extraction may cause wrong pose estimation, which affects the execution of grasping by the controller. To prevent wrong estimation and to exploit also partial estimation, we developed a pose tracking algorithm that provides the target pose $\mathbf{s}_t = [\mathbf{s}_{r,t}, \mathbf{s}_{q,t}]$ to the control system, where $\mathbf{s}_{r,t} \in \mathbb{R}^3$ is the position vector and $\mathbf{s}_{q,t} \in SO(3)$ the unit quaternion corresponding to the orientation w.r.t. the camera. The measurement $\mathbf{z}_t$ extracted from the stereo camera is either a frame or a frame fixed on the cylinder axis, when no terminal is observed. The measurements of AUV inertial sensor are used to predict the pose $\bar{\mathbf{s}}_t$ from previous target pose $\mathbf{s}_{t-1}$. In both cases, the algorithm computes the position and angular displacement of $\mathbf{z}_t$ with $\bar{\mathbf{s}}_t$. If the position or angular displacement are above a given threshold (in our experiments respectively 1 m and 30 deg), then the measurement is rejected and the previous estimation is kept $\mathbf{s}_t = \bar{\mathbf{s}}_t$. Otherwise, the novel object pose $\mathbf{s}_t$ is computed by interpolating $\bar{\mathbf{s}}_t$ and $\mathbf{z}_t$ as

$$\mathbf{s}_{r,t} = (1-\alpha)\bar{\mathbf{s}}_{r,t} + \alpha \mathbf{s}_{r,t} \qquad \mathbf{s}_{q,t} = Slerp(\alpha, \bar{\mathbf{s}}_{q,t}, \alpha \mathbf{z}_{q,t}) \tag{8}$$

where $\alpha \in [0,1]$ is an innovation coefficient and $Slerp()$ is the quaternion spherical linear interpolation. When only the cylinder axis is estimated, the frame origin $\mathbf{z}_{r,t}$ is computed by projecting $\bar{\mathbf{s}}_{r,t}$ on the estimated axis. More advanced approaches could have been used (e.g. an extended Kalman filter) at the expense of more accurate system modeling, calibration of additional parameters and careful management of outlier measurements. The proposed tracking algorithm provides a trade-off between robustness, real-time computation constraints and easiness in design and modeling.

5. Experiments

Experiments have been performed in an outdoor water pool with the system described in Section 3, with the goal of testing its underwater manipulation and grasping capabilities based on the developed vision system. Tests have

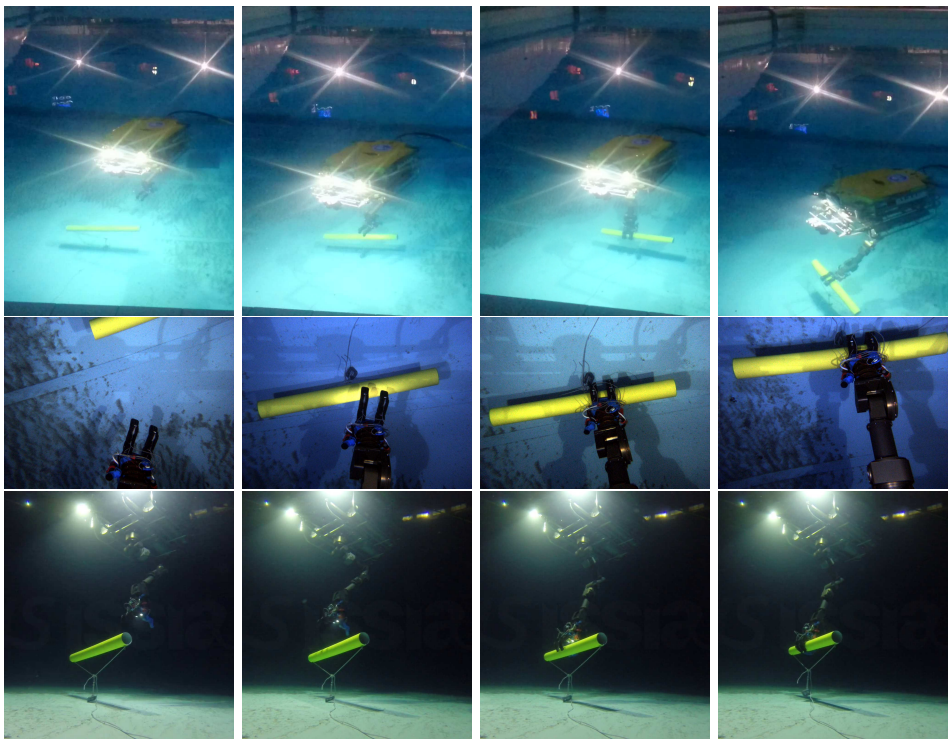


Figure 6: Successful pipe grasping sequence in night-time conditions, from pipe approach (left) to pipe recovery (right). Top line: view from the pool border. Center line: underwater images recorded by the AUV camera (post-processed to enhance scene visibility). Bottom line: frames recorded by camera mounted on a small ROV (VideoRay).

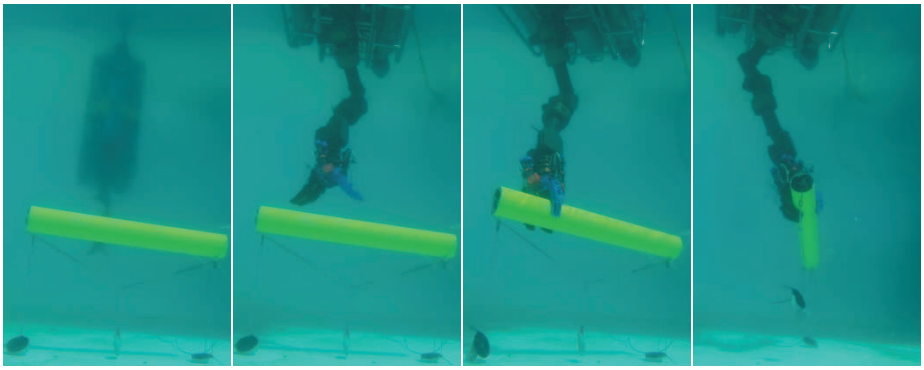
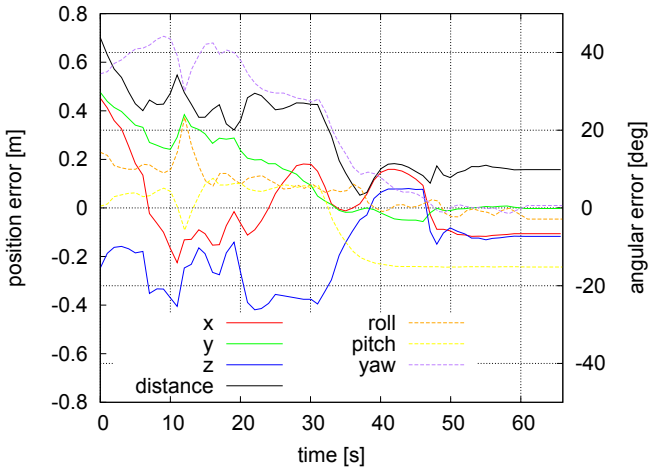
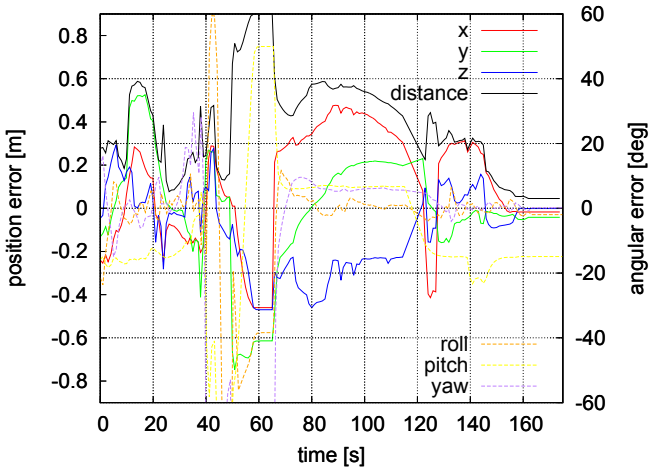


Figure 7: Successful pipe grasping sequence in daylight conditions.

taken place in different weather and light conditions (sunny, cloudy, night using vehicle lights), in reasonably clear, shallow water (from $3m$ to $4m$ depth). For additional safety in pool-based experiments, the pipe to be detected is suspended



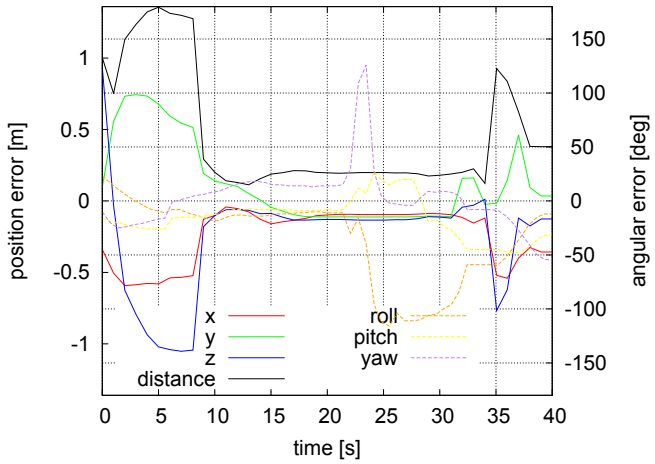
(a)



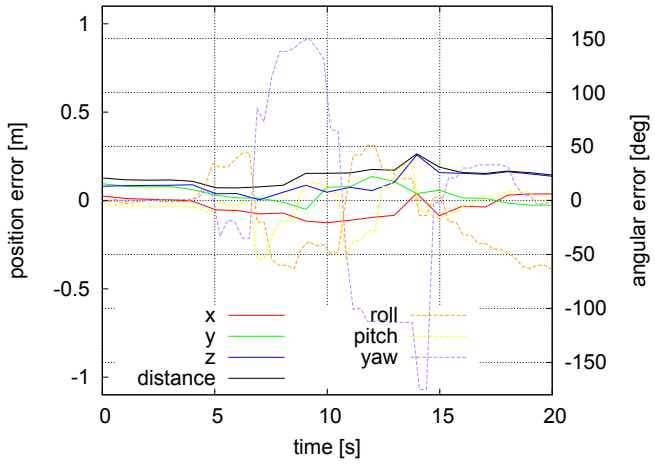
(b)

Figure 8: *Night trials*: robot end-effector position and angular errors with respect to the grasping goal in two pipe-approaching trials. Object is grasped at about time $53s$ in trial (a) and $162s$ in trial (b).

slightly above the floor. The pipe length is about $1m$ and its radius $5cm$. In the experiments reported in this section a yellow pipe is used as a target, although different colors have been used in other experimental sessions. At the beginning of the task the vehicle is operated in a remotely guided mode to explore the nearby environment until the target pipe is detected by the vision system. As soon as the pipe pose is consistently estimated, the vision system returns the goal frame for camera tracking and the vehicle becomes autonomous. The AUV



(a)



(b)

Figure 9: *Day-time trials*: robot end-effector position and angular errors with respect to the grasping goal in two pipe-approaching trials. Object is grasped at about time 14*s* in trial (a) and 4*s* in trial (b).

control switches to *free-floating* mode, which means that the system comprising vehicle and arm is considered as a single entity with appropriate goals. In particular, the free-floating control aims at reaching two objectives: to keep the left camera centered on its tracking reference frame, in order to maintain the best feasible view, and to let the robot end-effector reach the grasping goal frame.

Figures 6 and 7 show snapshot sequences of two successful pipe grasping executed respectively in night and daylight conditions. Videos of the experi-

ments can also be found in http://rimlab.ce.unipr.it/Maris.html. In particular, the night sequence shows the views from the swimming pool border, from the left camera of the AUV and from an underwater external camera. During the initial phase, the target object is not occluded by the robot arm, which stays in pre-grasping configuration, and a larger portion of the object is visible. Once the AUV approaches the target, automatic removal of the arm shape is essential to prevent error. Moreover, in the last stage the object is often not completely visible (e.g. Figure 4(g)) and in some frames the terminal edge of the pipe is either outside the camera FoV or occluded by the arm. The cylinder terminals are not visible on average $10 - 20\%$ of stereo camera frames and up to 80%, when the AUV approaches the target object along its axis. In these cases, the proposed pose estimation methods can only provide the cylinder axis and the tracking algorithm is essential to provide the reference frame based on previous estimation. In spite of these difficulties, the robot has successfully accomplished several vision-guided grasping trials in the different conditions discussed above.

Figures 8 and 9 illustrate the trend of position and angular errors between the arm end-effector frame and the goal frame in four successful grasping experiments. The results in Figure 8(a)-(b) refer to night conditions similar to Figure 6, whereas those in Figure 9(a)-(b) refer to daylight conditions similar to Figure 7. The contribution of AUV headlights to the scene luminance is important only during night-time experiments, whereas in presence of natural light it is negligible. In the latter case, the value of camera exposure is smaller than in the night-condition case, but the algorithm parameters have not been changed. Experiments have been successfully performed in both conditions.

In the grasping experiments, the goal frame is fixed w.r.t. the object frame (see Figure 5(a)) and close to the object terminal. All the errors (Figures 8 and 9) tend to decrease during a grasping trial as the robot approaches the object until their values are below a grasping tolerance set for the robotic hand. At that time a grasp command is issued by the AUV control system. After a successful grasp, the errors in pipe pose estimation tend to increase due to hand occlusion and object motions (see 9(a)-(b)), but this pose estimation is no longer required by the AUV control system. The residual values of position and angular errors are due to several causes, including inaccurate estimation of calibration parameters (camera frame w.r.t. robot base frame, end-effector frame w.r.t. joint, etc.), errors in object pose estimation and control system policy. Moreover, the tolerance on the pitch angle, i.e. the angle around the cylinder axis, is higher due to the symmetry of the object. In all trials, the roll and yaw errors are less than 3°. The position displacement is about or less than $15cm$ for all trials at the robotic hand grasp. The accuracy obtained in pipe pose estimation has proven sufficient to achieve object grasps with the MARIS AUV system.

6. Conclusion

In this work, a stereo vision system was presented for an AUV devoted to grasping and transportation of pipes, typical tasks of the underwater off-shore

industry. Successful underwater grasping experiments have been reported in different light conditions, even in the case of partially observable objects.

In future work, it is planned to perform experiments using visual feedback in more challenging tasks such as cooperative floating manipulation between two AUVs that require coordination and exchange of information about the common object to be grasped and the state of the vehicles.

Acknowledgment

This research has been funded by the Italian PRIN 2010FBLHRJ 007 through the research project MARIS. The authors would like to thank all the MARIS consortium for the fruitful collaboration.

7. References

[1] G. Casalino, M. Caccia, S. Caselli, C. Melchiorri, G. Antonelli, A. Caiti, G. Indiveri, G. Cannata, E. Simetti, S. Torelli, A. Sperindè, F. Wanderlingh, G. Muscolo, M. Bibuli, G. Bruzzone, E. Zereik, A. Odetti, E. Spirandelli, A. Ranieri, J. Aleotti, D. Lodi Rizzini, F. Oleari, F. Kallasi, G. Palli, U. Scarcia, L. Moriello, E. Cataldi, Underwater Intervention Robotics: An Outline of the Italian National Project MARIS, The Marine Technology Society Journal 50 (4) (2016) 98–107.

[2] M. Narimani, S. Nazem, M. Loueipour, Robotics vision-based system for an underwater pipeline and cable tracker, in: Proc. of the IEEE/MTS OCEANS, 2009, pp. 1–6.

[3] R. Eustice, H. Singh, J. Leonard, M. Walter, R. Ballard, Visually Navigating the RMS Titanic with SLAM Information Filters, in: Proceedings of Robotics: Science and Systems, 2005, pp. 57–64.

[4] M. Johnson-Roberson, O. Pizarro, S. Williams, I. Mahon, Generation and Visualization of Large-Scale Three-Dimensional Reconstructions from Underwater Robotic Surveys, Journal of Field Robotics 27 (1) (2010) 21–51.

[5] M. Prats, J. Garcia, S. Wirth, D. Ribas, P. Sanz, P. Ridao, N. Gracias, G. Oliver, Multipurpose autonomous underwater intervention: A systems integration perspective, in: Mediterranean Conference on Control & Automation, 2012, pp. 1379–1384.

[6] S. Bazeille, I. Quidou, L. Jaulin, Color-based underwater object recognition using water light attenuation, Intel Serv Robotics 5 (2012) 109–118.

[7] M. Carreras, A. Carrera, N. Palomeras, D. Ribas, N. Hurtós, Q. Salvi, P. Ridao, Intervention Payload for Valve Turning with an AUV, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer International Publishing, 2015, pp. 877–884.

[8] D. Lodi Rizzini, F. Kallasi, F. Oleari, S. Caselli, Investigation of vision-based underwater object detection with multiple datasets, International Journal of Advanced Robotic Systems (IJARS) 12 (77) (2015) 1–13.

[9] J. Pérez, J. Sales, A. Peñalver, J. Fernández, P. Sanz, J. García, J. Martí, R. Marín, D. Fornas, Robotic Manipulation Within the Underwater Mission Planning Context, in: G. Carbone, F. Gomez-Bravo (Eds.), Motion and Operation Planning of Robotic Systems, Vol. 29 of Mechanisms and Machine Science, Springer International Publishing, 2015, pp. 495–522.

[10] V. Rigaud, E. Coste-Maniere, M. Aldon, P. Probert, M. Perrier, P. Rives, D. Simon, D. Lang, J. Kiener, A. Casal, J. Amar, P. Dauchez, M. Chantler, UNION: underwater intelligent operation and navigation, IEEE Robotics & Automation Magazine 5 (1) (1998) 25–35.

[11] D. Lane, J. Davies, G. Casalino, G. Bartolini, G. Cannata, G. Veruggio, M. Canals, C. Smith, D. O'Brien, M. Pickett, G. Robinson, D. Jones, E. Scott, A. Ferrara, M. Coccoli, R. Bono, P. Virgili, R. Pallas, E. Gracia, AMADEUS: advanced manipulation for deep underwater sampling, IEEE Robotics Automation Magazine 4 (4) (1997) 34–45.

[12] P. Sanz, P. Ridao, G. Oliver, G. Casalino, Y. Petillot, C. Silvestre, C. Melchiorri, A. Turetta, TRIDENT An European project targeted to increase the autonomy levels for underwater intervention missions, in: Proc. of the IEEE/MTS OCEANS, 2013, pp. 1–10.

[13] D. Lane, F. Maurelli, P. Kormushev, M. Carreras, M. Fox, K. Kyriakopoulos, PANDORA - Persistent Autonomy Through Learning, Adaptation, Observation and Replanning, Proc. of the World Congr. of the International Federation of Automatic Control (IFAC) 48 (2) (2015) 238 – 243.

[14] D. Ribas, P. Ridao, A. Turetta, C. Melchiorri, G. Palli, J. J. Fernandez, P. J. Sanz, I-AUV Mechatronics Integration for the TRIDENT FP7 Project, IEEE/ASME Transactions on Mechatronics 20 (5) (2015) 2583–2592.

[15] F. Oleari, F. Kallasi, D. Lodi Rizzini, J. Aleotti, S. Caselli, An underwater stereo vision system: from design to deployment and dataset acquisition, in: Proc. of the IEEE/MTS OCEANS, 2015, pp. 1–5.

[16] E. Simetti, G. Casalino, N. Manerikar, A. Sperind, S. Torelli, F. Wanderlingh, Cooperation between autonomous underwater vehicle manipulations systems with minimal information exchange, in: Proc. of the IEEE/MTS OCEANS, 2015, pp. 1–6.

[17] E. Simetti, G. Casalino, A novel practical technique to integrate inequality control objectives and task transitions in priority based control, Journal of Intelligent & Robotic Systems (2016) 1–26.

[18] A. Agrawal, S. Ramalingam, Y. Taguchi, V. Chari, A theory of multi-layer flat refractive geometry, in: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 3346–3353.

[19] T. Dolereit, U. von Lukas, A. Kuijper, Underwater stereo calibration utilizing virtual object points, in: Proc. of the IEEE/MTS OCEANS, 2015, pp. 1–7.

[20] G. Buchsbaum, A spatial processor model for object colour perception, Journal of the Franklin institute 310 (1) (1980) 1–26.

[21] H. Lu, Y. Li, L. Zhang, S. Serikawa, Contrast enhancement for images in turbid water, J. Opt. Soc. Am. A 32 (5) (2015) 886–893.

[22] M. Bryson, M. Johnson-Roberson, O. Pizarro, S. B. Williams, True color correction of autonomous underwater vehicle imagery, Journal of Field Robotics, 33 (6) (2015) 735–874.

[23] C. Akinlar, C. Topal, Edlines: A real-time line segment detector with a false detection control, Pattern Recognition Letters 32 (13) (2011) 1633–1642.

[24] F. Oleari, F. Kallasi, D. Lodi Rizzini, J. Aleotti, S. Caselli, Performance Evaluation of a Low-Cost Stereo Vision System for Underwater Object Detection, in: Proc. of the World Congr. of the International Federation of Automatic Control (IFAC), 2014, pp. 3388–3394.

[25] F. Kallasi, D. Lodi Rizzini, F. Oleari, J. Aleotti, Computer vision in underwater environments: a multiscale graph segmentation approach, in: Proc. of the IEEE/MTS OCEANS, 2015, pp. 1–6.

Dario Lodi Rizzini received the PhD degree from the University of Parma in 2009, and since 2015 is assistant professor at the Information Engineering Department of the same University. He has been a visiting scholar at the University of Freiburg in 2007. His research interests include localization, mapping and navigation of mobile robots, 3D perception and object detection.

Fabjan Kallasi received the Master degree in Computer Engineering from the University of Parma in 2013, and is currently a PhD candidate at the same institution. His main research interests are in the area of computer vision and advanced perception for robot navigation, with particular focus on ground and underwater robotics and automation.

Jacopo Aleotti is fixed-time assistant professor at the University of Parma with the Department of Information Engineering. He owns a Laurea degree in Electronic Engineering and a Ph.D. degree in Information Technologies. In 2003, he was Marie Curie Fellow at the Learning System Laboratory, Orebro University, Sweden. His research interests include robot learning, robot manipulation, range sensing and virtual reality.

Fabio Oleari received the PhD degree from the University of Parma in 2016, and is currently member of the R&D division of the Elettric80 Spa company. His main research interests are in computer vision, embedded processing, and sensor fusion for object detection and robot navigation. His research is focused on industrial and field robotics applications.

Stefano Caselli is full professor at the Department of Information Engineering of the University of Parma and director of the Robotics and Intelligent Machines Laboratory. His scientific interests include autonomous robots, real-time architectures and advanced robot programming paradigms. Current research projects focus on novel aerial, underwater, and agricultural robot applications as well as on robot perception in industrial automation.