Active stereovision using invariant visual servoing

Vincent Brandou
and Anne-Gaelle Allais
and Michel Perrier
IFREMER Centre de Mediterranee - BP 330
FR-83507 La Seyne-Sur-Mer Cedex
Email: Vincent.Brandou@ifremer.fr

Ezio Malis
and Patrick Rives
INRIA Sophia-Antipolis
2004 route des lucioles - BP 93
FR-06902 Sophia Antipolis

***Abstract*— The objective of this paper is to propose an innovative visual servoing method in order to improve the 3D reconstruction of objects for quantitative measurements. The method uses a stereovision system that allows us to obtain various shots of an object, at regular intervals according to a predefined trajectory. In our case, the stereo rig is equipped with two different cameras, the first one is fixed while the other one is mounted on pan and tilt. So, the intrinsic parameters are not the same for the two cameras. To validate our approach, first experiments have been conducted by simulation and under laboratory conditions.**

I. INTRODUCTION

The aim of our research is to develop, implement and test an original robotic method dedicated to compute a metric 3D reconstruction in order to describe and quantify biodiversity in deep-sea fragmented habitats [17]. The objective consists in applying techniques of computer vision to create tools adapted to the exploitation of underwater images. Today, some deep-sea vehicles are equipped with measuring devices and a manipulator arm. Vision systems are often used to complete information provided by acoustic sensors [12]. Taking advantage of possibilities allowed by underwater vehicles, our goal is to develop a methodology based on a vision system, to obtain quantitative measurements through a 3D reconstruction of underwater scenes.

The images used for the reconstruction are acquired when the vehicle is deployed on the sea floor at a fixed and stable attitude. The images are subject to several constraints linked to the underwater environment. First of all, the observed scenes are unknown, and the objects to be reconstructed in these scenes are made up of random textures and shapes. We only know that the objects are rigid and have a vertical overall shape. Moreover, the refraction, the presence of particles, the absorption and the problems of lighting in an underwater environment considerably alter the image quality. Noisy images and an unknown model of the object have a knock-on effect on the 3D reconstruction accuracy. The idea is thus to define a trajectory adapted to the type of object in order to improve the reconstruction accuracy. The originality of the method lies in the use of an active stereo rig (i.e. with variable geometry) mounted on a 6 DOF manipulator arm effector, which equips the underwater vehicle. The stereo rig is equipped with two different underwater cameras, the first one is fixed while the other one is mounted on pan & tilt. The geometry configuration of the stereo rig allows us to generate predefined trajectories and to acquire images at regular intervals. Thus, the geometry configuration is fixed in function of the desired task. This constraint yields to a reduction of unknown variables in the reconstruction process by bundle adjustment technique [19].

We show hereafter how the stereo rig geometry determines the trajectory and the number of shots taken around the object to be reconstructed. Consequently, we carried out research related to the use of visual servoing techniques to improve the 3D object reconstruction. Many approaches have already been proposed to solve the problem of vision-based control [2], [3], [14]–[16]. All of them are based on the same scheme: target definition, image feature selection, 2D/3D modeling, robot geometric modeling and Jacobian matrix calculation. But even if the principle basis is the same, the methods vary according to the prior knowledge on the environment and the sensors.

In this paper, we propose a visual servoing method adapted to the stereo rig in order to improve a 3D reconstruction. Indeed, the intrinsic parameters of the two cameras are not the same. Moreover, they are influenced by the characteristics of the underwater environment such as the optical index, which varies as a function of temperature, salinity, pressure, or wavelength [12]. For example, the underwater vehicle used for the final experiments is able to dive up to a depth of 6000 meters. In applications using different cameras, the intrinsic parameters should be determined by a calibration process. It is quite difficult to use a calibration pattern to calibrate the cameras in such conditions. On the other hand, we would like to avoid on-line self-calibration techniques which imply new constraints on the robot displacement (i.e. a strong rotation of the stereo rig is needed). Consequently, we have implemented a visual servoing method which is invariant to camera intrinsic parameters [8]. The visual servoing scheme proposed in this paper has been completely validated by experiments.

II. TRAJECTORIES INDUCED BY THE GEOMETRY OF THE STEREO RIG

Visual control is carried out with a system of stereovision mounted at the tip of an instrumented arm of a robot (eye-in-hand robotic system). It consists in capturing a reference image with the right camera on a given position, and then converging towards this position with the left camera (see Figure (1)). We have shown that setting the geometry between both cameras is equivalent to moving the coordinate system associated to each camera on the surface of a cylinder. The demonstration is based on the integration of infinitesimal displacements to obtain a discrete displacement.

Let $\mathbf{T}$, defined in the Euclidean special group of dimension three $SE(3)$, be the transformation matrix between the

position of the two cameras (see Figure (2)):

$$\mathbf{T} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \qquad \mathbf{T} \in SE(3) \tag{1}$$

where $\mathbf{R}$ is the (3×3) rotation matrix, and $\mathbf{t}$ is the (3×1) translation vector. The rotation matrix can be written as a function of α:

$$\mathbf{R}(\alpha) = e^{[\mathbf{r}(\alpha)]_\times} \tag{2}$$

where $\mathbf{r}$ is the rotation vector corresponding to an angle $\alpha \in \mathbb{R}$ around a fixed axis specified by the vector $\boldsymbol{\delta}$:

$$\mathbf{r}(\alpha) = \frac{\boldsymbol{\delta}}{||\boldsymbol{\delta}||}\alpha \tag{3}$$

with:

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_{tilt} \\ \delta_{pan} \\ 0 \end{pmatrix} \tag{4}$$

Similarly, the translation vector depends on α and the distance l between the cameras:

$$\mathbf{t}(\alpha, l) = -\mathbf{R}(\alpha) \begin{pmatrix} l \\ 0 \\ 0 \end{pmatrix} = -e^{[\mathbf{r}(\alpha)]_\times} \begin{pmatrix} l \\ 0 \\ 0 \end{pmatrix} \tag{5}$$

$\mathbf{T}$ can always be written as:

$$\mathbf{T} = e^{\mathbf{A}} \qquad \mathbf{A} \in se(3) \tag{6}$$

where $se(3)$ is the Algebra of Lie associated to the group of Lie, and:

$$\mathbf{A} = \begin{pmatrix} [\boldsymbol{\omega}]_\times & \boldsymbol{\nu} \\ \mathbf{0} & 0 \end{pmatrix} \tag{7}$$

where $\boldsymbol{\omega}$ is the (3×1) rotation speed vector, such that $||\boldsymbol{\omega}|| < \pi$, and $\boldsymbol{\nu}$ is a (3×1) translation speed vector. Let $[\boldsymbol{\omega}]_\times$ be the (3×3) skew matrix associated to vector $\boldsymbol{\omega}$.

An infinitesimal displacement of the frame origin $\mathcal{O}$, which coincides with the center of projection $\mathcal{C}$ on the trajectory according to time t, standardized between 0 and 1, is defined by:

$$\dot{\mathcal{C}}(t) = [\boldsymbol{\omega}]_\times \mathcal{C}(t) + \boldsymbol{\nu} \tag{8}$$

It can be shown that the integral of this equation can be written in the general form of the parametric equations of a cylinder:

$$\mathcal{C}(t) = \boldsymbol{r}_0 + cos(\theta)\boldsymbol{u} + sin(\theta)\boldsymbol{v} + \boldsymbol{w}t \tag{9}$$

with:

$$\theta = ||\boldsymbol{\omega} t||, \qquad \boldsymbol{r}_0 = \frac{\boldsymbol{\gamma}}{||\boldsymbol{\omega}||^2},$$

$$\boldsymbol{u} = \begin{pmatrix} -\frac{\boldsymbol{\gamma}}{||\boldsymbol{\omega}||^2} \end{pmatrix}, \ \boldsymbol{v} = \begin{pmatrix} -\frac{\boldsymbol{\zeta}}{||\boldsymbol{\omega}||^3} \end{pmatrix}, \ \boldsymbol{w} = \begin{pmatrix} \boldsymbol{\nu} + \frac{\boldsymbol{\zeta}}{||\boldsymbol{\omega}||^2} \end{pmatrix}$$

where:

$$\boldsymbol{\gamma} = \begin{pmatrix} -\nu_y\omega_z + \nu_z\omega_y \\ \nu_x\omega_z - \nu_z\omega_x \\ -\nu_x\omega_y + \nu_y\omega_x \end{pmatrix}$$

$$\boldsymbol{\zeta} = \begin{pmatrix} -\nu_x(\omega_y^2 + \omega_z^2) + \nu_y\omega_x\omega_y + \nu_z\omega_x\omega_z \\ \nu_x\omega_x\omega_y - \nu_y(\omega_z^2 + \omega_x^2) + \nu_z\omega_y\omega_z \\ \nu_x\omega_x\omega_z + \nu_y\omega_y\omega_z - \nu_z(\omega_y^2 + \omega_x^2) \end{pmatrix}$$

The three vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ are unit and form a direct trihedron. Including $\boldsymbol{r}_0$, they characterize the cylinder which defines the trajectory. Since $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ and $\boldsymbol{r}_0$ depend on the various components of $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$, which depend themselves on the transformation matrix $\mathbf{T}(\mathbf{R}, \mathbf{t})$ linking the position of the two cameras (see Equations 1,6,7), we can say that the stereo rig geometry induces the trajectory.

Because point $\boldsymbol{r}_0$ is located on the cylinder axis, if we set the geometry of the stereo rig, we can also determine the distance to be applied between the cameras and the object under study.

Hence, it is possible to choose the geometry of the stereo rig with respect to the shape, the volume, and the orientation of the underwater object to be reconstructed. For example, with a fixed geometry of the stereo rig, the trajectory corresponds to a straight line in case of parallel cameras (no angle), a circle if we apply a pan angle, or an helicoid in case of a pan and tilt angle. For instance (see Figure (1)), a pan angle α and a distance l force the stereo rig to describe a circular trajectory. We can also compose more complex trajectories if the geometry of the stereo rig is changed during the visual servoing.

Moreover, if one geometric configuration of the stereo head leads to one trajectory, a particular trajectory can be induced by various value combinations of angles and baseline lengths which have only an effect on the number of shots taken along the trajectory. It is thus necessary to ensure an overlapping of the camera fields of view to make possible visual servoing. Concerning the number of views necessary for the 3D reconstruction, we can mention research carried out by [13], [18] which makes possible to select a relevant subset of images from image sequences to estimate the epipolar geometry between the views and then to compute the 3D reconstruction. But these methods are only adapted for post-processing when a great deal of images have already been acquired. So they cannot be applied to determine the stereo rig geometry which is a requirement to visual servoing and image acquisition.

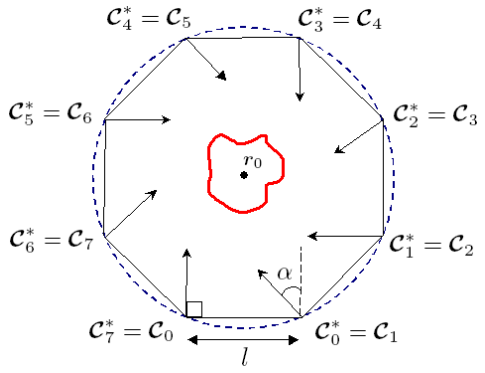


Fig. 1. Trajectory induced by the stereo rig geometry

III. Visual servoing approach

The aim of visual servoing is to control the movement of the robot's end-effector using the information provided by vision sensors. A typical task consists in putting an "eye-in-hand" system in a given position with respect to an observed object [10]. Most visual servoing techniques are based on a "teaching-by-showing" approach [6]. In our case, given a reference image $\mathcal{I}^*$ taken by the right camera at the reference

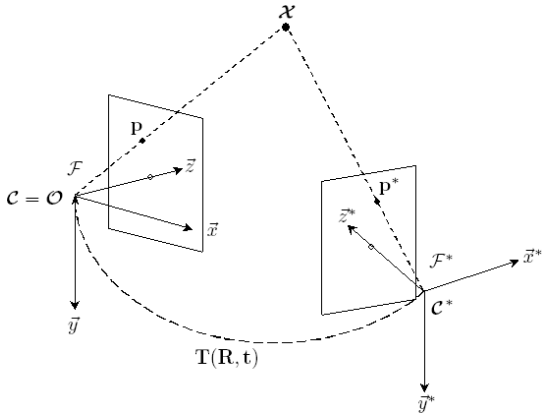


Fig. 2. Perspective projections of a 3D point $\boldsymbol{\mathcal{X}}$ on the two stereo rig cameras

position $\mathcal{F}^*$, the goal is to reach the same position with the left camera.

Let us consider the pinhole camera model to perform a perspective projection of a 3D point $\boldsymbol{\mathcal{X}} = (X, Y, Z, 1) \in \mathbb{P}^3$ to a virtual point $\mathbf{m} = (x, y, 1) \in \mathbb{P}^2$ in the frame $\mathcal{F}$ from the center of projection $\boldsymbol{\mathcal{C}}$ (see Figure (2)). The relationship is defined by:

$$\zeta \mathbf{m} = \frac{1}{Z} [\, \mathbf{R} \;\; \mathbf{t} \,] \boldsymbol{\mathcal{X}} \tag{10}$$

where ζ is the positive depth, $\mathbf{R}$ and $\mathbf{t}$ are respectively the rotation and the translation between frames $\mathcal{F}^*$ and $\mathcal{F}$ which set the stereo rig geometry. Point $\mathbf{m}$ gives the corresponding point $\mathbf{p} = (u, v, 1)$ measured in pixels in image $\mathcal{I}$:

$$\mathbf{p} = \mathbf{K}\mathbf{m} \qquad \mathbf{p} \in \mathcal{I}(\mathcal{F}, \mathbf{K}) \tag{11}$$

where $\mathbf{K}$ is the intrinsic parameter matrix of the left camera:

$$\mathbf{K} = \begin{bmatrix} f & s & u_0 \\ 0 & rf & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

where u_0 and v_0 are the coordinates of the principal point (in pixels), f is the focal length (in meters), s is the skew and r is the aspect ratio. In the same way, points $\mathbf{p}^*$ are obtained in frame $\mathcal{F}^*$, but with different intrinsic parameters if the cameras are not the same.

Generally, using "camera-dependent" visual servoing methods [9] implies that the same camera is used to learn the reference image and to perform the visual servoing. If the cameras are different, the intrinsic parameters of these cameras must be precisely identified. Otherwise, the current image $\mathcal{I}$ converges towards the reference image $\mathcal{I}^*$, but with different camera positions. Note that the intrinsic parameters may significantly vary during the life of a vision system, and they can be changed intentionally when using zooming cameras [10]. We saw in section I, that it is difficult to calibrate cameras under underwater experimental conditions. So, we have used a visual servoing method allowing us to compute an error function invariant to camera intrinsic parameters $\mathbf{K}$ [8].

Let consider n non-collinear 3D points of the observed object. These points are projected respectively in frames $\mathcal{F}^*$ and $\mathcal{F}$ to give points $\mathbf{p}_i^* = (u_i^*, v_i^*, 1)$ and $\mathbf{p}_i = (u_i, v_i, 1) \;\; \forall i \in 1, 2, \ldots, n$. The latter are then projected into spaces $\mathcal{Q}^*$ and $\mathcal{Q}$ invariant to intrinsic parameters. To carry out these projections, we need to compute the following (3×3) symmetric matrix:

$$\mathbf{S}_p = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p}_i \mathbf{p}_i^\top \tag{13}$$

If the observed points are not collinear and $n > 3$, matrix $\mathbf{S}_p$ is symmetric positive definite and can be written, using a Cholesky decomposition:

$$\mathbf{S}_p = \mathbf{T}_p \mathbf{T}_p^\top \tag{14}$$

where $\mathbf{T}_p$ is a (3×3) non-singular upper triangular matrix, which allows us to compute points $\mathbf{q}_i \in Q$:

$$\mathbf{q}_i = \mathbf{T}_p^{-1} \mathbf{p}_i = (a_i, b_i, 1) \tag{15}$$

In the same way, points $\mathbf{q}_i^*$ are computed with points $\mathbf{p}_i^*$. These new points are then contained in two vectors: $\mathbf{s}^* = (\mathbf{q}_1^*, \mathbf{q}_2^*, \ldots, \mathbf{q}_n^*)$ and $\mathbf{s} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n)$. The camera has converged to the reference position when $\mathbf{s} = \mathbf{s}^*$. The derivative of vector $\mathbf{s}$ is written:

$$\dot{\mathbf{s}} = \mathbf{L}\mathbf{v} \tag{16}$$

where $\mathbf{L}$ is the (3n×6) matrix of interaction, and the (6×1) vector $\mathbf{v}$ represents the Cartesian velocity of the camera.

The task function is:

$$\mathbf{e} = \widehat{\mathbf{L}}^+ (\mathbf{s} - \mathbf{s}^*) \tag{17}$$

where $\widehat{\mathbf{L}}^+$ is an approximation of the pseudo-inverse of $\mathbf{L}$ since $\mathbf{K}$ and ζ are unknown.

In order to control the movement of the camera by imposing the exponential convergence of the task function, the control law is:

$$\mathbf{v} = -\lambda \mathbf{e}, \quad \text{where} \;\; \lambda > 0 \tag{18}$$

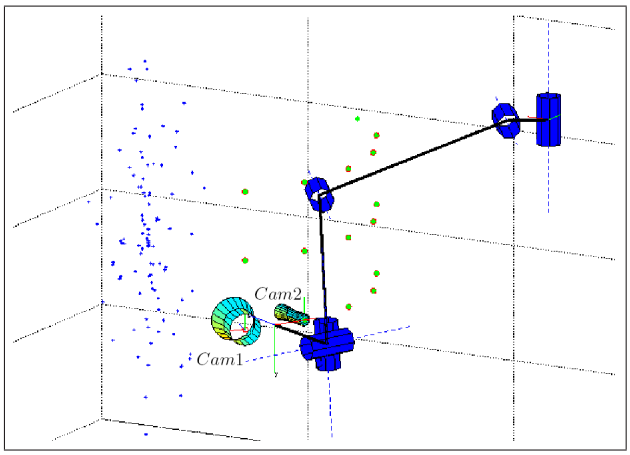
This method has been tested with a stereo system, first by simulation, and afterwards by experiment, in order to generate large trajectories by repeating visual servoing. As it is shown in [9], the control law is stable even in the presence of calibration errors.

IV. Simulation

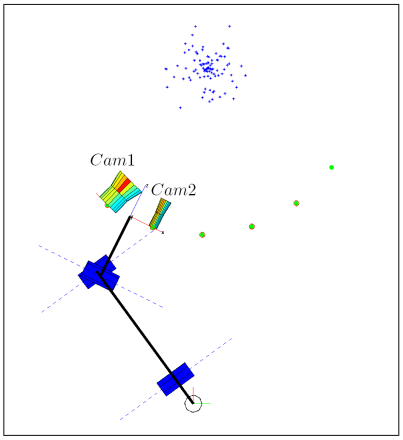
The simulation allowed us to test the complete process of visual servoing. We have taken into account the real dimensions and parameters of the cameras (sensor dimension, pixel size and focal length), and of the robotic arm (such as joint limits, kinematic singularities, maximum articular velocities, axes dimension). The simulation of the arm makes possible to determine which trajectories can be performed within the robot workspace.

Figure (3) represents a trajectory simulation of the stereo rig carried by the robotic arm of the underwater vehicle. In Figure (3)(a) and Figure (3)(b), camera $Cam1$ is mounted on pan and tilt and is used to acquire reference images whereas camera $Cam2$ is fixed and is controlled during visual servoing. Both cameras have different intrinsic parameters. In these figures, the scatter plot stands for the 3D object such that it fits the shape of an underwater hydrothermal vent.

The camera trajectories, hence the stereo rig geometry, are chosen according to the size and the shape of the object under study. In this example, the trajectory is induced by a geometry of 0.3 meters between the cameras and a pan angle

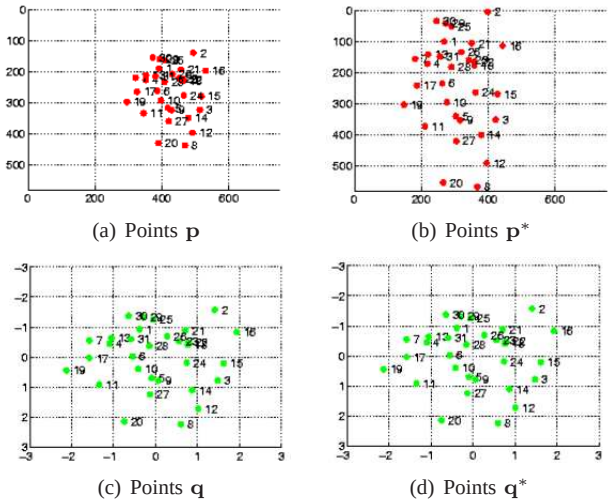


(a) side view



(b) top view

Fig. 3. Simulation of trajectories



(a) Points $\mathbf{p}$ (b) Points $\mathbf{p}^*$

(c) Points $\mathbf{q}$ (d) Points $\mathbf{q}^*$

Fig. 4. Projection of the scatter plot seen by the two cameras (red) in the invariant space (green)

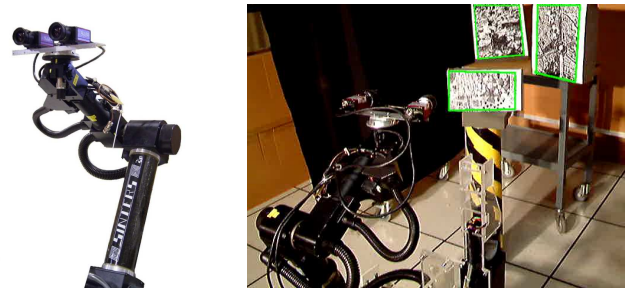
of 18 degrees. Figure (3)(a) and Figure (3)(b) show the side and the top view of the trajectory used for the simulation. The trajectory is marked by the successive positions of the reference camera in green points. It is composed of three quarters of a circle in order to cover the whole height of the object. To ensure the continuity of the trajectory, the stereo rig is U-turned at the end of a quarter of a circle. We can see that the arm cannot be manoeuvred all around the object if the vehicle is at a fixed location. So, to make the 3D reconstruction of an overall object, the underwater vehicle has to be landed at several positions.

Figures (4)(a) and (b) represents the scatter plot seen by both cameras, at the end of visual servoing once the controlled camera has converged towards the reference position. The picture on the right is image $\mathcal{I}^*$ taken by the reference camera whereas the picture on the left is image $\mathcal{I}$ taken by the controlled camera. We can notice that, because the intrinsic parameters are not the same for the two cameras, points $\mathbf{p}$ do not match points $\mathbf{p}^*$.

On the contrary, Figures (4)(c) and (d), the green scatter plot represents points $\mathbf{q}$ and $\mathbf{q}^*$ which are the projection of points $\mathbf{p} \in \mathcal{I}$ and $\mathbf{p}^* \in \mathcal{I}^*$ in space $\mathcal{Q}$ and $\mathcal{Q}^*$ invariant to intrinsic parameters. So, these intrinsics-free spaces are perfectly adapted to quantify the convergence of camera positions. These points are then used to compute the control law $\mathbf{v}$ applied to the controlled camera.

Different trajectories adapted to various object shapes (cylinder, sphere, several plans, ...) have been simulated. Noise has been added to the point coordinates. The final aim for the use of simulation is to improve the robustness of the control law, and to observe the robot behaviour in order to prepare the experiments, and thus to prevent the real arm from damage.

V. Experiments

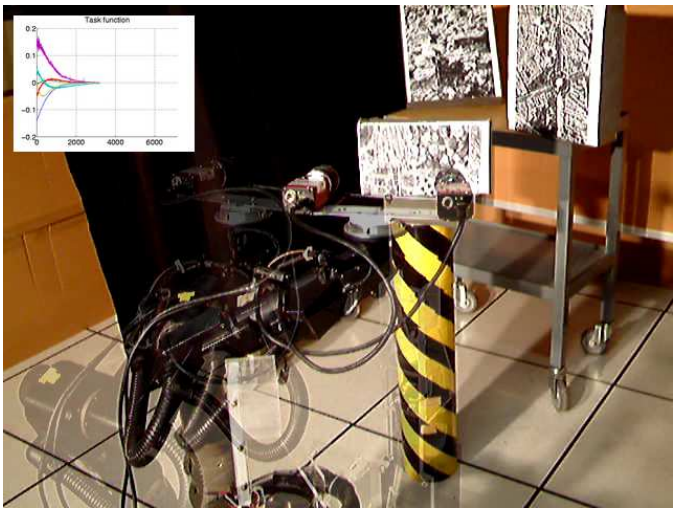


(a) Robot used with a 6 DOF arm *(b) Initial conditions*

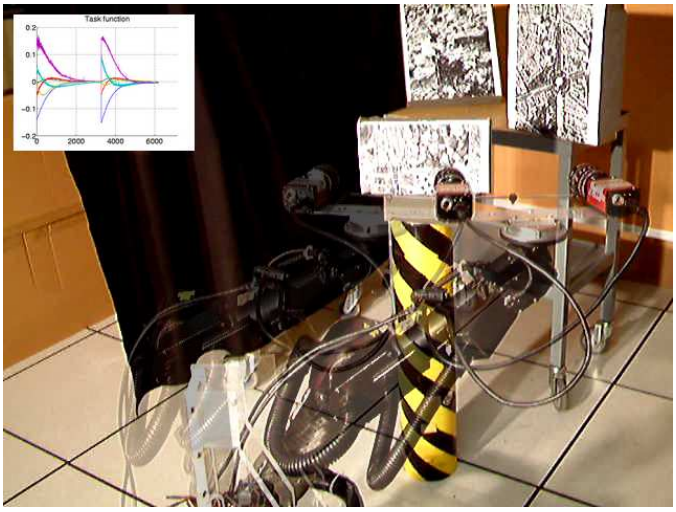
Fig. 5. Experimental conditions

In this part we have chosen to present the first experiments led in laboratories which allowed us to validate the results obtained by simulation.

Experiments have been conducted using a 6 DOF arm to test the visual servoing loop under optimum conditions. Figure (5) represents an experiment carried out in laboratory with a baseline of 20 cm between the cameras, and a right camera pan angle of 15 degrees. With this geometry the stereo rig describes a circle around the object. Both cameras have different intrinsic parameters. The 3D target, located at approximately 1.3 m from the cameras, is composed of three plans with different normal vectors (Figure (5)(b)).



(a) End of the 1^{st} servoing

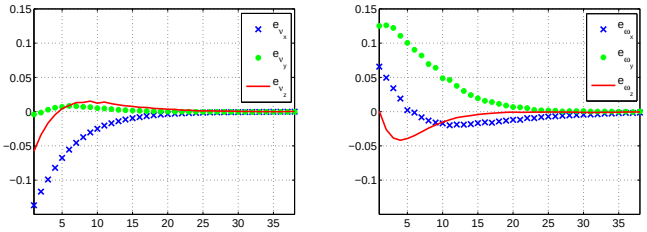


(b) End of the 2^{nd} servoing

Fig. 6. Visual servoing

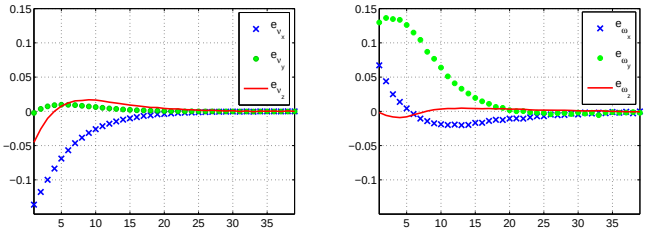
In order to compute the robot control law, features have to be extracted from images. We have focused our choice on a method based on a robust point extraction such as the SIFT algorithm [7]. These points are particularly robust to solve the difficult matching correspondence problem in underwater images of vent chimneys. Keypoints extracted from the reference and the current images are matched. A RANSAC algorithm [4] is applied to remove false matches. We then obtain points $\mathbf{p}^*$ (see section III) in the reference image, and points $\mathbf{p}$ matched in the current image of the controlled camera.

Points $\mathbf{p}$ are tracked in the current image until the left camera reaches the right camera position. The ESM tracking [1] algorithm has been chosen for these first experiments. But this method allows us to track only planar targets, therefore in the first experiments, the 3D target is made up of several plans. As soon as a new image is acquired, points are tracked and the command is computed and applied to the robot. To generate the command, the coordinates of points $\mathbf{p}^*$and $\mathbf{p}$ are projected in invariant spaces $\mathcal{Q}^*$ and $\mathcal{Q}$, to obtain respectively points



(a) Task function: $\mathbf{e}_\nu$ (b) Task function: $\mathbf{e}_\omega$

Fig. 7. Task function of the 1^{st} servoing



(a) Task function: $\mathbf{e}_\nu$ (b) Task function: $\mathbf{e}_\omega$

Fig. 8. Task function of the 2^{nd} servoing

$\mathbf{q}^*$ and $\mathbf{q}$, invariant to intrinsic parameters. At each iteration, points $\mathbf{q}$ are computed with the new coordinates of points $\mathbf{p}$.

When points $\mathbf{q}^*$ and $\mathbf{q}$ are superposed, task function $\mathbf{e}$ has converged exponentially towards zero (Figure (7) and Figure (8)). This is also illustrated in Figure (6), which represents the end of two successive displacements of the stereo rig, when the left camera position has reached the reference camera position. The starting position is represented in transparency, in order to show the superposition of the left camera on the reference position. Finally, we can see in Figure (7) and Figure (8) that the control law is stable in spite of noise and calibration error.

In pool trials are currently carried out in order to prepare the at sea trials. In Figure (9), we can see on the left the stereovision system made up of a frame in aluminium, two LED lights and two underwater cameras that are mounted on a baseline. One of these is a pan and tilt camera. On the right, the object is designed so that it is similar by its size, its shape and its texture to a section of a deep-sea vent. The results concerning these experiments will be presented in a future paper.

Fig. 9. In pool trials with the stereovision system and a 3D target

VI. Discussion

Our application relates to supervised exploration of ocean floors but does not concern autonomous exploration as it is defined by [20]. Our aim is to collect in real time with the most effectiveness the data concerning an object identified by an operator in order to improve the 3D reconstruction in a post-processing stage. In our case, time constraint is of great importance since deep-sea experiments are generally heavy to set up and to carry out. That is why we cannot afford to spend much time merging different sources of information during visual servoing and image acquisition. The solution we propose is thus to use only vision sensors to perform a predefined trajectory. But some uncertainties are induced by noise in the image that leads to a shift of the camera positions from their theoretical positions. It is thus necessary to take these uncertainties into account and to correct them in order to improve the quality of the final reconstruction result. To reduce these uncertainties, the noise can be first reduced by filtering during visual servoing. Then, in a post-processing stage, we use an algorithm of minimization, such as a sparse bundle adjustment algorithm [19], initialized with the theoretical camera positions.

Although the method of visual servoing we propose does not require the knowledge of camera intrinsic parameters, the latter have to be estimated in order to carry out a metric reconstruction. They can be estimated by a self-calibration method which does not imply the use of additional material. This is a great advantage for applications using robots whose payload capacity is limited. Some methods are well-suited to our application as [5] because the constrained nature of the motion to perform the self-calibration corresponds to the type of trajectories described by visual servoing.

VII. Conclusion

In this paper, we have presented an innovative visual servoing scheme in order to improve the 3D reconstruction of an unknown underwater object. The stereovision system we have specified allows us to generate predefined trajectories around the object under study, and thus to take pictures at regular intervals that are determined by the stereo rig geometry. Because the cameras are not necessarily the same and the intrinsic parameters may vary according to the environment, we have focused our attention on an intrinsics-free visual servoing method [9].

After the validation of our technique under laboratory conditions, tests under real underwater conditions are in progress. The at sea trials will bring additional constraints, yielding to a strong degradation of the images. These deteriorations are principally due to the presence of suspended particles and the problems of illumination in deep-sea environment associated to absorption that attenuates light. Moreover, the phenomenon of diffusion introduces a blur into the image [12]. Besides, the illumination of the objects by a lighting source rigidly linked to the cameras involves a variation of textures during the visual servoing because of the dropped shadow phenomenon. The noise in the image, the loss of the tracked points, the number of points and their distribution in the image can disturb the control law stability. But the first results obtained during in pool trials with the underwater stereovision system are pretty good and are promising for the future at sea trials. It thus will be very interesting to compare in a future paper the experiment results obtained under optimal laboratory conditions with those obtained in the complex world of the ocean depths.

Future work will focus on a novel technique for metric reconstruction adapted to underwater objects. Owing to a regular distribution of the shots taken around the object, the camera position corresponding to the images can be used as prior knowledge for the 3D reconstruction. Indeed, the number of unknown variables is reduced, increasing the accuracy and the speed of the optimization process, hence improving the reconstruction [11].

References

[1] S. Benhimane, E. Malis, *Real-time image-based tracking of planes using efficient second-order minimization*, IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'04, vol. 1, pp. 943–948, Sendai, Japan, October 2004.

[2] F. Chaumette, P. Rives, B. Espiau, *Positioning a robot with respect to an object tracking it and estimating its velocity by visual servoing*, Proceedings of the 1991 IEEE International Conference on Robotics and Automation, pp. 2248–2253, Sacramento, California, April 1991.

[3] B. Espiau, F. Chaumette and P. Rives, *A New Approach to Visual Servoing in Robotics*, IEEE Trans. on Robotics and Automation, pp. 313-326, 1992.

[4] M. A. Fischler, R. C. Bolles, *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*, Comm. of the ACM, Vol. 24, pp 381–395, 1981.

[5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, pp. 471-475, 2000.

[6] S. Hutchinson, G. D. Hager, and P. I. Corke, *A tutorial on visual servo control*," *IEEE Trans. on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, October 1996.

[7] D.G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, Vol. 60, 2, pp. 91–110, 2004.

[8] E. Malis, *A Unified Approach to Model-based and Model-free Visual Servoing*, European Conference on Computer Vision, Vol. 4, pp. 433–447, Copenhagen, Denmark, May 2002.

[9] E. Malis, *Visual servoing invariant to changes in camera intrinsic parameters*, IEEE Transaction on Robotics and Automation, 20(1):72–81, February 2004.

[10] E. Malis, *Vision-based control using different cameras for learning the reference image and for servoing*, IEEE/RSJ International Conference on Intelligent Robots Systems, Vol. 3, pp. 1428-1433, Hawaii, USA, November 2001.

[11] E. Malis and A. Bartoli, *Euclidean Reconstruction Independent on Camera Intrinsic Parameters* , IEEE/RSJ International Conference on Intelligent Robots Systems, Sendai, Japan, October 2004.

[12] N. Pessel, *Camera Self-Calibration in underwatrer environment*, Proceedings of The Fourteenth International Offshore and Polar Engineering Conference, ISOPE-2004, vol. 1 pp 738–745, Toulon, France May 23-28, 2004.

[13] J. Repko, M. Pollefeys, *3D Models from Extended Uncalibrated Video Sequences: Addressing Key-Frame Selection and Projective Drift*, 3DIM, pp. 150-157, 2005.

[14] P. Rives, F. Chaumette, B. Espiau, *Visual servoing based on a task function approach*, First International Symposium on Experimental Robotics, Montral, Canada, June 1989.

[15] C. Samson, M. Le Borgne, B. Espiau, *Robot Control: The Task Function Approach*, vol. 22 of Oxford Engineering Science Series. Clarendon Press, Oxford, UK, 1990.

[16] A.C. Sanderson and L.E. Weiss, *Image based visual servo control using relational graph error signal*, In Proc. Int. Conf. Cybern. and Soc. (Cambridge, MA), Oct. 1980, pp. 1074–1077.

[17] P.M. Sarradin et al., *EXtreme Ecosystem Studies in the Deep OCEan : Technological Developments*, Proceedings of The Fourteenth International Offshore and Polar Engineering Conference, ISOPE-2004, vol. 1 pp 738–745, Toulon, France May 23-28, 2004.

[18] P. Torr, A. W. Fitzgibbon, and A. Zisserman, *Maintaining multiple motion model hypotheses over many views to recover matching and structure*, IEEE Int. Conf. on Computer Vision, pp. 485–491, Bombay, India, January 1998.

[19] B. Triggs, P. McLauchlan, R. Hartley, A. Fiztgibbon, *Bundle Adjustment: A Modern Synthesis*, In B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice, LNCS Vol.1883, pp.298-372, Springer-Verlag, 2000.

[20] Peter Whaite, Frank P. Ferrie, *Autonomous Exploration: Driven by Uncertainty*,IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 19(3), pp. 193-205, 1997.