Proc. of CVPR 94, pp. 347-354, Seattle, June 1994

Optimal Estimation of 3D Structures Using Visual Servoing

F. Chaumette*, S. Boukir*, P. Bouthemy*, D. Juvin**

* IRISA / INRIA Rennes, Campus de Beaulieu, 35042 Rennes, France
** CEA-LETI / DEIN-SLA Saclay, 91191 Gif-sur-Yvette, France
e-mail: chaumett@irisa.fr

Abstract

This paper deals with the recovery of 3D information using a single mobile camera in the context of active vision. We propose a general revisited formulation of the structure-from-motion issue, and we determine adequate camera configurations and motions which lead to a robust and accurate estimation of the 3D structure parameters. We apply the visual servoing approach to perform these camera motions. Real-time experiments dealing with the 3D structure estimation of points and cylinders are reported, and demonstrate that this active vision strategy can very significantly improve the estimation accuracy.

1 Introduction

Recovering 3D structure from images is one of the main issue in computer vision. Among others, an appealing way of solving this problem is to use 2D motion computed in image sequences acquired by a monocular camera. Basically, two main approaches have been investigated to solve the problem of structure from known motion: long range motion-based and short range motion-based ones. In the former approach, images are considered at distant time instants. This approach is based on the extraction of a set of relatively sparse, distinguishable two-dimensional features in the successive images [9], [16], [17] . Inter-frame correspondence is first established between these features. Then, the 3D structure is determined. In the latter approach, images are considered at video rate [12], [15]. In this case, the emphasis is placed on the estimation of the optic flow field in the image. A usually dense flow field is computed and used in conjunction with a measure of camera motion to determine the 3D structure of the scene.

However, to correctly compute correspondence (or optic flow) is a difficult problem requiring the development of sophisticated algorithms. Furthermore, the optic flow field is generally corrupted by noise and partially incorrect especially near occlusion or motion boundaries.

To alleviate these problems, hybrid approaches have been proposed like in [8]. Such methods are based on a formulation in terms of continuous velocities (i.e., use of the instantaneous kinematic screw) in the 3D reconstruction process, while relying on the tracking of 2D sparse image features. These algorithms avoid the intermediate stage of optic flow computation and involve simple matching process. But, they still suffer from several shortcomings: their sensitivity to noise, their numerical unstability and their unsatisfactory accuracy. An attractive way to cope with these problems is to follow an active vision approach, which can be defined as an intelligent data acquisition process [1-3]. Our concern is to deal with the problem of recovering 3D spatial structure using a single mobile camera by means of an active vision scheme, and to show that 3D reconstruction can thus be solved in a much more efficient way compared to an usual dynamic vision approach. This issue has already motivated some investigations [2] [14], but only for the case of 3D points. Furthermore, effective comparison between dynamic and active vision has not yet been performed through real experiments. In this paper, the problem of the 3D reconstruction of geometrical primitives is handled at two levels:

- modeling aspect: we propose a general revisited formulation of the structure-from-motion issue. The same framework can handle various kinds of 3D geometrical primitives, i.e., points, lines, cylinders, spheres, ...
- optimization aspect: we derive sufficient conditions to minimize the effects of the different measurement errors which may occur in this process. More precisely, we determine the adequate camera locations and motions which provide a robust and accurate estimation of the 3D structure. We apply the visual

servoing approach to perform these motions using a control law in closed-loop with respect to visual data.

Finally, we demonstrate with real-time experiments that our active vision scheme significantly improves the accuracy of the structure estimation.

2 Structure from motion using dynamic vision

Let us consider an usual perspective projection model (see Figure 1). Without loss of generality, the focal length is assumed to be equal to 1. The relation between the 3D point $\underline{x} = (x\,y\,z)$ and its projection $\underline{X} = (X\,Y\,1)$ on the image plane is given by:

$$\underline{X} = \frac{1}{z}\,\underline{x} \tag{1}$$

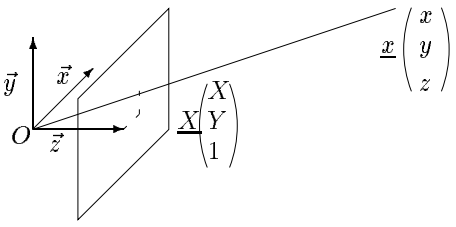


Figure 1: *Camera model.*

The velocity screw of the camera frame $(O, \vec{x}, \vec{y}, \vec{z})$ with respect to the scene is denoted by $T = (V(O), \Omega)$ where $V(O) = (V_x\,V_y\,V_z)^T$ and $\Omega = (\Omega_x\,\Omega_y\,\Omega_z)^T$ are respectively the translational and rotational velocities. If point $\underline{x}$ is static, we get:

$$\dot{\underline{x}} = -V(O) - \Omega \times \underline{x} \tag{2}$$

Differentiating (1) and using (2) lead to the well known relations [10]:

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = 1/z \begin{pmatrix} -V_x + XV_z \\ -V_y + YV_z \end{pmatrix} + \begin{pmatrix} XY\Omega_x - (1+X^2)\Omega_y + Y\Omega_z \\ (1+Y^2)\Omega_x - XY\Omega_y - X\Omega_z \end{pmatrix} \tag{3}$$

From these two equations, we can easily derive the expression of the unknown depth z; we obtain [11]:

$$1/z = \frac{\alpha_x(XV_z - V_x) + \alpha_y(YV_z - V_y)}{(XV_z - V_x)^2 + (YV_z - V_y)^2} \tag{4}$$

with $\alpha_x = XY\Omega_x - (1+X^2)\Omega_y + Y\Omega_z - \dot{X}$ and $\alpha_y = (1+Y^2)\Omega_x - XY\Omega_y - X\Omega_z - \dot{Y}$. Let us note that no information on the depth z can be retrieved if the camera motion is such that $V_x = XV_z$ and $V_y = YV_z$.

We will now adopt in the remainder of this section the same approach in order to estimate the parameters describing more complex primitives such as straight line, circle, sphere, cylinder.

Let us consider the geometrical primitive $\mathcal{P}_s$ of the scene; its configuration is specified by an equation of the type:

$$h(\underline{x}, \underline{p}) = 0\ , \quad \forall \underline{x} \in \mathcal{P}_s \tag{5}$$

where h expresses the type of the considered primitive and the value of parameters $\underline{p}$ defines its corresponding configuration. The representation $\underline{p}$, of dimension n, is chosen complete and minimal in order that any position of the primitive can be represented by only one value of $\underline{p}$.

Using (1), equation (5) becomes:

$$h'(\underline{X}, 1/z, \underline{p}) = 0 \tag{6}$$

Under the trivial condition $\frac{\partial h'}{\partial z} \neq 0$ which is satisfied in all the non-degenerated cases (a degenerated case occurs for example when a line is projected onto the image plane as a point, or a circle as a segment), the implicit function theorem ensures the existence of a unique function μ such that:

$$1/z = \mu(\underline{X}, \underline{p}_0) \tag{7}$$

where the representation $\underline{p}_0$, function of $\underline{p}$, is chosen complete and minimal (its dimension is denoted n_0).

Let us denote $\mathcal{P}_i$ the projection in the image plane of $\mathcal{P}_s$. Substituting (7) for $1/z$ in (6), the configuration of $\mathcal{P}_i$ can be written as follows:

$$g(\underline{X}, \underline{P}) = 0\ , \quad \forall \underline{X} \in \mathcal{P}_i \tag{8}$$

where g defines the type of the image primitive and the value of $\underline{P}$, function of $\underline{p}$, its configuration. Once again, the representation $\underline{P}$, of dimension m, is chosen complete and minimal in order that any position in the image of $\mathcal{P}_i$ can be represented by only one value of $\underline{P}$.

- **Remark:** For a planar primitive, the function μ represents the plane in which the primitive lies. For a tri-dimensional primitive (sphere, cylinder, torus,...), the function $g(\underline{X}, \underline{P})$ is the limb equation (we only consider the contour of $\mathcal{P}_i$). Matching between 3D points and 2D contour points provides us with the function $\mu(\underline{X}, \underline{p}_0)$ which is thus called the limb surface.

The time variation of parameters $\underline{P}$, which links the motion of the primitive in the image to the camera motion T, can be explicitly derived [7], and we get:

$$\dot{\underline{P}} = L^T_{\underline{P}}(\underline{P}, \underline{p}_0)\,T \tag{9}$$

where $L_{\underline{P}}^T$, called the interaction matrix related to $\underline{P}$, fully characterizes the interaction between the camera and the considered primitive.

We are now able to present a general method to reconstruct a 3D geometrical primitive using dynamic vision (i.e. to compute the value of $\underline{p}$ using the measure, along an image sequence, of the camera velocity T and of the image parameters $\underline{P}$ and $\dot{\underline{P}}$).

Let us denote $\mathcal{H}(\underline{P}, \dot{\underline{P}}, \underline{p_0}, T)$ the following function derived from (9):

$$\mathcal{H}(\underline{P}, \dot{\underline{P}}, \underline{p_0}, T) = \dot{\underline{P}} - L_{\underline{P}}^T(\underline{P}, \underline{p_0})\, T = 0 \qquad (10)$$

Under the condition $\dfrac{\partial \mathcal{H}}{\partial \underline{p_0}}$ (of dimension $m \times n_0$) is full rank n_0, the implicit function theorem allows us to express $\underline{p_0}$ with respect to the other parameters involved in (10). Thus, we obtain:

$$\underline{p_0} = \underline{p_0}\,(T, \underline{P}, \dot{\underline{P}}) \qquad (11)$$

More precisely, for all the primitives that we have studied (lines, circles, spheres and cylinders), the parameters $\underline{p_0}$ are simply determined from the resolution of a linear system derived from (10).

Furthermore, let us note that it is possible to find the camera motions which do not provide any information on the spatial configuration of the primitive: they are such that $\dfrac{\partial \mathcal{H}}{\partial \underline{p_0}}$ is not of full rank.

Finally, knowing $g(\underline{X}, \underline{P})$ and $\mu(\underline{X}, \underline{p_0})$, we can solve for the parameters $\underline{p}$ which completely define the configuration of the considered primitive:

$$\underline{p} = \underline{p}\,(\underline{P}, \underline{p_0}) \qquad (12)$$

Let us note that our method is based on a continuous approach since it uses the measure of the camera velocity. It is basically different from the discrete ones [16], [17] which consider a camera displacement (described by a translation and a rotation matrix) instead of camera velocity.

From a geometrical point of view (see Figure 2), our method consists in determining the intersection between a generalized cone (defined by its vertex O and the function $g(\underline{X}, \underline{P})$) and the limb surface (derived as explained above). On the other hand, the discrete approach, equivalent to a stereovision paradigm, is based on the intersection between two generalized cones (corresponding to each camera position). Intersecting two volumes, instead of a volume and a surface, seems more complicated to derive closed-formed expressions (and therefore robust estimations) in the case of complex primitives. For a circle for example, the discrete method proposed in [13] is based on the resolution of a complex non linear system, whereas our method is based on the simple resolution of two linear systems.

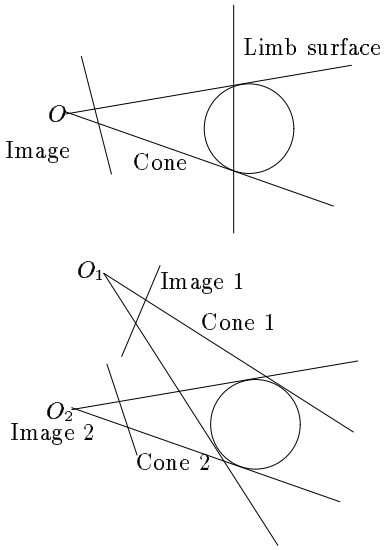


Figure 2: *Difference between continuous (on the top) and discrete (on the bottom) approaches.*

The estimation of the depth z of a point, which has been described at the beginning of this section, can of course be obtained using this formalism (it is the simplest case). The proposed formalism has been successfully used for the 3D structure estimation of lines, circles, spheres and cylinders [6]. For all of these primitives, the closed-formed expressions of the 3D parameters to be estimated are simply determined from the resolution of two linear systems (the first one to determine the parameters of the limb surface using the interaction matrix, the second one to determine the parameters $\underline{p}$). Furthermore, this approach can probably be used for more complex primitives, such as torus for example.

3 Structure from motion using active vision

It has been observed that the 3D reconstruction from monocular image sequence is very sensitive to the nature of the successive camera motions [8]. The experimental results reported in the next section confirm that important errors on the structure estimation appear when no particular strategy concerning the camera motion is defined. Besides, we have seen that some given motions are not able to provide any

3D-information. Therefore, one of the goal of an active vision scheme is to find an optimal camera motion which could lead to a robust and non biased estimation of the 3D spatial structure. In this section, we state the problem in terms of the minimization of the errors occurring in the process. Two kinds of errors are of particular concern: the first one is due to the discretization step that affects the continuous method that we have proposed in the previous section, the second one is due to the unavoidable measurement errors on the image data and on the camera motion.

3.1 Suppression of the discretization error

The main error encountered in the recovery of structure from known motion using dynamic vision comes from the discretization error. Indeed, our method is based on the measure of $\underline{\dot{P}}$, i.e., the time variation of the parameters representing the considered image primitive. The exact value of $\underline{\dot{P}}$ is generally unreachable and the image measurements only supplies $\Delta\underline{P}$, the "displacement" of $\underline{P}$ during the time interval Δt between two successive images. If we use $\Delta\underline{P}/\Delta t$ instead of $\underline{\dot{P}}$ in the derivation described above, this induces discretization errors which can be important as it will be seen in the experimental results. On the other hand, if we can ensure that $\underline{\dot{P}} = \Delta\underline{P}/\Delta t, \forall t$, the discretization step will have no effect. Such a condition is satisfied if and only if:

$$\underline{\ddot{P}} = \cdots = \underline{P}^{[n]} = 0, \forall t \tag{13}$$

From (9), we have $\underline{\dot{P}} = f(\underline{P}, \underline{p_0}, T)$. Thus:

$$\underline{\ddot{P}} = \frac{\partial f}{\partial \underline{P}}\underline{\dot{P}} + \frac{\partial f}{\partial \underline{p_0}}\underline{\dot{p_0}} + \frac{\partial f}{\partial T}\dot{T} \tag{14}$$

If we consider that the camera velocity is constant between two image acquisitions, then $\dot{T} = 0$, and a sufficient and general condition to verify (13), is to constrain the camera motions to be such that:

$$\underline{\dot{P}} = \underline{\dot{p_0}} = 0, \forall t \tag{15}$$

In other words, a solution to suppress the discretization error is that the equation of the limb surface remains the same, and that the primitive constantly appears at the same position in the image while the camera is moving.

We can show that, except for points and lines, the first condition $\underline{\dot{P}} = 0$ implies the second one $\underline{\dot{p_0}} = 0$, which reduces the problem to a **fixation** situation. We will see in Section 3.3 that the visual servoing approach is perfectly appropriate to generate such camera motions.

Let us note that the condition that we have exhibited is only sufficient, and not necessary. Indeed, camera motion exists such that $\underline{\ddot{P}} = 0$ with $\underline{\dot{P}} \neq 0$ or $\underline{\dot{p_0}} \neq 0$. For a point for example, we can easily show by differentiating (4) that $\underline{\ddot{P}} = 0$ when the camera motion is a pure translation parallel to the image plane, i.e., $V_z = \Omega_x = \Omega_y = \Omega_z = 0$. More generally, these motions are such that (see relation (14)):

$$\frac{\partial f}{\partial \underline{P}}\underline{\dot{P}} + \frac{\partial f}{\partial \underline{p_0}}\underline{\dot{p_0}} = 0 \tag{16}$$

Determining all the solutions of such a non linear system is a very difficult task. Moreover, they deeply depend on the considered primitive since they require the knowledge of $\frac{\partial f}{\partial \underline{P}}$ and $\frac{\partial f}{\partial \underline{p_0}}$. On the other hand, the condition (15) is valid for any kind of primitives. It has the supplementary advantage that the primitive will remain in the field of view of the camera during the estimation process.

3.2 Minimizing the effects of the measurement errors

An other important point in an active vision context is to select configurations of the camera with respect to the primitive of interest, likely to provide an estimation as robust as possible. More precisely, we show in this section that the effect of the measurements errors on the estimation of the 3D spatial structure of the primitive depends on the position of the projection of the primitive in the image. Therefore, we propose to constrain the camera motion in order to **focus** on the primitive in order that the primitive to be reconstructed is located at the position in the image that minimizes the effect of the measurement errors.

Let us denote a parameter of the 3D primitive by p. Recall that p depends on information extracted from the image $(\underline{P}, \underline{\dot{P}})$ and on the measured camera velocity T. If we suppose that the measurement errors on $\underline{P}, \underline{\dot{P}}$ and T are not correlated, the uncertainty σ_p on the estimation of p can be written in the form:

$$(\sigma_p)^2 = \sum_{i=1}^{m}\left(\frac{\partial p}{\partial P_i}\right)^2(\sigma_{P_i})^2 + \sum_{j=1}^{m}\left(\frac{\partial p}{\partial \dot{P}_j}\right)^2(\sigma_{\dot{P}_j})^2 + \sum_{k=1}^{6}\left(\frac{\partial p}{\partial T_k}\right)^2(\sigma_{T_k})^2 \tag{17}$$

Minimizing σ_p is equivalent to minimizing each term $p_{a_i} = (\frac{\partial p}{\partial a_i})^2$ where a_i denotes any of the variables $\underline{P}, \dot{\underline{P}}$ and T. We thus have to find the value of $\underline{P}$ such that $\left(\frac{\partial p_{a_i}}{\partial P_j}\right) = 0, \forall a_i$ and $\forall j = 1$ to m. To find all the solutions of this non linear system in an analytical way seems unreachable. However, we have derived the following particular solutions of interest:

- for a point, the effect of the measurement errors are minimized (i.e., $\frac{\partial p_{a_i}}{\partial X} = \frac{\partial p_{a_i}}{\partial Y} = 0, \forall a_i$) if the point constantly appears at the center of the image ($X = \dot{X} = Y = \dot{Y} = 0, \forall t$) during the estimation time interval, and if we also have $V_z = \Omega_z = 0$. The camera must therefore be displaced on a sphere the center of which is the point to be reconstructed. It is interesting to notice that we get conclusions similar as those obtained in [4] and [14] where the interest of locating the fixation point of an active binocular head in the center of the image is demonstrated.
- for a sphere, the effect of the measurement errors are minimized if the image of the sphere constantly remains a circle centered on the image center and if $\Omega_z = 0$. The optimal trajectory of the camera is thus the same as in the previous case.
- for a straight line, the effect of the measurement errors are minimized if the line always appears centered in the image ($\rho = \dot{\rho} = 0$), vertical ($\theta = \dot{\theta} = 0$), and if $V_y = V_z = \Omega_x = 0$ (or horizontal with $V_x = V_z = \Omega_y = 0$).
- for a cylinder, similarly, the effect of the measurement errors are minimized if the projections of its two limbs lie astride the image center in a symmetric manner ($\rho_1 = -\rho_2$), vertically ($\theta_1 = \theta_2 = 0$), and if $V_y = 0$ (or horizontally with $V_x = 0$). Therefore, the camera must be displaced on a circle around the cylinder (see Figure 3).

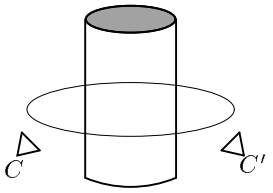


Figure 3: *Optimal camera motion in the cylinder case.*

Unfortunately due to the complexity of the stated problem, we have not been able to prove that these solutions are unique. On the other hand, we have checked that numerous configurations really do not minimize all the p_{a_i} terms and thus are not likely to provide a robust estimation (for example, let us just consider $X = 1, Y = 0$ in the case of the point).

We now describe how it is possible to automatically compute the camera motion satisfying the constraints described above.

3.3 Image-based closed-loop control

Active vision aims at improving the knowledge of the environment by means of adequate camera motions. A control law in closed-loop with respect to visual data is perfectly suitable to generate such motions. This **visual servoing** approach is based on the regulation to zero of a task function $\underline{e}$ which can be written as follows [7]:

$$\underline{e} = W^+ W C \, (\underline{P} - \underline{P}^*) + (\mathbb{I}_6 - W^+ W)\underline{e}_2 \qquad (18)$$

where:

- $\underline{P}$ is the value of the 2D parameters describing the projection in the image of the primitive on which the camera is fixing or focusing. $\underline{P}$ is measured at each iteration of the control law.
- $\underline{P}^*$ is the target value of $\underline{P}$ to be obtained. To suppress the discretization error, we have to satisfy $\dot{\underline{P}} = 0$; in this fixation task, $\underline{P}^*$ is therefore set to the initial measured value of $\underline{P}$. In the focusing task, $\underline{P}$ has to reach a given value to obtain a robust estimation ($X = Y = 0$ for a point for example); in that case, $\underline{P}^*$ must then be equal to this particular value.
- C is a matrix which represents the inverse jacobian of the vision-based task. Ideally, this matrix is chosen as the pseudo-inverse of the interaction matrix related to $\underline{P}$: $C = L_{\underline{P}}^{T^+}(\underline{P}, \underline{p}_0)$. But, since the real value of $\underline{p}_0$ is unknown, we choose $C = L_{\underline{P}}^{T^+}(\underline{P}, \hat{\underline{p}}_0)$ where $\hat{\underline{p}}_0$ is the current estimation of the parameters of the limb surface obtained by the method described in Section 2.
- $\underline{e}_2$ is a secondary task which allows us to move the camera along a desired trajectory (on a sphere or a circle for example). $\underline{e}_2$ also permits to satisfy the additional constraint $\dot{\underline{p}}_0 = 0$ for the point and line cases.
- W^+W and $\mathbb{I}_6 - W^+W$ are two projection operators which guarantee that the camera motion due to the secondary task is compatible with the regulation of $\underline{P}$ to $\underline{P}^*$ ($\mathbb{I}_6$ is the 6×6 identity matrix and W is a full rank matrix such that Ker W = Ker $L_{\underline{P}}^T$. More details are given in [7]).

Once the task function $\underline{e}$ is defined, a simple control law, which computes the camera velocity T and ensures an exponential decrease of $\underline{e}$, is given by [7]:

$$T = -\lambda \, \underline{e} - (\mathbb{I}_6 - W^+ W)\frac{\partial \underline{e}_2}{\partial t} \qquad (19)$$

where λ (> 0) is the factor that controls the speed of the decay, and where the term $(\mathbb{I}_6 - W^+W)\frac{\partial e_2}{\partial t}$ is tied to the generation of a non zero camera motion when the vision-based task is realized (i.e. when $\underline{P} = \underline{P^*}$).

3.4 Experimental results

We present in this section the experimental results obtained for the 3D structure estimation of a point and a cylinder. More detailed experiments are described in [6]. For each of these primitives, we compare the results supplied by a dynamic vision approach (i.e., general camera motion) with those given by the active vision approach in order to demonstrate the improvement brought by the latter.

Our experimental system is composed of a camera mounted on the end effector of a six d.o.f. robot arm. The image processing part is realized on a commercial board. Each iteration of our algorithm is realized in 100 milliseconds.

3.5 Case of the 3D point reconstruction

The first image acquired by the camera is depicted in Figure 4.a. The point that we consider is the center of gravity of the white ball which is in the field of view of the camera. The image processing step simply consists in extracting and tracking along the image sequence the center of gravity of the ellipse corresponding to the projection of the ball in the image. From the initial position, the camera moves with constant velocity (in the reported experiment, $V_x = V_y = V_z = 50\ mm/s, \Omega_x = \Omega_y = \Omega_z = 3\ dg/s$) and the results obtained are shown in Figure 5.a. The plots represent the 3D coordinates (x, y, z) of the considered point, estimated at each iteration of our algorithm. The position (x, y, z) of the point is expressed in a world reference frame (corresponding to the first camera position); the plots should then correspond to constant values over time. Important errors can be observed in that experiment, where no particular strategy, as far as camera motion is concerned, has been used (errors can reach 10 cm). Let us note that similar results were obtained with other general camera motion.

As proposed in Section 3.1 to improve these results, we first constrain the camera movement in order to suppress the discretization error. Consequently, the point projection remains static in the image, and the distance between the camera and the point is maintained constant. To perform that task, we use the control law described in Section 3.3. The results are shown in Figure 5.b. During the first iterations, errors are still large. This is due to the fact that several iterations are required to perfectly achieve the fixation task by the control law. After the short transient period, the results become stable and errors on the estimated depth z are less than 1 cm. By comparing these results with the previous ones, we can observe the important improvement brought by the first part of the active vision strategy.

As explained in Section 3.2, results can be improved further by positioning the camera in such a way that the point constantly appears in the center of the image (see Figure 4.b). Indeed, the effects of the measurement errors are minimized for that position. The estimated 3D point coordinates obtained after the realization of the focusing task are shown in Figure 5.c. These values are particularly stable and accurate. Errors on the depth z are only about 2 mm (that is 2.5%).

3.6 Case of the cylinder reconstruction

Let us now apply this approach to the 3D reconstruction of a cylinder, the equation of which is given by:

$$\begin{aligned} h(\underline{x}, \underline{p}) &= (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \\ &- (ax + by + cz)^2 - r^2 = 0 \\ \text{with } a^2 &+ b^2 + c^2 = 1 \text{ and } ax_0 + by_0 + cz_0 = 0 \end{aligned} \quad (20)$$

where r is the radius of the cylinder, (a, b, c) represents the direction of its axis and (x_0, y_0, z_0) is the point of the cylinder axis the nearest to the camera. These are the parameters to be estimated.

The initial image acquired by the camera is shown in Figure 6.a (note the superimposed two white lines corresponding to the two selected limbs of the cylinder). The image processing step now consists in tracking theses two limbs along the image sequence and in determining the (ρ, θ) parameters describing their position in the image. The extraction, maintenance and tracking of the contour segment (in fact a list of edge points) are achieved in 80 ms. The method we have used is described in [5]. It is based on a local and robust matching of the moving edge-points constituting the selected line.

Results obtained using a not constrained camera motion are plotted in Figure 7.a, and results obtained once the focusing task has been achieved (see Figure 6.b) are plotted in Figure 7.b. Once again, we can point out the very significant improvement resulting from our estimation scheme using active vision. Let us note that, after the first iterations, the error

between the real value of the cylinder radius (40 mm) and the estimated one is lower than 1 mm and generally around 0.2 mm, which demonstrates the robustness and the validity of the proposed method.

4 Conclusion

We have described an original formulation of the problem of reconstructing 3D parametric geometrical primitives using a mobile monocular camera. The introduction of the so-called interaction matrix related to the primitive under concern allows us to define a general and attractive framework which can be applied to usual primitives such as points and straight lines, but also to more complex primitives such as cylinders, spheres, ..., without additional complexity in the derivation of the solution.

Since the nature of the camera motion affects the accuracy of the results, we have focused on this critical aspect of dynamic vision. We have mathematically and experimentally shown that very noticeable improvements can be obtained in the 3D recovery of a large class of geometrical primitives, if the camera is properly positioned, and if optimal camera motions are generated. Our approach consists in particular in fixing and focusing on the 3D primitive to be determined. This confirms the point of view of previous works on the promising strength of active vision paradigms [2], [3], [4], [14]. We have stressed that this active vision approach can be adequately performed using a control law in closed loop with respect to visual data. A real-time version of this visual servoing scheme has been implemented on an experimental system, and it turns out to be powerful and efficient. Future work will be devoted to the development of global perceptual strategies able to appropriately combine a succession of such optimal individual primitive reconstruction to recover the complete spatial structure of complex scenes.

Acknowledgements: This work was done in collaboration with CEA-LETI DEIN-SLA under contract 1 91 C 244 00 31315 01 1. It was also partly supported by Région Bretagne (Brittany County Council) under contribution to student grant.

References

[1] Abbot L., Ahuja N. : Active surface reconstruction by integrating focus, vergence, stereo, and camera calibration. 3^{rd} *ICCV*, Osaka, pp. 489-492, December 1990.

[2] Aloimonos J., Weiss I., Bandopadhay A. : Active vision. 1^{st} *ICCV*, London, pp. 35-54, June 1987.

[3] Bajcsy R. : Active perception. *Proc. IEEE*, Vol. 76, No. 8, pp. 996-1005, August 1988.

[4] Bandopadhay A., Chandra B., Ballard D. : Egomotion using active vision. *CVPR*, Miami, pp. 498-503, June 1986.

[5] Boukir S., Bouthemy P., Chaumette F., Juvin D. : Real-time contour matching over time in an active vision context, 8^{th} *SCIA*, Vol. 1, pp. 113-120, Tromso, Norway, May 1993.

[6] Boukir S. : Reconstruction 3d d'un environnement statique par vision active. *Thèse de l'Université de Rennes 1*, October 1993.

[7] Espiau B., Chaumette F., Rives P. : A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, Vol 8, No. 3, pp. 313-326, June 1992.

[8] Espiau B., Rives P. : Closed-loop recursive estimation of 3D features for a mobile vision system. *IEEE Int. Conf. on Robotics and Automation*, Raleigh, Vol. 3, pp. 1436-1443, April 1987.

[9] Faugeras, O. : *Three-dimensional computer vision: a geometric viewpoint.* MIT Press, 1993.

[10] Horn B. : *Robot Vision.* MIT Press, 1987.

[11] Matthies L., Kanade T., Szeliski R. : Kalman filter-based algorithms for estimating depth from image sequences. *IJCV*, Vol. 3, pp. 209-236, 1989.

[12] Negahdaripour S., Horn B. : Direct passive navigation. *IEEE Trans. on PAMI*, Vol. 9, No. 1, pp. 168-176, January 1987.

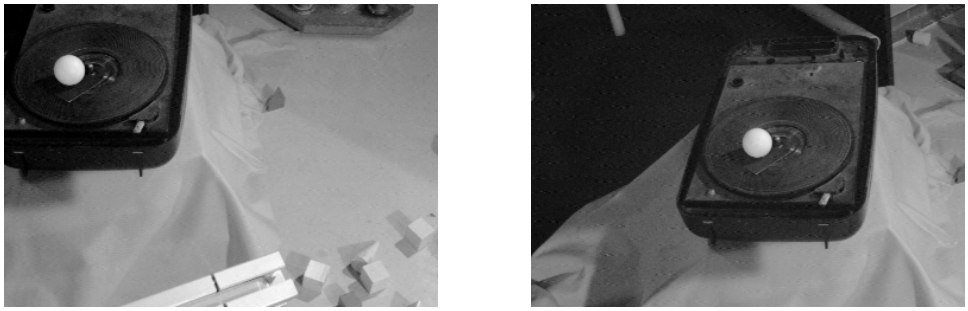
[13] Safaee-Rad R., Tchoukanov I., Benhabib B., Smith K. : 3D-pose estimation from a quadratic-curved feature in two perspective views. 11^{th} *ICPR*, The Hague, pp. 341-344, August 1992.

[14] Sandini G., Tistarelli M. : Active tracking strategy for monocular depth inference over multiple frames. *IEEE Trans. on PAMI*, Vol. 12, No. 1, pp. 13-27, January 1990.

[15] Vernon D., Tistarelli M. : Using camera motion to estimate range for robotic part manipulation. *IEEE Trans. on Robotics and Automation*, Vol. 6, No. 5, pp. 509-521, October 1990.

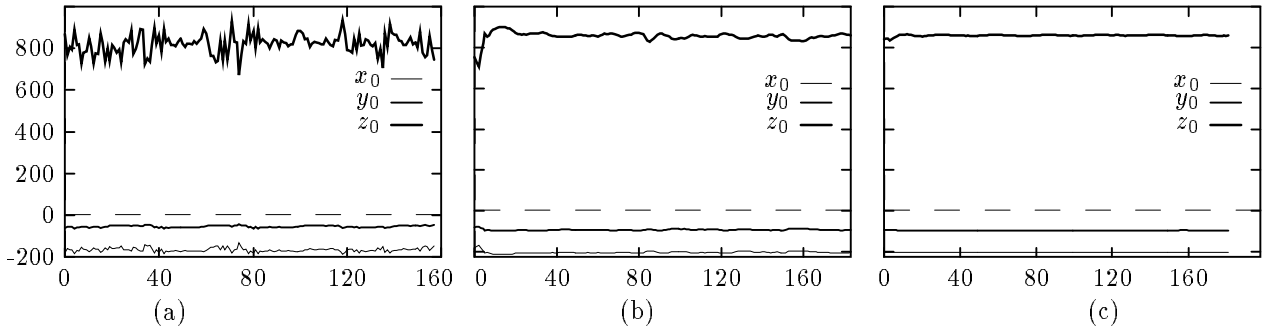
[16] Viala M., Faye C., Guerin J.P., Juvin D. : Cylindrical object reconstruction from a sequence of images. *Conf. SPIE Intelligent Robots and Visual Communications*, Boston, November 1992.

[17] Weng J., Huang T., Ahuja N. : Motion and structure from line correspondences: closed-form solution, uniqueness, and optimization. *IEEE Trans. on PAMI*, Vol. 14, No. 3, pp. 318-336, March 1992.



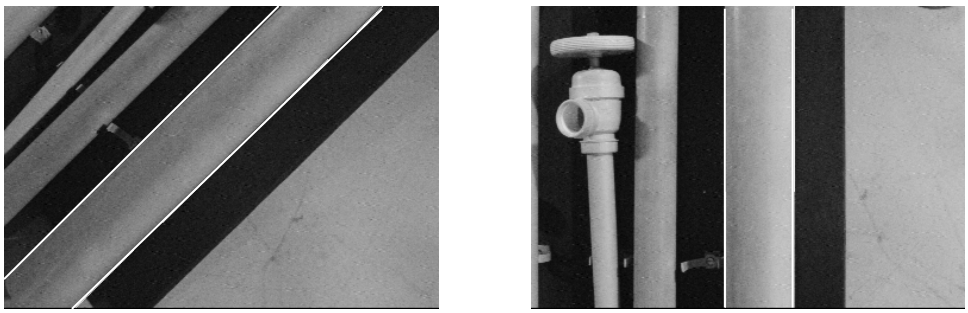
(a) (b)

Figure 4: Images acquired at the initial camera position (a) and after the realization of the focusing task (b).



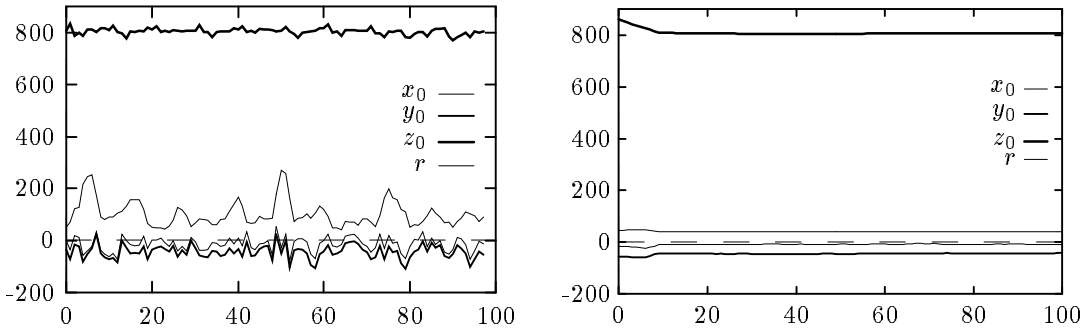
(a) (b) (c)

Figure 5: Successive values of the 3D point coordinates x_0, y_0, z_0 (expressed in mm) estimated using a dynamic vision approach (a), obtained with camera motion allowing to suppress the discretization error (b) and using the active vision scheme (c)



(a) (b)

Figure 6: Images acquired at the initial camera position (a) and after the realization of the focusing task (b).



(a) (b)

Figure 7: Successive values of the 3D cylinder parameters x_0, y_0, z_0 and r estimated using dynamic vision (a) and using the active vision scheme (b).