IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'96, pp. 169–176, San Francisco, California, june 1996

Controlled camera motions for scene reconstruction and exploration

Éric Marchand, François Chaumette

IRISA - INRIA Rennes - Université de Rennes I
Campus universitaire de Beaulieu - F-35042 Rennes Cedex, France
Email {marchand, chaumett}@irisa.fr

Abstract

This paper deals with the 3D structure estimation and exploration of a scene using active vision. Our method is based on the structure from controlled motion approach which consists in constraining the camera motion in order to obtain a precise and robust estimation of the 3D structure of a geometrical primitive. Since this approach involves to gaze on the considered primitive, we present a method for connecting up many estimations in order to recover the complete spatial structure of scenes composed of cylinders and segments. We have developed perceptual strategies able to perform a succession of robust estimations without any assumption on the number and on the localization of the different objects. Furthermore, the proposed strategy ensures the completeness of the reconstruction. An exploration process centered on current visual features and on the structure of the previously studied primitives is presented. This leads to a gaze planning strategy that mainly uses a representation of known and unknown areas as a basis for selecting viewpoints. Finally, experiments carried out on a robotic cell have proved the validity of our approach.

1 Introduction

Many applications in robotics involve a good knowledge of the robot environment. For such applications, the aim of this paper is to obtain a complete and precise description of a scene using the visual data provided by a camera mounted on the end effector of a robot arm. A recent expansion of computer vision and image analysis is related to the estimation of 3D structure from image sequences [1][7][17]. The approach we have chosen to get an accurate three-dimensional geometric description of a scene is based on the active vision paradigm and consists in controlling the camera motion. The idea of using active schemes to address vision issues has been recently introduced [2][3]. Active vision is defined in [3] as an intelligent data acquisition process. Since the major shortcomings which limit the performance of vision systems are their sensitivity to noise and their low accuracy, the aim of active vision is generally to elaborate control strategies for adaptively setting camera parameters (position, velocity,...) in order to improve the knowledge of the environment [2]. Here, the purpose of active vision is handled at two levels: a **local aspect** where active vision is used to constrain the camera motion in order to improve the quality of the reconstruction results, and a **global aspect** which is used to explore the unknown areas.

The measure of the camera motion, which is necessary for the 3D structure estimation, characterizes a domain of research called dynamic vision. Approaches for 3D structure recovery may be divided into two main classes: the discrete approach, where images are acquired at distant time instants [7][17] and the continuous approach, where images are considered at video rate [1]. The method used here is a continuous approach which stems on the camera velocity and on the motion of the considered primitive in the image. More precisely, we use a "*structure from controlled motion*" method which consists in constraining the camera motion in order to obtain a precise and robust estimation of 3D geometrical primitives such as points, straight lines and cylinders [6]. Such constraints are automatically ensured using *visual servoing* [10]. Simplifying and improving shape estimation by viewpoint control is also reported in [11].

As far as the **global aspect** of our reconstruction scheme is concerned, active vision is used to determine the location of the next camera position in order to obtain a complete model of the scene. Previous works have been done in order to answer the "*where to look next*" question. Differences can be done if an *a priori* knowledge about the scene is available or not. If the complete geometrical description about the scene is known, many approaches about automatic sensor placement are described in [9][15]. The problem is different if no *a priori* information about the scene is available *i.e.,* if the sensor is in an unknown environment. It raises the problem of autonomous exploration [8][14][16][19][18][4]. In [8], the sensor placement is computed from a local map of the scene which is described by an octree. The first proposed solution, called the "planetarium algorithm", gives for all the camera positions on a sphere located around the scene, the viewpoint from which the maximal amount of unexamined area will be visible.

A second solution, the "normal algorithm", which uses the internal structure of the octrees is also proposed. In [14], Maver and Bajcsy use informations given by occlusions to plan the next viewing direction. In [19], Wixson describes strategies to search for a known object in a cluttered area. Three strategies for sensor placement are studied and compared: the "model-based strategy" based on the Connoly's algorithm, the "occlusion-based strategy" which uses occluding edges to restrict attention to areas that have not been checked yet, and a strategy which simply rotates the camera around the scene with a fixed rotation increment. In [12], Kutulakos presents an approach for exploring a 3D surface, using a mobile monocular camera, which is based on the use of the occlusion boundary. In [18], Whaite and Ferrie present a system which creates a 3D model of the environment using the data gathered by a laser range-finding system through a sequence of exploratory probes. In order to minimize the uncertainty of the parametric forms (such as superquadric) used to describe the scene, a feedback based on the model uncertainty is used as a basis for selecting viewpoints.

Our concern is to deal with the problem of recovering the 3D spatial structure of a whole scene without any knowledge on the localization and the dimension of the different geometrical primitives of the scene (assumed to be composed of polygons, cylinders and segments). Since the proposed structure estimation method involves to successively focus on each primitive of the scene, developing perception strategies to get the complete spatial organization of complex scenes is thus necessary. Integrating knowledge on 3D data previously gathered, and current 2D information into an exploration process allows us to determine the next primitive to be estimated or the next camera viewpoint.

The remainder of this paper is organized as follows: Section 2 is devoted to the local aspect of our reconstruction scheme and briefly describes the structure from controlled motion framework. Section 3 is devoted to the global aspect and deals with the development of perception strategies. Exploration strategies based on a partial 3D model of the scene, 2D visual features extracted from the images and a representation of the observed areas are proposed. We demonstrate with various real time experiments that the implemented active vision system allows the reconstruction of complex scenes with a very good accuracy.

2 Structure From Controlled Motion

The method we have used to estimate the 3D structure of the primitives assumed to be present in the scene is described in [6]. It is based on the measure of the camera velocity and the corresponding motion of the primitive in the image. More precisely, if $\underline{p}$ is the set of parameters describing the 3D structure of a primitive, we have:

$$\underline{\widehat{p}} = \underline{\widehat{p}}(\underline{P}, \underline{\dot{P}}, T_c) \tag{1}$$

where:

- $\underline{\widehat{p}}$ is the estimated value of $\underline{p}$;
- $\underline{P}$ is the set of parameters describing the 2D position of the perspective projection of the primitive in the image;
- and $\underline{\dot{P}}$ is the measured time variation of $\underline{P}$ due to the applied camera velocity T_c.

This approach has been applied to the most representative primitives (*i.e.*, point, straight line, circle, sphere and cylinder) [6]. As far as cylinders are concerned, this method provides the 3D orientation and position of their axis, as well as their radius. For a segment, it provides the 3D orientation and position of the straight line to which the segment belongs.

When no particular strategy concerning camera motion is defined, important errors on the 3D structure estimation can be observed. This is due to the fact that the quality of the estimation is very sensitive to the nature of the successive camera motions. An active vision paradigm is thus necessary to improve the accuracy of the estimation results by generating adequate camera motions. It has been shown in [6] that two vision-based tasks (called fixation and gazing tasks) have to be realized in order to obtain a robust and non biased estimation. The visual servoing approach [10] is perfectly suitable to perform such tasks. Dealing with cylinders or segments, they must appear centered and vertical (or horizontal) in the image [6]. For a cylinder, this estimation scheme can be applied using only the projection of one limb, but a two limbs-based estimation provides more robust and precise results.

Length Estimation In order to determine the length and the position of the primitive along its axis, its vertices have to be observed in the image, which generally implies a complementary camera motion. A motion, around the $\vec{x}$ (or $\vec{y}$) axis in the camera frame, is thus performed until one of the two endpoints of the primitive appears at the image center (see Figure 1). Once the camera has reached its desired position, the 3D position of the corresponding end point is simply computed as the intersection between the primitive axis and the camera optical axis. A motion in the opposite direction is then generated to determine the position of the other endpoint. Such camera motions, based on visual data, are again performed using the visual servoing approach.

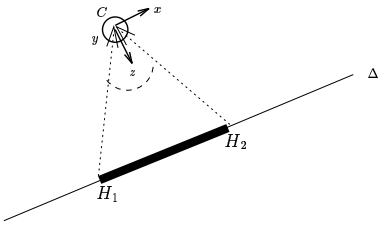


Figure 1: *Camera motion for length estimation*

A Maximum Likelihood Ratio Test for Primitive Recognition The only information we initially have on the considered scene is composed by the set of 2D segments observed by the camera at its initial position.

We assume that each segment corresponds to the projection in the image of either a limb of a cylinder, either a 3D segment. Since the structure estimation method is specific to each kind of primitives, a preliminary recognition process is required. To determine the nature of the observed primitive, we first assume that it is a cylinder, and a one limb-based estimation is performed. When this estimation is done, two competing hypotheses can be acting. Respectively:

- H_0: the observed primitive is a straight line. This hypothesis implies that we have to find a radius r close to 0 ;
- H_1: the observed primitive is a cylinder. This hypothesis implies that we have to find $r = r_1$ with $r_1 > 0$;

A maximum likelihood ratio test is used to determine which one of these two hypotheses is the right one [13]. The likelihood ratio ξ is given by $\xi = \frac{N\bar{r}^2}{2\sigma^2}$ where N is the number of estimations, $\bar{r}$ is the mean value of the estimated radius, and σ^2 its variance. Hypothesis H_1 (cylinder) is selected versus hypothesis H_0 (segment) if the obtained value for the likelihood ratio ξ is greater than a given threshold (which can be easily determined by experiment). Indeed, when the primitive is a segment, the reconstruction process using one limb gives a low radius, with a very high variance. On the other hand, when the primitive is a cylinder, the estimated radius is close to its real value and its variance is small. A two limbs-based estimation is then performed using for the initial detection of the second limb in the image a simple matching between the segments observed in the image and the projection of the estimated cylinder.

Experimental results The whole application presented in this paper has been implemented on an experimental testbed composed of a CCD camera mounted on the end effector of a six degrees of freedom cartesian robot. The image processing part is performed in real time on a commercial image processing board. It consists in tracking the projection of the selected straight lines or limbs along the image sequence [5].

We here presents the results obtained for the structure estimation of a cylinder. Figure 2.a represents the initial image acquired by the camera and the selected cylinder. Figure 2.b contains the image acquired by the camera after the convergence of the visual servoing task.

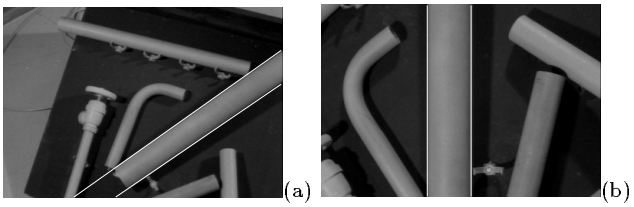
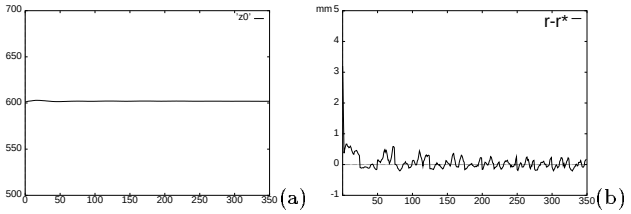

(a) (b)

FIGURE 2: *Position of the cylinder in the image before and after the focusing task*



(a) (b)

FIGURE 3: *Estimation of the parameters of a cylinder in the camera frame (a) estimated cylinder depth (in mm) (b) error between the real and estimated radius of the cylinder (in mm)*

Figure 3.a describes the evolution of the estimated depth of the cylinder displayed on Figure 2. Figure 3.b reports the error between the real value of its radius and the estimated one. These results underline the fact that our estimation algorithm is particularly robust, stable and accurate.

3 Camera Control Strategies

We are now interested in investigating the problem of recovering a precise and complete description of a 3D scene containing several objects using the visual reconstruction scheme presented above. As already stated, this scheme involves fixating at and gazing on the different primitives in the scene. This can be done on only one primitive at a time, hence reconstructions have to be performed in sequence for each primitive of the scene. After each estimation of a primitive, an exploration process is required to determine the next selection and to ensure the completeness of the scene reconstruction.

3.1 Incremental scene exploration

Our concern is to deal with the problem of recovering the 3D spatial structure of a whole scene without any knowledge on the number, the localization and the dimensions of the different geometrical primitives of the scene. Thus, we have to determine viewpoints able to bring new primitives in the field of view of the camera. Such viewpoints will be computed using the previously estimated 3D map and the part of the 3D scene which has not been already observed.

Completeness of the reconstruction. From a viewpoint ϕ_t, we first compute the camera field of view. Then, using the current 3D map of the scene, we can compute the volume $V(\phi_t)$ observed from this viewpoint. We denote $\mathcal{V}(\Phi_t)$ the area observed by the camera from the beginning of the reconstruction process. We have:

$$\mathcal{V}(\Phi_t) = \bigcup_{i=1}^{t} V(\phi_i), \text{ with } \quad \Phi_t = \bigcup_{i=1}^{t} \phi_t$$

The scene reconstruction process takes end when:

$$\forall \phi_{t+1}, \mathcal{V}(\Phi_t) \cup V(\phi_{t+1}) = \mathcal{V}(\Phi_t).$$

This means that the exploration process is as complete as possible if for all reachable viewpoints, the camera looks at a known part of the scene. We thus can be sure that all

the areas of the scene are either free space, either an object which has been reconstructed, either an un-observable area. We now describe how ϕ_{t+1} is determined.

A two levels algorithm for scene exploration. Our incremental strategy leads to an exploration process which is handled at two levels:

- When a new primitive appears in the field of view of the camera, or has been previously observed, it is estimated. In that case, we do not need to compute explicitly new viewpoints. This level is called **local exploration**. It allows to split the observed areas into free-space and reconstructed objects.
- When a local exploration ends, a more complex strategy has to be implemented in order to focus on parts of the 3D space which have not been already observed. This level is called **global exploration**.

3.2 Local exploration

As already stated, the scene is assumed to be only composed of polyhedral objects and cylinders, so that the contours of all the objects projected in the image plane form a set of segments. The first step in the scene reconstruction process is to obtain the list of these segments. It is simply obtained by extracting the edges in the image with a Shen Castan filter, and applying a Hough transform on the edges which computes the equation of the different segments. We denote these lists ω_{ϕ_t}, where ϕ_t is the corresponding camera location. The image processing algorithm we use during the active 3D estimation allows us to only track a limited number of segments at the video rate. Thus, for real time issue, we cannot create a list at each iteration of the estimation process. So, they are created after each reconstruction, and are used for the selection of the next considered segment.

An other list, denoted Ω_Φ, is used. It contains all the unestimated segments previously observed, and the camera position ϕ_t from which they have been observed. More precisely, we have $\Omega_\Phi = \{(\mathcal{S}_i, \phi_k), i = 1 \dots N, k \in [0,t]\}$ where $\mathcal{S}_i$ represents a 2D segment, $\Phi = \bigcup_t \phi_t$ and N is the number of untreated segments.

Step 0 Initialization. We consider that the camera is located in ϕ_0 and ω_{ϕ_0} is acquired. We do not have any information on the parameters of the corresponding 3D primitives. Therefore the 3D map of the scene is initially empty. Thus, initially, $\Omega_\Phi = \omega_{\phi_0} = \{(\mathcal{S}_i, \phi_0), i = 1 \dots n\}$ and $\Phi = \phi_0$. We extract from Ω_Φ a segment $\mathcal{S}_i$ to be estimated. In fact, we choose the segment $\mathcal{S}_i$ which is the nearest from the image position corresponding to the gazing task (horizontal or vertical and centered in the image).

Step 1 Active 3D estimation and 3D map creation. An estimation based on $\mathcal{S}_i$ is performed, including the recognition process and the structure estimation process. The obtained parameters $\widehat{p}$ of the primitive are introduced into the 3D global map of the scene.

Step 2 Local and global 2D lists generation. After the active estimation, because of the camera motion implied by this process, the camera is located in ϕ'_t. A new local database $\omega_{\phi'_t}$ corresponding to this position is constructed.

Using the projection of the current 3D map into the image, a simple matching algorithm is performed in order to determine the set of segments of $\omega_{\phi'_t}$ which have been previously estimated (see Figure 4.b where the dashed lines correspond to segments which are the projection of estimated primitives). The matched segments are then suppressed from $\omega_{\phi'_t}$ which is merged with the global 2D list Ω_Φ (thus, $\Phi = \cup_{i=1}^{t} \phi_t \cup \phi'_t$, and Ω_Φ contains all the segments which have been observed from all the previous viewpoints, and which have not been estimated yet).

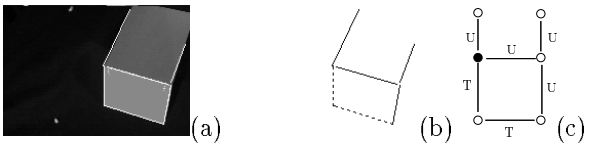


FIGURE 4: *(a) Image acquired (b) 2D database and results of the matching (c) neighboring graph*

Step 3 Segment selection. If one (ore more) unestimated segment is in the current list $\omega_{\phi'_t}$, the new camera position is chosen as $\phi_{t+1} = \phi'_t$ and a new segment $\mathcal{S}_i$ is chosen. An **active estimation** (step 1) based on this segment is then performed. In the case where several unestimated segments are in the current list, a choice is performed in order to select the next chosen segment. Using the ω'_{ϕ_t} and the current 3D map, a neighboring graph is computed where the nodes are composed of the junctions between segments and where the vertices represent the state of the segment *i.e.*, treated (T) or untreated (U) (see Figure 4c). Using this graph, we look for an unestimated segment connex to the last estimated one. If such a segment exists, it is selected. Otherwise, we choose the untreated segment the nearest from the optimal position for its robust 3D estimation. We iterate the steps **estimation, 2D lists creation** and **selection** until one of the segments present in the current list $\omega_{\phi'_t}$ has not been estimated.

Backtracking. If all the segments of $\omega_{\phi'_t}$ have been considered and if at least one of the 2D segments previously observed have not been estimated (*i.e.*, $\omega_{\phi'_t}$ empty and Ω_Φ not empty), we look in Ω_Φ for the couple $(\mathcal{S}_i, \phi_k)$, for which the distance between the current camera location ϕ'_t and the location ϕ_k (from which the segment $\mathcal{S}_i$ has been observed) is minimal. Then, the camera moves to the position ϕ_k (thus, $\phi_{t+1} = \phi_k$). An active estimation (step 1) is then performed. Finally, if Ω_Φ is empty (*i.e.*, all the 2D segments observed from any previous camera positions have been treated), a new viewpoint must be found. A **global exploration** is thus necessary.

Results The example reported here (see Figure 5) deals with a scene composed of a cylinder (whose radius is 40 mm) and five polygons which lie in different planes. In Figure 6.a is displayed the initial image acquired by the camera. Note that the whole scene is not in the camera field of view for that position.

Figure 6 shows the images acquired before each optimal estimation and the corresponding list of segments. Fig-

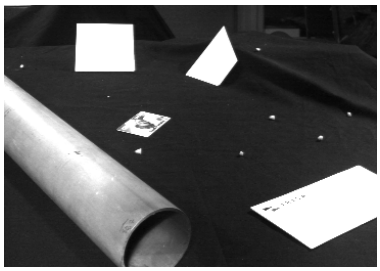


Figure 5: *External view of the scene*

ure 6.a shows the image acquired from the initial position ϕ_0 of the camera. At this time, no reconstruction has been performed and only three segments appear in the field of view of the camera. The first segment extracted from ω_{ϕ_0} is the right limb of the cylinder. After the recognition process and the estimation based on the two limbs, the camera is located in ϕ_1 (see image 6.b). All the segments in the list ω_{ϕ_1} have been treated. Using Ω_Φ, the segment which has been previously observed from the position ϕ_0 and has not been treated yet is selected. Thus the camera moves to ϕ_0 and gazes on this segment. After the reconstruction of the corresponding primitive, camera is located in ϕ_2 (see image 6.c). One of the two segments connex to the last estimated one is selected and reconstructed. The process is iterated until all the primitives observed during this local exploration process are reconstructed (*i.e.*, obtained at the position ϕ_6, see image 6.d). Note that several primitives which did not appear in the initial field of view of the camera have been detected and reconstructed. The 3D model of the scene at this step of the reconstruction process is displayed on Figure 7.

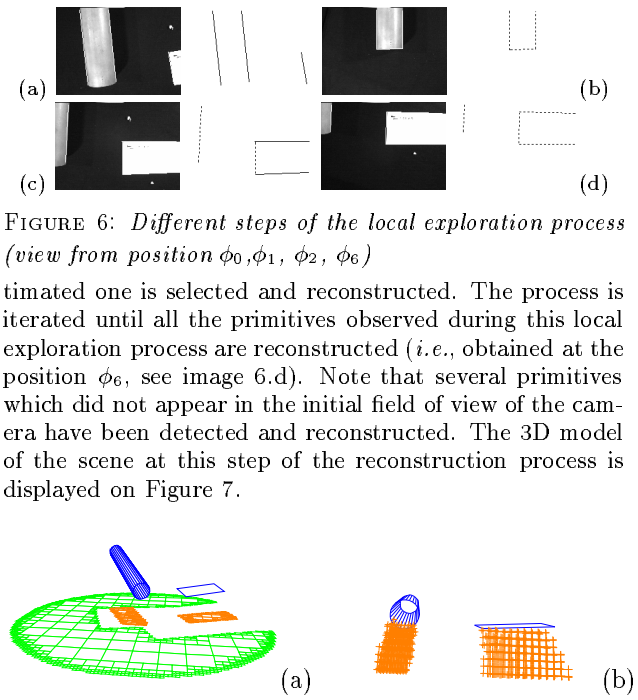


FIGURE 6: *Different steps of the local exploration process (view from position ϕ_0,ϕ_1, ϕ_2, ϕ_6)*

FIGURE 7: *Results of the first local exploration process: (a) Reconstructed scene and projection on a virtual plane of the unknown area (b) Reconstructed scene and volumetric representation of the occluded area*

Since this exploration strategy is local, it avoids computing explicitly new viewpoints. Let us note that some objects or complex scenes can not be completely recovered using this local exploration. The composition of simple primitives, such as polygons, can be treated by this algorithm but more complex combinations raise new problems: an object can be occluded by another one (or by itself) or may not have been observed from the different viewpoints. Exploration probes are thus necessary to make sure that the whole scene has been reconstructed.

3.3 Global exploration

We have chosen to formulate the probing strategy as a function minimization. We define a function to be minimized which reflects the quality of a new viewpoint. It integrates the expected gain of the new position, and the cost of the displacement from the previous to the new position. This leads to a gaze planning strategy that mainly uses a representation of known and unknown areas, obtained after the local exploration, as a basis for selecting viewpoints.

Viewpoint Selection. The function $\mathcal{F}$ to be minimized must integrate the constraints imposed by the robotic system and evaluate the quality of the viewpoint. Thus, we define a set of independent measures which define the quality or the badness of a viewpoint. As in [16], each result of a given measure belongs to $[0, 1]$ (or has an infinitive value for unreachable positions). A value near 0 results from an ideal situation. The function $\mathcal{F}$ to be optimized is taken as a weighted sum of this set of measures.

- The quality of a new position ϕ_{t+1} is defined by the volume of the unknown area which appears in the field of view of the camera. The new observed area is given by $\mathcal{G}(\phi_{t+1})$ where (see Figure 8) :

$$\mathcal{G}(\phi_{t+1}) = \mathcal{V}(\phi_{t+1}) - \mathcal{V}(\phi_{t+1}) \cap \mathcal{V}(\Phi_t) \qquad (2)$$

  where $\mathcal{V}(\phi_{t+1})$ defines the part of the scene observed from the position ϕ_{t+1} and $\mathcal{V}(\phi_{t+1}) \cap \mathcal{V}(\Phi_t)$ defines the sub-part of $\mathcal{V}(\phi_{t+1})$ which has been already observed.

  The measure of the quality of the position ϕ_{t+1} is then given by :

$$g(\phi_{t+1}) = 1 - \frac{volume(\mathcal{G}(\phi_{t+1}))}{volume(\mathcal{V}(\phi_{t+1}))} \qquad (3)$$

  Remark: In fact, $\mathcal{G}(\phi_t)$ defines the potential volume of unknown area using the current knowledge on the 3D scene. If a new object appears in the camera field of view, the new observed area will be smaller than the expected one ($\mathcal{G}'(\phi_t) \subseteq \mathcal{G}(\phi_t)$).

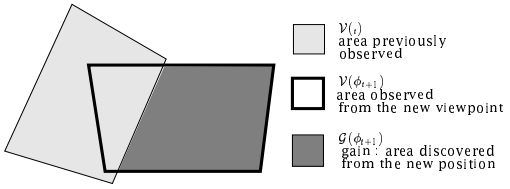


Figure 8: *Quality of a new position (2D projection)*

- To compute the cost of the camera displacement between viewpoints ϕ_t and ϕ_{t+1}, we use the distance between these two positions. More precisely, this measure

is given by:

$$\mathcal{C}(\phi_t, \phi_{t+1}) = \frac{1}{N_{ddl}} \sum_{i=1}^{N_{ddl}} \beta_i \frac{\mid q_{i_t} - q_{i_{t+1}} \mid}{\mid Q_{i_{Max}} - Q_{i_{Min}} \mid} \quad (4)$$

where

- N_{ddl} is the number of robot degrees of freedom ;
- q_i is the position of the robot joint i (note that : $\phi = (q_0, q_1, \ldots, q_{N_{ddl}})$).
- $\mid Q_{i_{Max}} - Q_{i_{Min}} \mid$ gives the distance between the joint limits on axis i.
- β_i are weights which allow to fix the relative importance of an axis with respect to the others (for instance, rotationnal motions may be preferred to translationnal ones).

- Furthermore, additional constraints are associated to camera locations. The goal of these constraints is :
  - to avoid unreachable viewpoints (*i.e.*, camera location out of the joint limits of the robot. This is a binary test which returns an infinite value when the position is unreachable:

$$\mathcal{A}(\phi) = \begin{cases} 0 & \text{if } \phi \text{ is reachable} \\ \infty & \text{else} \end{cases} \quad (5)$$

  - to avoid positions near the robot joint limits. The measure associated to this constraint is optimal (equal to 0) if the camera is located at the middle of the extension of each axis of the robot:

$$\mathcal{B}(\phi) = \frac{1}{N_{ddl}} \sum_{i=1}^{N_{ddl}} \frac{4(q_i - \frac{Q_{i_{Max}} + Q_{i_{Min}}}{2})^2}{(Q_{i_{Max}} - Q_{i_{Min}})^2} \quad (6)$$

The function $\mathcal{F}(\phi_{t+1})$ to be minimized is thus defined as :

$$\mathcal{F}(\phi_{t+1}) = \mathcal{A}(\phi) + \alpha_1 g(\phi_{t+1}) + \alpha_2 \mathcal{C}(\phi_t, \phi_{t+1}) + \alpha_3 \mathcal{B}(\phi) \quad (7)$$

Determining the weights α_i is not a simple problem. Here, the weights are predetermined in order to reflect the relative importance of the different measures. For example, the gain of a new position is more important than the cost of the camera displacement. We have defined a priority order of the coefficients α_i such that $\alpha_1 > \alpha_2 > \alpha_3$.

Optimization. We have to determine the position ϕ_{best} which minimize the energy function $\mathcal{F}(\phi)$. Each position $\phi \in SE_3$ could be *a priori* a solution of this optimization problem. In order to avoid possible obstacles, we allow the camera to only move on an hemisphere located around the scene (assumed to be inside the hemisphere). Thus, the camera location is described by a vector with five parameters $(\theta, \varphi, \Omega_x, \Omega_y, \Omega_z)$ where θ is the latitude and φ the longitude of the camera on the hemisphere, and Ω_x, Ω_y and Ω_z are the angles which define the camera orientation.

To minimize $\mathcal{F}(\phi)$, we chose to use a fast deterministic relaxation scheme corresponding to a modified version of the ICM algorithm. First, $\mathcal{F}(\phi)$ is minimized using large variation steps of the parameters. When the minimum is found, the process is iterated with smaller variation steps. Unlike stochastic relaxation methods such as simulated annealing, we cannot ensure that the global minimum of the function is reached. However our method is not timecomsuming and experimental results show that we always get a correct minimum in a low number of iterations. Furthermore, in our problem, finding the global minimum at each iteration of the exploration is not really necessary as long as the new viewpoint discovers a large part of the scene.

Results. Figure 9 shows the different steps of the global exploration of the scene. Each figure shows the obtained 3D scene, the camera trajectory and the projection on a virtual plane of the unknown areas. Figure 9.a corresponds to the camera position ϕ_6 obtained just after the local exploration process described in the previous paragraph. The first camera displacements (see Figure 9.b and 9.c) allows to reduce significantly the unknown areas. At position ϕ_{13} (see Figure 9.d) a new primitive is detected. A new local exploration process is performed. It ends at position ϕ_{24} (Figure 9.f). At this step, the two polygons on the "top" of the scene have been reconstructed. A new global exploration is then performed. It moves the camera to position ϕ_{25} (Figure 9.g) where a new segment which belongs to the last object of the scene appears in the field of view of the camera. After a last local exploration process, the four segments of this polygon are reconstructed and the camera is located in ϕ_{30} (Figure 9.h). At this step, 99% of the space has been observed, which ensures that the reconstruction of the scene is complete. Figure 10 shows the final 3D model of the scene (to be compared to Figure 5 and the camera trajectory.

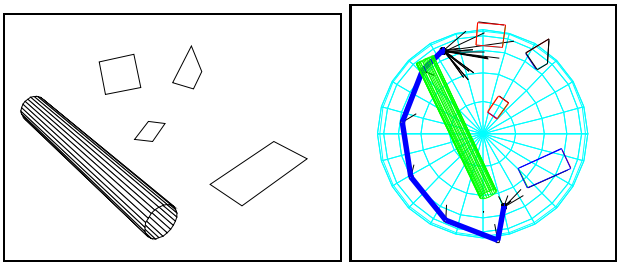


FIGURE 10: *3D model of the reconstructed scene and polar view of the camera trajectory*

Influence of the weight in the energy function. We want here to analyze the influence of the weights α_i involved in (7) on the camera trajectory. We consider a scene with only two objects: a cylinder and a polygon which have been reconstructed during a first local exploration/reconstruction process. In the first strategy (Figure 11.a), the distance between two viewpoints is not taken into account, thus this strategy is mainly based on the maximization of the new observed area (the weight α_2 in (7) is null). The second strategy (Figure 11.b) uses the distance between this two viewpoints in order to minimize the total distance covered by the camera. Figure 11.c shows the distance covered by the camera versus the number of

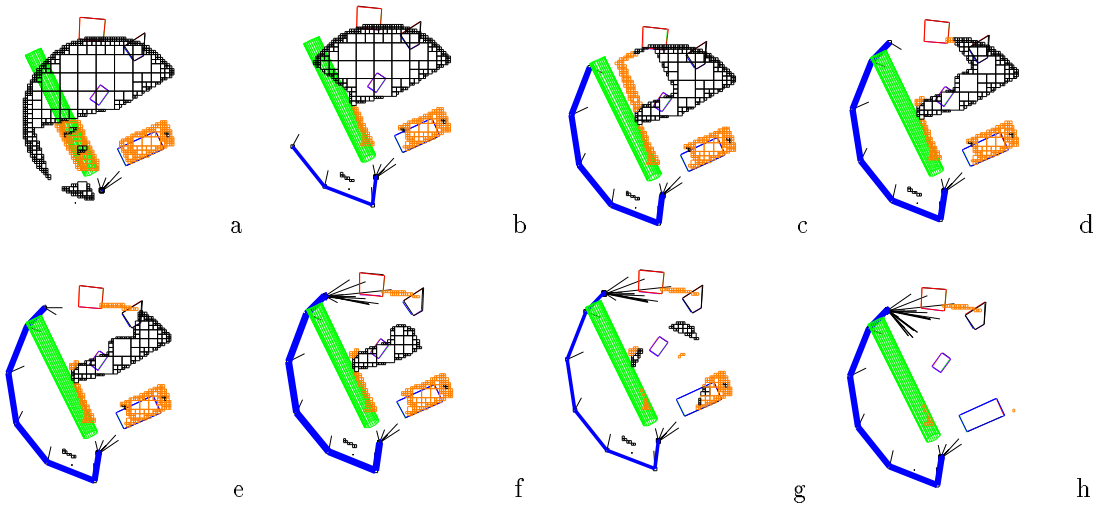


FIGURE 9: *Different steps of the global exploration process (camera trajectory, 3D model of the final reconstructed scene and projection on an virtual plane of the unknown area)*

viewpoints for both strategies. We note that if the distance between two viewpoints is not taken into account, the camera motion behaves like a "bee flight". Such motion does not occur if the distance cost is introduced into the energy function (the camera motion is more continuous). This underlines the interest in introducing the distance parameter into the energy function.

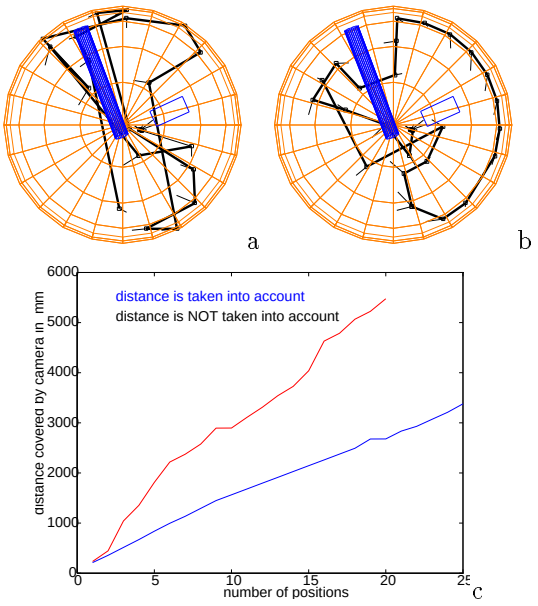


FIGURE 11: *Global exploration of the scene. Camera motion with (a) $\alpha_2 = 0$ (b) $\alpha_2 > 0$. (c) Distance covered by the camera for the both strategies*

Seeing behind occlusions Due to the constraints imposed on the camera positions (located on an hemisphere around the scene), our method is not able to ensure the absolute completness of the reconstruction (especially for occluded areas). However, different approaches can be used in order to cope with the occlusions problem: we can use the same global exploration algorithm but restricted to small areas located around the unknown/occluded areas (see the reconstruction of a polyhedron on Figure 12). We thus have to take care about possible obstacles (which can be introduced in the cost function to be minimized). Furthermore, it seems to be useful to gaze on the occluding contours/segments and then move to observe the area occluded by these contours (see for example [12][14]).

4 Conclusion

In this paper, we have proposed a method for 3D environment perception using a sequence of images acquired by a mobile camera. We have described a reconstruction process which provides an accurate and robust estimation of the parameters of a geometrical primitive. As this method is based on peculiar camera motions, perceptual strategies able to appropriately perform a succession of such individual primitive reconstruction have been proposed in order to recover the complete spatial structure of complex scenes. An important feature of our approach is its ability to easily determine the next primitive to be estimated without any knowledge or assumption on the number, the localization and the spatial relation between objects. Furthermore, we do not have to define complex planning strategies. Our strategy can be defined as an on-line strategy (VS an off-line strategy where a plan must be previously computed). Our approach is entirely bottom-up and does not use any *a priori* on the environment except the nature of the considered primitives.

Finally, experiments carried out on a robotic cell have proved the validity of our approach (accurate, stable and robust results, efficient exploration algorithms), but have also shown its limitations (the constraints on the camera motion, which are necessary to obtain precise results, im-

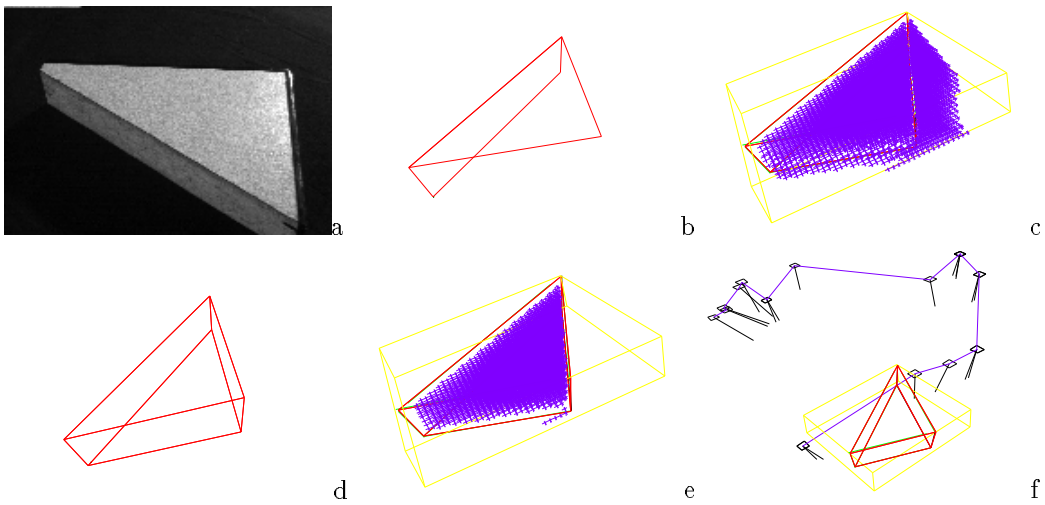


FIGURE 12: *Polyhedron reconstruction (a) image of the scene, (b) two polygons have been reconstructed after the local exploration (c) unobserved areas after the local, the other areas are occluded (d) Model computed at the end of the reconstruction process (e) unobserved areas after the global exploration (all the remaining unobserved areas are inside the polyhedron) (f) camera trajectory*

ply the sequencing of visual estimations and we cannot perform several active estimations in parallel). Future work will thus be devoted to determine optimal camera motions for a simultaneous structure estimation of a set of geometrical primitives.

References

[1] G. Adiv. Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(5):477–489, May 1989.

[2] Y. Aloimonos. Purposive and qualitative active vision. In *IAPR Int. Conf. on Pattern Recognition*, volume 1, pages 346–360, Atlantic City, New Jersey, June 1990.

[3] R. Bajcsy. Active perception. *Proc. of the IEEE*, 76(8):996–1005, August 1988.

[4] A. Blake, A. Sisserman, and R. Cipolla. Visual exploration of free-space. *Active Vision* (A. Blake and A. Yuille, eds), MIT Press, 1994.

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

[6] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin. Optimal estimation of 3D structures using visual servoing. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 347–354, Seattle, USA, June 1994.

[7] C. Chien and J.K. Aggarwal. Model construction and shape recognition from occluding contour. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(4):372–389, February 1989.

[8] C. Connoly. The determination of next best views. In *IEEE Int. Conf. on Robotics and Automation*, pages 432–435, St Louis, Missouri, USA, March 1985.

[9] C.K. Cowan and P.D. Kovesi. Automatic sensor placement from vision task requirements. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(3):407–416, May 1988.

[10] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.

[11] K.N. Kutulakos and C.R. Dyer. Recovering shape by purposive viewpoint adjustment. *International Journal of Computer Vision*, 12(2):113–136, February 1994.

[12] K.N. Kutulakos, C.R. Dyer, and V. Lumelsky. Provable strategies for vision-guided exploration in three dimensions. In *IEEE Int. Conf. on Robotics and Automation*, pages 1365–1372, San Diego, California, USA, June 1994.

[13] E. Marchand and F. Chaumette. Active visual 3D perception. In *IEEE Workshop on Vision for Robots*, pages 10–17, Pittsburgh, USA, August 1995.

[14] J. Maver and R. Bajcsy. Occlusions as a guide for planning the next view. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(5):417–433, May 1993.

[15] K. Tarabanis, R. Tsai, and P.K. Allen. Automated sensor planning for robotic vision tasks. In *IEEE Int. Conf. on Robotics and Automation*, volume 1, pages 76–82, Sacramento, California, April 1991.

[16] B. Triggs and C. Laugier. Automatic camera placement for robot vision. In *IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 1732–1738, Nagoya, Japon, May 1995.

[17] J. Weng, T.S. Huang, and N. Ahuja. Estimation and structure from line matches: Performance obtained and beyond. In *IAPR Int. Conf. on Pattern Recognition*, volume 1, pages 168–172, Atlantic City, New Jersey, USA, June 1990.

[18] P. Whaite and F. Ferrie. Autonomous exploration: Driven by uncertainty. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 339–346, Seattle, USA, June 1994.

[19] L.E. Wixson. Viewpoint selection for visual search. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 800–805, Seattle, USA, June 1994.