Active sensor placement for complete scene reconstruction and exploration

Eric Marchand[1], Franois Chaumette
IRISA - INRIA - Université de Rennes 1
Campus de Beaulieu, 35042 Rennes Cedex, France
E-mail {marchand, chaumett}@irisa.fr

Abstract

This paper deals with the 3D structure estimation and exploration of a scene using active vision. We have used the structure from controlled motion approach to obtain a precise and robust estimation of the 3D structure of geometrical primitives. Since it involves to gaze successively on the considered primitives, 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. 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. The proposed strategy ensures the completeness of the reconstruction.

1 Overview

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 descripcvption of a static scene using the visual data provided by a camera mounted on the end effector of a robot arm. The idea of using active schemes to address vision issues has been recently introduced [1][2]. Active vision is defined in [2] 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 [1]. 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 [10].

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

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

where:

- $\widehat{\underline{p}}$ 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 $\dot{\underline{P}}$ 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.*, points, straight lines, circles, spheres and cylinders) [3]. 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 [3] 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 [6] is perfectly suitable to perform such tasks. Dealing with cylinders or segments, they must appear centered and vertical (or horizontal) in the image [3].

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 the complete geometrical description about the scene is known [5][13] and if the sensor is in an unknown environment. In that case, it raises the problem of autonomous exploration [4][7][11][15][17][16]. In [4], the sensor placement is computed from a local map

[1] Éric Marchand is currently at Yale University, Dpt of Computer Science, PO Box 208-285, New Haven, CT 06520-8285, E-mail: marchand@cs.yale.edu

of the scene which is described by an octree. The proposed solution, called the "planetarium algorithm", gives for all camera positions on a sphere located around the scene, the viewpoint from which the maximal amount of unexamined area will be visible. In [11], Maver and Bajcsy use information given by occlusions to plan the next viewing direction. In [17], 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 [8], 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 [16], Whaite and Ferrie present a system which creates a 3D model of the environment using the data gathered by a laser range-finder system through a sequence of exploratory probes. Our concern is to deal with the problem of recovering the 3D spatial structure of a whole scene without any knowledge on the localization, the number, and the dimensions of the different geometrical primitives of the scene (assumed to be composed of polygons, cylinders and segments). Since the proposed structure estimation method 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. 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 are occluded or have not been already observed. This level is called **global exploration**.

2 Incremental scene 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. We denote these lists $\omega_{\phi_t} = \{\mathcal{S}_i, i = 1 \dots M\}$, where ϕ_t is the corresponding camera location from which the M segments $\mathcal{S}_i$ are observed. 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 $\Omega_{\mathcal{T}_{t_1}^{t_2}}$, is used. It contains all the untreated segments previously observed, and the camera positions ϕ_k from which they have been observed. More precisely, we have:

$$\Omega_{\mathcal{T}_{t_1}^{t_2}} = \{(\mathcal{S}_i, \phi_k), i = 1 \dots N, k \in [t_1, t_2]\}$$

where

- $\mathcal{T}_{t_1}^{t_2} = \{\phi_{t_1}, \phi_{t_1+1}, \dots, \phi_{t_2}\}$ is the set of all viewpoints between t_1 and t_2. Initially, $\mathcal{T}_0^0 = \{\phi_0\}$ is the first camera location.
- $\mathcal{S}_i$ is a 2D segment associated with an unestimated primitive and ϕ_k is the camera location from which it has been observed. ϕ_k belongs to $\mathcal{T}_{t_1}^{t_2}$. N is the number of untreated segments.

Using these two sets of segments it is possible to define an incremental reconstruction strategy:

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 about the parameters of the corresponding 3D primitives. Therefore the 3D map of the scene is initially empty: $\Omega_{\mathcal{T}_0^0} = \omega_{\phi_0} = \{(\mathcal{S}_i, \phi_0), i = 1 \dots n\}$ and $\mathcal{T}_0^0 = \{\phi_0\}$. We extract from ω_{ϕ_0} a segment $\mathcal{S}_i$ to be estimated.

Step 1 Active 3D estimation and 3D map creation. Let us consider now that the camera is located in ϕ_t, ($\phi_t = \phi_0$ if this is the beginning of the exploration process). An estimation based on $\mathcal{S}_i$ is performed, including a recognition process [9] (does this segment correspond to a 3D segment or to a cylinder ?), the structure estimation process (see previous section and [3]) and the estimation of the primitive length [10]. The obtained parameters $\widehat{\underline{p}}$ of the primitive are introduced into the 3D global map of the scene. We then remove from $\Omega_{\mathcal{T}_0^t}$ all the 2D segments, corresponding to this estimated primitive.

Step 2 Local and global 2D lists generation. After the active estimation, because of the camera motion implied by this process (see [3]), the camera is located in ϕ_{t+1}. A new local set of segments $\omega_{\phi_{t+1}}$ corresponding to this position is constructed and merged with $\Omega_{\mathcal{T}_0^t}$:

$$\Omega_{\mathcal{T}_0^{t+1}} = \Omega_{\mathcal{T}_0^t} \cup \omega_{\phi_{t+1}}$$

Step 3 Segment selection. Three different cases may occur:

1. In the case where several unestimated segments are in the current list $\omega_{\phi_{t+1}}$, a choice is performed in order to select the next chosen segment $\mathcal{S}_i$ (see [9] for more details). An **active estimation** (step 1) based on this segment is then performed. We iterate the steps **estimation, 2D lists creation** and **selection** until one of the segments present in the current list $\omega_{\phi_{t+1}}$ has not been estimated.
2. If all the segments of $\omega_{\phi_{t+1}}$ have been considered and if at least one of the 2D segments previously observed have not been estimated (*i.e.*, $\omega_{\phi_{t+1}}$ empty and $\Omega_{\mathcal{T}_0^{t+1}}$ not empty), we look in $\Omega_{\mathcal{T}_0^{t+1}}$ for the couple $(\mathcal{S}_i, \phi_k)$, for which the distance between the current camera location ϕ_{t+1} 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.
3. Finally, if $\Omega_{\mathcal{T}_0^{t+1}}$ 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**, which is described in the next section, is thus necessary.

Experimental results The example reported here (see Figure 1.a) deals with a scene composed of a cylinder (whose radius is 40 mm) and five polygons which lie in different planes. In Figure 2.a is displayed the initial image acquired by the camera.

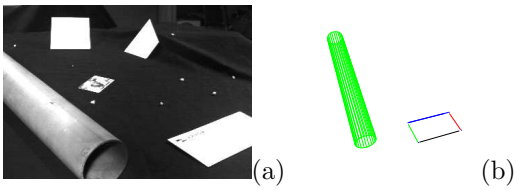

Figure 1: *External view of the scene and results of the first local exploration process.*

Figure 2 shows the images acquired before each optimal estimation and the corresponding list of segments. On Figure 2.a, the segment at the middle of the image is chosen and an estimation of the cylinder parameters is performed, first using one limb and then using both limbs (see [9] for details). After this reconstruction, all the segments observed from the current camera viewpoint belong to estimated primitives (see Figure 2.b). Thus, we look in $\Omega_{\mathcal{T}_0^{t+1}}$ for an untreated segment. Here, the camera moves to the previous position (Figure 2.a) and the camera gazes on the segment on the right of the image. The process is iterated

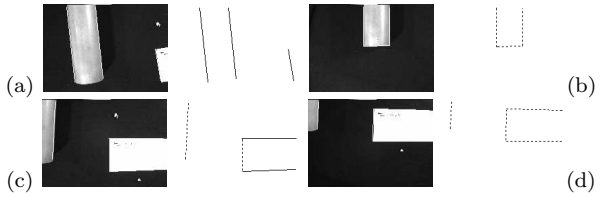

Figure 2: *Different steps of the local exploration process: image and list of segments from position ϕ_0,ϕ_1, ϕ_2 and ϕ_6 (dashed lines correspond to segments which are the projection of estimated primitives).*

until all the primitives observed during this local exploration process are reconstructed which is obtained at the position ϕ_6 (see Figure 2.d). Note that several primitives which did not appear in the initial camera field of view have been detected and reconstructed. At this step of the reconstruction process, the 3D model of the scene, quite incomplete, is displayed on Figure 1.b.

Since this exploration strategy is local, it avoids computing explicitly new viewpoints. Furthermore, the composition of simple primitives, such as polygons, can be treated by this algorithm. However, more complex combinations raise new problems: an object can be occluded by another one (or by itself). Finally, as shown on the previous results, several objects 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 Global exploration

3.1 Exploration Strategy

We consider a scene composed of a set $\mathcal{O}$ of initially unknown primitives. At the end of the local exploration, a subset $\mathcal{O}(\mathcal{T}_0^t) \subseteq \mathcal{O}$ has been observed and reconstructed. Thus, we have to determine viewpoints able to bring more information about the scene. By *information*, we mean either a new primitive, either the certainty that a given area is object-free. Such viewpoints will be computed using the previously estimated 3D map and the part of the 3D scene which has not been already observed. If a new primitive is observed from the computed viewpoint, the local exploration process is used to estimate its parameters.

Knowing the set $\mathcal{T}_0^t$ of viewpoints since the beginning of the reconstruction process, it is possible to maintain a map of the observed and unexplored areas. The knowledge is thus composed by:

- the primitive already estimated: $\mathcal{O}(\mathcal{T}_0^t)$;
- the known free space, denoted $\mathcal{V}(\mathcal{T}_0^t)$. Knowing the position of the primitives, and the position ϕ of the camera, it is possible to compute the area $\mathcal{V}(\phi)$ observed from this position (using a ray tracing scheme). Thus, knowing the trajectory $\mathcal{T}_0^t$, we can determined $\mathcal{V}(\mathcal{T}_0^t)$ the area observed from the beginning of the reconstruction process. We have $\mathcal{V}(\mathcal{T}_0^t) = \bigcup_{i=0}^{t} \mathcal{V}(\phi_i)$. This area can also be determined incrementally if we consider only the additional information given by the last viewpoint. Thus we have:

$$\left\{ \begin{array}{lcl} \mathcal{V}(\mathcal{T}_0^t) & = & \mathcal{V}(\mathcal{T}_0^{t-1}) \cup \mathcal{V}(\phi_t) \\ \mathcal{V}(\mathcal{T}_0^0) & = & \mathcal{V}(\phi_0) \end{array} \right. \quad (2)$$

- the unknown area $\mathcal{U}(\mathcal{T}_0^t)$: knowing the location of the reconstructed objects and the known free space, $\mathcal{U}(\mathcal{T}_0^t)$ is computed as:

$$\mathcal{U}(\mathcal{T}_0^t) = \overline{\mathcal{V}(\mathcal{T}_0^t) \cup \mathcal{O}(\mathcal{T}_0^t)} \quad (3)$$

We want to ensure the completeness of the reconstruction. In theory, the reconstruction must end when all the space has been observed, *i.e.*, if at instant t:

$$\mathcal{U}(\mathcal{T}_0^t) = \emptyset \quad (4)$$

However, this condition is usually unreachable. Ensuring the completeness of the reconstruction is not always possible. Some areas may be observed only from a set of viewpoints unreachable by the camera. Furthermore, due to the objects topology, some areas may be unobserved whatever the position of the camera. Thus we use the following termination condition:

$$\forall \phi_{t+1}, \left\{ \begin{array}{l} \mathcal{V}(\mathcal{T}_0^t) \cup \mathcal{V}(\phi_{t+1}) = \mathcal{V}(\mathcal{T}_0^t) \\ \mathcal{O}(\mathcal{T}_0^t) \cup \mathcal{O}(\phi_{t+1}) = \mathcal{O}(\mathcal{T}_0^t) \end{array} \right. \quad (5)$$

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, at the end of the exploration process, all the areas of the scene are either free-space, either an object which has been reconstructed, either an unobservable area.

3.2 Viewpoint Selection.

A simple strategy able to compute the "next best view" ϕ_{t+1} is to consider the viewpoint which maximizes the volume of the new observed areas [4] [17]. However the final trajectory may be ineffective dealing with the final distance covered by the camera. Furthermore, such a strategy do not take into account some problems such as the manipulator kinematics constraints or geometric constraints. We have thus to introduce some constraints in the strategy in order to reflect these different aspects of the exploration problem. We have chosen to formulate the probing strategy as a function minimization problem. Like in [13][15][12], we define a function to be minimized which integrates the constraints imposed by the robotic system and evaluates the quality of the viewpoint. Thus, we use a set of independent measures which determine the quality or the badness of a viewpoint. 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.

Quality of a new position 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:

$$\mathcal{G}(\phi_{t+1}) = \mathcal{V}(\phi_{t+1}) - \mathcal{V}(\phi_{t+1}) \cap \mathcal{V}(\mathcal{T}_0^t) \tag{6}$$

where $\mathcal{V}(\phi_{t+1})$ defines the part of the scene observed from the position ϕ_{t+1} and $\mathcal{V}(\phi_{t+1}) \cap \mathcal{V}(\mathcal{T}_0^t)$ defines the sub-part of $\mathcal{V}(\phi_{t+1})$ which has been already observed (see Figure 3). If the position ϕ_{t+1} does not give any wage of information (*i.e.* $\mathcal{G}'(\phi_t) = \emptyset$), we must reject this position. Thus, the value of the measure g is fixed to infinity in that case. The measure of the quality of the position ϕ_{t+1} is then given by:

$$g(\phi_{t+1}) = \begin{cases} \infty & \text{if } \mathcal{G}(\phi_{t+1}) = \emptyset \\ 1 - \frac{volume(\mathcal{G}(\phi_{t+1}))}{volume(\mathcal{V}(\phi_{t+1}))} & \text{else} \end{cases} \tag{7}$$

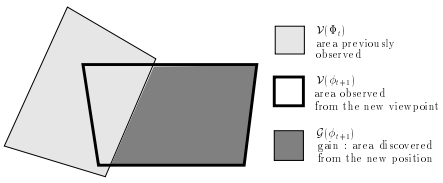

Figure 3: *Quality of a new position (2D projection).*

Displacement Cost. A term reflecting the cost of the camera displacement between two viewpoints ϕ_t and ϕ_{t+1} is introduced in the cost function $\mathcal{F}$, in order to reduce the total camera displacement. It is defined using the following relation:

$$\mathcal{C}(\phi_t, \phi_{t+1}) = \frac{1}{N_{dof}} \sum_{i=1}^{N_{ddl}} \beta_i \frac{\mid q_{i_t} - q_{i_{t+1}} \mid}{\mid Q_{i_{Max}} - Q_{i_{Min}} \mid} \tag{8}$$

where:

- N_{dof} is the number of robot degrees of freedom ;
- q_i is the position of the robot joint i and $\mid Q_{i_{Max}} - Q_{i_{Min}} \mid$ gives the distance between the joint limits on axis i ;
- β_i are weights setting the relative importance of an axis with respect to the others ($\beta_i \in [0, 1]$). For instance, rotational motions may be preferred to translational ones.

Additional Constraints. Additional constraints are associated to camera locations. The goal of these constraints is:

- to avoid unreachable viewpoints. This is a binary test which returns an infinite value when the position is unreachable:

$$\mathcal{A}(\phi) = \begin{cases} 0 & \text{if } \phi \text{ is reachable} \\ \infty & \text{else} \end{cases} \tag{9}$$

  A position is unreachable if it is not in the operational space of the manipulator, or if this position is located in an unknown area (leading to a collision risk).
- to avoid positions near the robot joint limits. When a new primitive is observed from the computed viewpoint ϕ_{t+1}, an optimal estimation of its parameters must be performed. This estimation requires camera motions performed by visual servoing which can not be realized if the robot encounters a joint limit. 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:

$$\mathcal{B}(\phi) = \frac{1}{N_{ddl}} \sum_{i=1}^{N_{dof}} \frac{4(q_i - \frac{Q_{i_{Max}} + Q_{i_{Min}}}{2})^2}{(Q_{i_{Max}} - Q_{i_{Min}})^2} \tag{10}$$

Note that unlike in mobile robotics we do not have odometry problems and due to the quality of the reconstruction, it is not necessary to perform other estimations of the primitives and develop merging processes which would have implied the introduction of other constraints (such as overlap constraint) in our objective function. However, this algorithm can also be used with reconstruction scheme such as stereovision or laser range finder. Constraints introduced in the optimization function can be changed function of the sensor or robot caracteristics (*e.g.,* scannings, tolerance or overlap constraints in the case of a laser range finder [12] or of a mobile robot).

The function $\mathcal{F}(\phi_{t+1})$ to be minimized is thus defined as a weighted sum of the different measures:

$$\mathcal{F}(\phi_{t+1}) = \mathcal{A}(\phi) + \alpha_1 g(\phi_{t+1}) + \alpha_2 \mathcal{C}(\phi_t, \phi_{t+1}) + \alpha_3 \mathcal{B}(\phi) \tag{11}$$

Here, the weights are predetermined in order to reflect the relative importance of the different measures. For example, the wage of information given by 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 $\alpha_1 > \alpha_2 > \alpha_3$. More precisely we have fixed $\alpha_1 = 0.6$, $\alpha_2 = 0.3$ and $\alpha_3 = 0.1$.

We have decided to constrain the camera viewpoints inside an hemisphere located around the scene (assumed to

be inside the hemisphere), but only in the region already observed and object-free (in order to avoid collision). At the beginning of the exploration process, as the observed area is null, the camera motion is limited to the surface of the sphere. To minimize $\mathcal{F}(\phi)$, we have used a fast deterministic relaxation scheme corresponding to a modified version of the ICM algorithm. 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 time-consuming and experimental results show that we always get a correct minimum in a low number of iterations.

3.3 Focusing on the regions of interest

At the end of the reconstruction process, some residual areas remain unexplored. This is mainly due to the termination criterion we have proposed (see equation (5)). First of all, it is not possible to ensure that 100% of the scene can be observed: the topology of the objects, the kinematic constraints of the manipulator prevent from observing the whole space. Thus, small parts of the scene usually remain unobserved. Second, the marginal gain of information decreases rapidly while the number of viewpoints increases. Thus, even if the whole scene is observable, the observation of the last residual areas requires a large number of viewpoints. For these different reasons, we decide to stop the exploration when a subset of the observable space has been really observed (typically, we define a threshold located between 95% and 97% of the observable space). However, it is necessary to verify that the remaining unobserved areas do not contain any objects (and if any to perform a reconstruction).

In a first time, we compute a segmentation of the residual areas considering the polyhedron incorporating a set of connex unobserved areas (defined here by small voxels). As the goal is to define small regions, if the volume of the computed polyhedrons is to important, they are subdivided. In a second time, considering sequentially each sub-scene, a focused global exploration algorithm is then performed. Like in the case of the global exploration algorithm we consider that the camera motion is limited inside an hemisphere located around this polyhedron. The knowledge previously acquired is used in order to consider a collision avoidance process. This strategy allows to decrease significantly the number of viewpoints while increasing the part of observed areas (leading to 99%).

3.4 Reducing the number of viewpoints

The technique proposed in the previous paragraphs to solve the "next best view" problem is a depth-first search algorithm. We do no try to consider the whole trajectory in order to reduce either the number of viewpoints, either the distance performed by the camera. The following algorithm aims at reducing the number of viewpoints needed to explore the scene. The problem can be stated as follow: at instant t the camera has already performed the trajectory $\mathcal{T}_0^t$; knowing the set of reconstructed objects $\mathcal{O}(\mathcal{T}_0^t)$ and the observed area $\mathcal{V}(\mathcal{T}_0^t)$, find the trajectory $\mathcal{T}_{t+1}^{t_{fin}}$ which ensures a complete reconstruction of the scene in a minimal number of viewpoints.

Even knowing a complete model of the scene, the problem is, in the general case, NP-complete [14]. Furthermore, to determine such a trajectory, we need to know the shape and the location of the objects. In our case, we do not have this information. Thus this scheme is efficient if all the objects present in the scene have been already reconstructed at time t. Then, the trajectory $\mathcal{T}_{t+1}^{t_{fin}}$ can be computed. However if a new primitive is observed from viewpoint $\phi_{t'}, t' \in [t+1, t_{fin}[$, it must be reconstructed and a new trajectory $\mathcal{T}_{t'}^{t_{fin}}$ must be determined.

It is possible, using the method proposed in Section 3.2 to obtain a set of viewpoints $\mathcal{T}_0^{n_0}$ such that the whole scene is observed from the $n_0 = t_{fin}$ viewpoints. As previously stated, at the end of the exploration process, the marginal gain of each viewpoint is weak. A local adjustment of the position of a viewpoint along the computed trajectory may be preferable to the addition of a new viewpoint. Furthermore, it is possible that the information given by a viewpoint at time t is also given by other viewpoints at time $t' > t$. This viewpoint ϕ_t is thus useless and can be deleted. If n_0 is the initial number of viewpoints, the iterative algorithm allowing to define a trajectory which ensures a reconstruction as complete as possible of the scene in n viewpoints ($n \leq n_0$) can be decomposed in two steps:

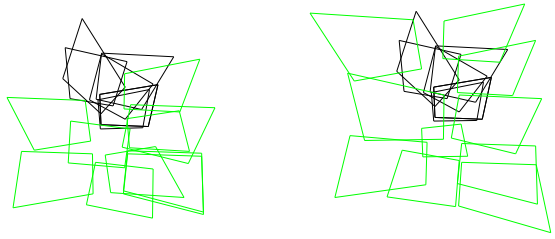

Figure 4: *Projection on a virtual plane of the view cone associated to each camera viewpoints $\phi \in \mathcal{T}_{t+1}^n$. (a) The volume of observed area has been deliberately limited to 75%. (b) After a set of local adjustments (the number of viewpoint remains constant), more than 95% of the scene is observed.*

In a first time, we want to know if some local adjustments of the trajectory could increase the quantity of available information. Thus we consider each viewpoint $\phi_i \in \mathcal{T}_{t+1}^{t_{fin}}$ and look for a viewpoint $\phi_{i'}$ such that:

$$volume(\mathcal{U}(\mathcal{T}_0^{t_{fin}})) \leq volume(\mathcal{U}(\mathcal{T'}_0^{t_{fin}}))$$

where $\mathcal{T'}_0^{t_{fin}} = \{\phi_0, \ldots, \phi_{i-1}, \phi'_i, \phi_{i+1}, \ldots, \phi_{t_{fin}}\}$ and where $\mathcal{U}$ represents the unknown area (see equation (3)). To compute $\phi_{i'}$, we use the proposed computing viewpoints algorithm considering in (6) $\mathcal{G}(\phi_{i'})$ as:

$$\mathcal{G}(\phi_{i'}) = \mathcal{V}(\phi_{i'}) - \mathcal{V}(\phi_{i'}) \cap \mathcal{V}(\mathcal{T}_0^{t_{fin}} - \phi_i).$$

If such a viewpoint exists, ϕ_i is replaced with $\phi_{i'}$ in the trajectory. This process is done for each viewpoint of the trajectory $\mathcal{T}_{t+1}^{t_{fin}}$ and iterated while the location of a viewpoint is modified. Figure 4 describes the result of this local adjustment algorithm.

In a second time, if the whole scene is observed, the viewpoint ϕ_{min}, bringing the lowest quantity of information is suppressed from the trajectory. The local adjustment process is then used again in order to compensate for this deletion. This algorithm is iterated while the observed areas remains greater than a predefined threshold (99% in our case).

3.5 Experimental results

First, we consider the case of a simple scene composed by a cylinder and a polygon in order to illustrate the influence of the different proposed strategies. All the objects have been reconstructed during a first local exploration/reconstruction process (see Figure 5). The observed/unobserved/occluded areas are computed using a ray tracing algorithm.

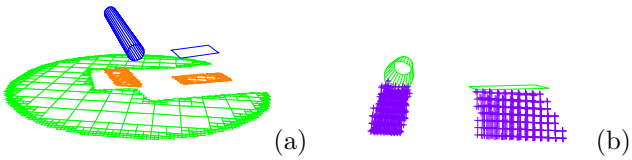

Figure 5: *Results of the first local exploration process: (a) Reconstructed scene and projection on a virtual plane of the unknown area (b) Reconstructed scene and volumetric representation of the occluded area.*

In this first experiment, we want to analyze the influence of the weights α_i involved in (11). In the first strategy (Figure 6.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 (11) is null). The second strategy (Figure 6.b) uses the distance between two successive viewpoints in order to reduce the total distance covered by the camera 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 and shorter (1.13 m versus 3.37 m with $\alpha_2 = 0$). This underlines the interest in introducing the distance parameter into the energy function.

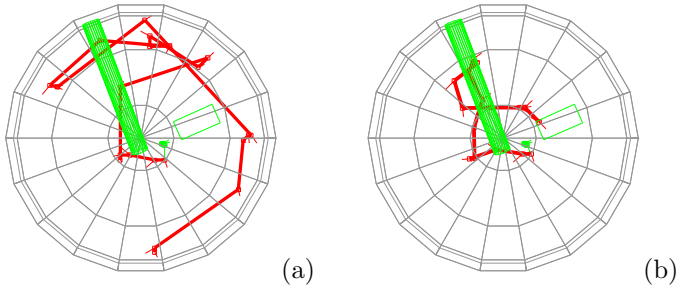

Figure 6: *Global exploration of the scene. Camera motion with (a) $\alpha_2 = 0$ (b) $\alpha_2 = 0.3$.*

Focusing on the regions of interest Using the same scene, we decide to stop the exploration when 95% of the scene has been observed. Then, we compute a segmentation of the residual areas (see Figure 7.a and 7.b). Four regions of interest are created. The camera focuses on each of these regions in order to ensure that no other objects can be observed in the scene. The remaining area unobserved (Figure 7.c) is less that 1% and mostly corresponds to occluded regions which cannot be observed due to robot kinematic constraints. Figure 8 depicts the results obtained for the reconstruction of another scene, a polyhedron, using this algorithm.

Reducing the number of viewpoints Using the algorithm proposed in Section 3.4, we can reduce the number of viewpoints from 29 (see Figure 9.a) to 15 (see Figure 9.b) without any waste of information. Figure 10 shows the percentage of observed area (after local adjustement) versus the number of viewpoint. Above 20 viewpoints, the percentage of observed area is greater than 99%, and below 15 viewpoints the percentage of observed area is smaller than the initial one. However, reducing to 15 the number of viewpoints requires approximatively 15 minutes on a SUN SS20 while only 10 seconds are necessary to compute a new viewpoint. Thus, we think that this optimization algorithm is interesting only if time is not critical. It is also more adapted to the inspection of a known scene than to the exploration of an unknown environment.

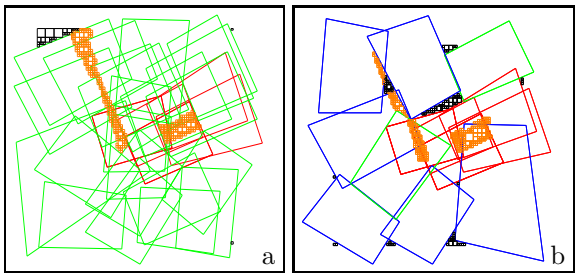

Figure 9: *Projection on a virtual plane of the view cone associated to each camera viewpoints $\phi \in \mathcal{T}_0^n$ (a) Initial trajectory ($n = 29$) (b) final computing trajectory ($n = 15$).*

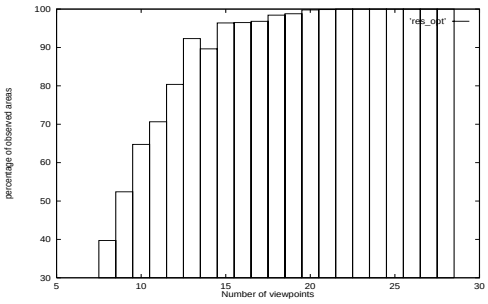

Figure 10: *Percentage of observed areas versus the number of viewpoints.*

Cylinder and polygons scene In this experiment, we consider the scene proposed in Section 2. Figure 11

presents 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 11.a corresponds to the camera position ϕ_6 obtained just after the local exploration process described in Section 2. The first camera displacements allows to reduce significantly the unknown areas. At position ϕ_{13} (see Figure 11.b), a new primitive is detected. A new local exploration process is performed. It ends at position ϕ_{24} (Figure 11.c). At this step, the two polygons on the "top" of the scene have been reconstructed. A new global exploration is then performed.

After a last exploration process, the last polygon is reconstructed and the camera is located in ϕ_{30} (Figure 11.d). At this step, 99% of the space has been observed, which ensures that the reconstruction of the scene is complete. Figure 12 shows the final 3D model of the scene (to be compared to Figure 1) and the camera trajectory.

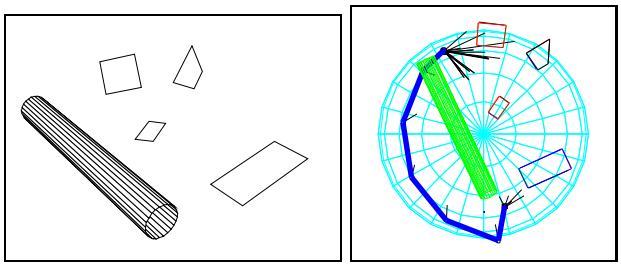

Figure 12: *3D model of the reconstructed scene and polar view of the camera trajectory*

4 Conclusion

In this paper, we have proposed a method for 3D environment perception using a sequence of images acquired by a mobile camera. Since the method used for reconstruction is based on peculiar camera motions, perceptual strategies able to appropriately perform a succession of individual primitive reconstruction have been proposed. 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. Our approach is entirely bottom-up and does not use any *a priori* on the environment except the nature of the considered primitives and a bounding volume located around the scene. 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, imply the sequencing of visual estimations.

References

[1] Y. Aloimonos, I. Weiss, and A. Bandopadhay. – Active vision. – *Int. Journal of Computer Vision*, 1(4):333–356, January 1987.

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

[3] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin. – Structure from controlled motion. – *IEEE Trans. on PAMI*, 18(5):492–504, May 1996.

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

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

[6] 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.

[7] G. Hager and M. Mintz. – Computational methods for task-directed sensor data fusion and sensor planning. – *Int. Journal of Robotics Research*, 10(4):285–313, August 1991.

[8] 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*, pp. 1365–1372, San Diego, June 1994.

[9] E. Marchand and F. Chaumette. – Controlled camera motions for scene reconstruction and exploration. – In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'96*, pp. 169–176, San Francisco, June 1996.

[10] E. Marchand. – *Stratégies de perception par vision active pour la reconstruction et l'exploration de scènes statiques.* – PhD Thesis, Université de Rennes 1, IRISA, June 1996.

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

[12] R. Pito. – A sensor based solution to the next best view problem. – In *IAPR Int. Conf. Pattern Recognition, ICPR'96*, pp. 941–945, Vienna, Austria, August 1996.

[13] K. Tarabanis, R. Tsai, and P.K. Allen. – The MVP sensor planning system for robotic vision tasks. – *IEEE Trans. on Robotics and Automation*, 11(1):72–85, February 1995.

[14] G.H. Tarbox and S.N. Gottschlich. – Planning for complete sensor coverage in inspection. – *CVGIP: Image Understanding*, 61(1):81–111, January 1995.

[15] B. Triggs and C. Laugier. – Automatic camera placement for robot vision. – In *IEEE Int. Conf. on Robotics and Automation*, pp. 1732–1738, Nagoya, Japan, May 1995.

[16] P. Whaite and F. Ferrie. – Autonomous exploration: Driven by uncertainty. – In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pp. 339–346, Seattle, June 1994.

[17] L.E. Wixson. – Viewpoint selection for visual search. – In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pp. 800–805, Seattle, June 1994.

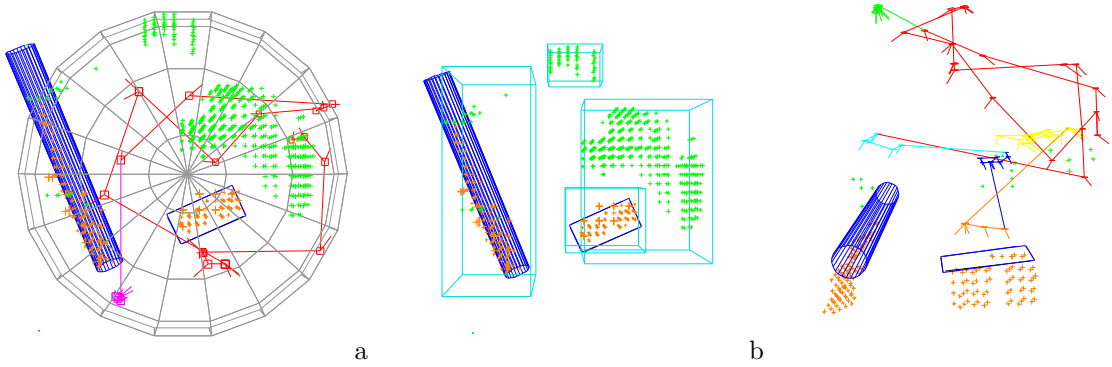

Figure 7: *Focusing on the regions of interest : (a) observed/unobserved areas and camera trajectory before focusing, (b) detected regions of interest, (c) observed/unobserved areas and camera trajectory after focusing.*

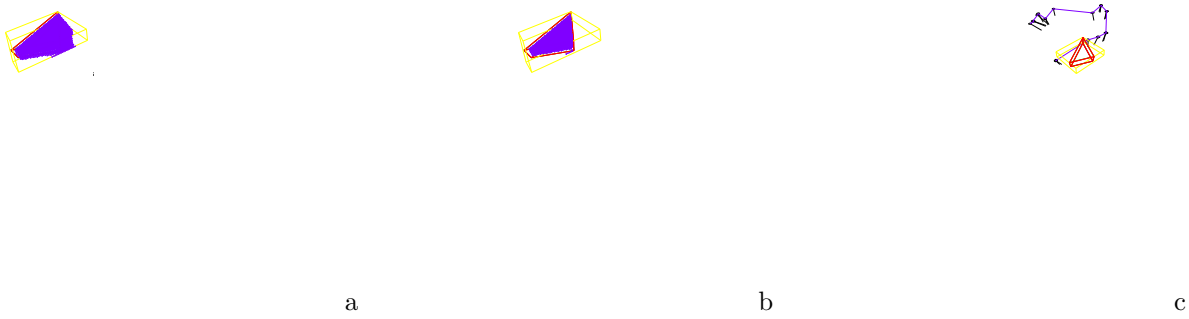

Figure 8: *Polyhedron reconstruction : (a) residual unobserved areas after a first exploration, (b) model computed at the end of the reconstruction process, (c) camera trajectory.*

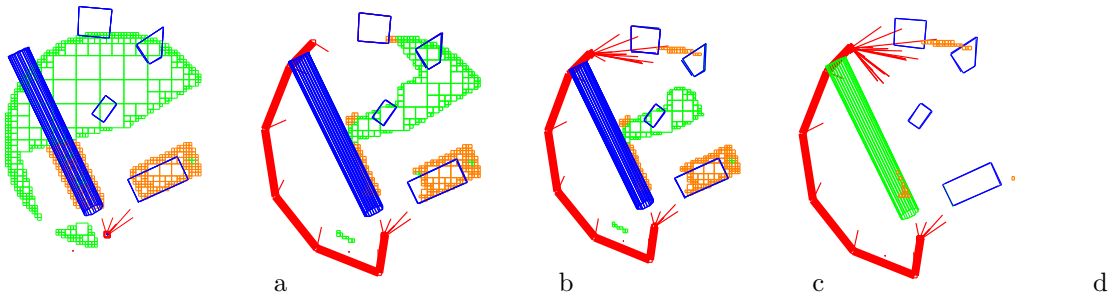

Figure 11: *Different steps of the global exploration process (camera trajectory, 3D model of the final reconstructed scene and projection on an virtual plane of the unknown area).*