RECHERCHE

Contrôle actif d'une caméra pour la reconstruction et l'exploration de scènes

Éric Marchand – François Chaumette

IRISA - INRIA Rennes - Université de Rennes I
Campus Universitaire de Beaulieu
35042 Rennes Cedex

RÉSUMÉ. *Nous proposons dans cet article une méthode de reconstruction et d'exploration de scènes fondée sur un concept de vision active. L'approche retenue est basée sur le contrôle du mouvement de la caméra, celui-ci étant contraint de manière à optimiser la reconstruction de chaque primitive de la scène. Cette approche nécessitant une focalisation sur les objets considérés (dans notre cas des segments et/ou des cylindres), nous présentons des stratégies de perception permettant l'enchaînement de ces focalisations et la complétude de la reconstruction. Aucune connaissance sur le nombre, la position et les dimensions des objets constituant la scène n'étant disponible, une méthode de recherche du positionnement optimal de la caméra utilisant les informations 2D ainsi que la carte 3D déjà acquise est présentée.*

ABSTRACT. *This paper deals with the 3D structure estimation of a set of geometrical objects in an active vision context. Our method is based on the structure from controlled motion approach which consists in constraining the camera motion in order to obtain an optimal 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, the shape and the localization of the different objects. Furthermore, the proposed strategy ensures the completeness of the reconstruction. This leads to a gaze planning strategy that mainly uses a representation of known and unknown areas as a basis for selecting viewpoints.*

MOTS-CLÉS : *Vision Active, Reconstruction 3D, Exploration de scènes*
KEY WORDS : *Active Vision, Structure from Controlled Motion, Scene exploration*

Technique et science informatiques. Volume 16 – n°1/1997, pages 101 à 129

1. Introduction

Une perception correcte de l'environnement est un élément clé de l'indépendance que l'on souhaite donner aux robots. Depuis quelques années, de nombreux travaux menés en vision artificielle se sont fixés pour objectif la réalisation de systèmes capables d'accéder à la géométrie spatiale d'une scène à partir de son observation par une ou plusieurs caméras mobiles [ADI 89], [CHI 89], [ESP 87] [FAU 93], [WEN 90], [WAX 87], [XIE 89]. Ces systèmes ont pour but de fournir une description géométrique claire et complète de la scène à partir de séquences d'images souvent bruitées et difficilement exploitables. L'étude proposée ici tente d'apporter sa contribution au problème de la reconstruction d'environnements assez restreints (objets statiques, hypothèses sur la nature des objets constituant la scène,...) avec comme ambition la reconstruction de scènes de plus en plus complexes. L'approche retenue pour l'obtention d'une carte précise de l'environnement est basée sur le concept de la **vision active** [ALO 87], [ALO 90], [BAJ 88], [SWA 93]. La vision active peut être vue comme un processus d'acquisition "intelligent" des données. Les problèmes soulevés lors de la conception de systèmes de vision sont principalement leur sensibilité au bruit et leur faible précision. Le principe de la vision active est d'élaborer des stratégies de contrôle des paramètres de la caméra (position, mise au point, vitesse,...) de manière à améliorer la perception de l'environnement. Aloimonos [ALO 87] aborde la vision active d'un point de vue théorique en proposant une formulation mathématique des problèmes de stabilité, de linéarité et d'unicité des solutions. Bajcsy [BAJ 88] définit la vision active comme une étude des stratégies de perception incluant la modélisation des capteurs, des objets, de l'environnement, ainsi que les interactions entre ces différents acteurs pour la réalisation d'une tâche visuelle précise. Elle se place ainsi dans le paradigme de la **vision intentionnelle** où l'observateur actif est capable de décider du type d'informations et du type de traitements à utiliser pour la résolution d'une tâche précise. Dans cet article, nous utilisons la vision active à deux niveaux différents : un **niveau local**, où les mouvements de la caméra sont contraints de manière à optimiser la qualité des résultats de reconstruction de chaque primitive 3D observée, et un **niveau global**, où les mouvements de la caméra sont commandés de manière à permettre la reconstruction de scènes composées de plusieurs objets, ainsi que l'exploration des zones de la scène non encore observées, assurant ainsi la complétude de la reconstruction.

La mesure du mouvement de la caméra qui est souvent essentielle en reconstruction 3D caractérise un domaine de recherches appelé *vision dynamique*. Deux principales approches ont été mises en œuvre dans la littérature pour traiter ce problème :

— l'approche discrète, dans le sens où une formulation sous forme de déplacement est utilisée dans la reconstruction 3D [CHI 89], [WEN 90], [FAU 93], [ZHA 95]. Un déplacement important de la caméra est généralement effectué entre l'acquisition de deux images successives afin d'obtenir par triangulation des résultats satisfaisants. Cette approche est cependant confrontée au déli-

cat problème de la mise en correspondance des attributs extraits des images acquises.

— l'approche continue, dans le sens où l'on considère une formulation en terme de vitesse. Les images sont alors acquises à une cadence proche de la cadence vidéo. D'un point de vue traitement d'images, cette approche repose soit sur une estimation du champ des vitesses apparent (*optic flow*), qui s'avère généralement peu robuste et instable, notamment autour des contours d'occultation [ADI 89], [WAX 87], soit sur le suivi d'objets d'intérêt (*token trackers*) [ESP 87], [XIE 89].

Pour notre part, nous avons utilisé une approche continue fondée sur le principe de la vision active et dont l'originalité repose sur la génération automatique de mouvements adéquats de la caméra [BOU 93], [CHA 96]. En effet, les résultats de reconstruction sont très sensibles à la nature des mouvements du capteur et aux différentes erreurs de mesure. Plus précisément, cette approche consiste à contraindre les mouvements de la caméra de manière à obtenir une reconstruction 3D précise et robuste de primitives géométriques paramétrables telles que les points, les droites, les cylindres, les sphères, etc. L'asservissement visuel [CHA 90], [ESP 92], qui consiste à considérer le système de vision comme un capteur particulier associé à une tâche et inséré dans une boucle de commande, permet de générer les mouvements de la caméra correspondant à ces contraintes.

Cependant, cette approche ne permet de reconstruire qu'une seule primitive à la fois et ne permet pas d'assurer la complétude de la reconstruction. De plus, une connaissance *a priori* sur la nature de la primitive est nécessaire afin de générer les mouvements adéquats de la caméra. Or, nous souhaitons obtenir une carte de l'environnement sans aucune connaissance sur le nombre, la position et la dimension des objets de la scène (composée par hypothèse de cylindres droits circulaires, de polygones et de segments). Il faut donc s'abstraire des contraintes imposées par la méthode de reconstruction en définissant des stratégies de perception qui permettent l'acquisition d'une carte précise et complète de la scène. Ce problème d'exploration et de focalisation de la caméra correspond à un **niveau global** dans le cadre de notre problème de perception active de l'environnement. La méthode retenue repose sur une reconstruction incrémentale de la scène : chaque primitive 3D sera reconstruite de manière indépendante en utilisant la méthode de reconstruction optimale. Cet aspect **stratégie de perception** peut être divisé en deux niveaux distincts. La première phase de l'exploration est locale au sens où seules les informations disponibles sont utilisées afin de déterminer sur quelle primitive la caméra doit se focaliser. Le second niveau vise à assurer la complétude de la reconstruction de la scène grâce à un calcul explicite de nouveaux points de vue. Plusieurs approches ont été proposées afin de résoudre le problème du calcul automatique de points de vue et de l'exploration de scène [TAR 95a]. Elles peuvent se diviser en deux classes principales en fonction des connaissances éventuelles sur l'environnement. Le cas où une modélisation géométrique complète de la scène est disponible est traité dans [COW 88], [TAR 95b] et [TRI 95]. Le problème de

l'*exploration autonome* est posé dans le cas où la caméra évolue dans un environnement totalement ou partiellement inconnu [CON 85], [KUT 94], [MAV 93], [WHA 94], [WIX 94]. Dans notre cas, le problème à résoudre consiste à choisir les positions et orientations de la caméra permettant d'obtenir le maximum d'informations possible sur les zones encore inconnues de la scène. L'ensemble des points de vue choisis doit conduire à une modélisation complète de la scène, tout en minimisant le nombre de points de vue et/ou la distance parcourue par la caméra.

L'article est structuré de la manière suivante : la première partie rappelle la méthode retenue pour l'estimation des paramètres 3D d'une primitive géométrique à partir de l'utilisation d'une caméra mobile commandable. La seconde partie est dédiée aux aspects globaux du processus de reconstruction et présente des stratégies autonomes d'exploration. Enfin la dernière partie présente une série d'expérimentations réalisées sur une cellule de vision robotique validant les méthodes proposées.

2. Reconstruction de primitives 3D

La méthode utilisée pour reconstruire les primitives 3D est détaillée dans [BOU 93] et [CHA 96]. Elle permet d'obtenir une estimation précise et robuste des paramètres 3D d'une primitive géométrique à partir de l'analyse d'une séquence d'images acquises par une caméra en mouvement. Les mouvements de la caméra nécessaires à la reconstruction sont générés par asservissement visuel [CHA 90], [ESP 92]. Notons qu'un processus de calibration de la caméra est nécessaire à une estimation correcte des paramètres 3D de la primitive considérée.

2.1. *Reconstruction 3D par vision dynamique*

La caméra est modélisée de manière classique par une projection perspective. Sans perte de généralité, la distance focale de la caméra est fixée égale à 1. Un point M de coordonnées $\underline{X} = (X, Y, Z)$ se projette donc en m de coordonnées $\underline{x} = (x, y, 1)$ avec :

$$\underline{x} = \frac{1}{Z}\,\underline{X}. \qquad [1]$$

Soit $\mathcal{P}_s$ une primitive géométrique paramétrable décrite par une équation de la forme :

$$h(\underline{X}, \underline{P}) = 0\ , \quad \forall \underline{X} \in \mathcal{P}_s \qquad [2]$$

où h définit la nature de la primitive et $\underline{P}$ son vecteur de paramètres. Le but de la reconstruction est d'estimer la valeur des paramètres $\underline{P}$ afin de reconstruire et localiser la primitive $\mathcal{P}_s$ définie par h. Soit $\mathcal{P}_i$ la projection dans l'image de $\mathcal{P}_s$. La primitive $\mathcal{P}_i$ peut s'écrire :

$$g(\underline{x}, \underline{p}) = 0\ , \quad \forall \underline{x} \in \mathcal{P}_i \qquad [3]$$

où g définit la nature de la primitive et où la valeur des paramètres $\underline{p}$, dépendant de $\underline{P}$, décrit sa configuration dans l'image.

Par ailleurs, en utilisant les équations de projection perspective (1), l'équation (2) devient :

$$h'(\underline{x}, 1/Z, \underline{P}) = 0. \quad [4]$$

Si l'on exclut les cas dégénérés (tels que $\frac{\partial h'}{\partial Z} = 0$, ce qui se produit par exemple lorsqu'un cercle se projette dans l'image sous la forme d'un segment au lieu d'une ellipse), le théorème des fonctions implicites assure l'existence d'une unique fonction μ tel que :

$$1/Z = \mu(\underline{x}, \underline{P}_L) \quad [5]$$

Pour des primitives planes (cercle,...), la fonction μ représente le plan dans lequel la primitive se situe. Dans le cas plus général de primitives volumiques (cylindre (voir figure 1), sphère, tore,...), la fonction $g(\underline{x}, \underline{p})$ représente la projection dans l'image des limbes de la primitive et la fonction μ, alors appelée *surface des limbes*, exprime la relation entre les points de $\mathcal{P}_i$ et les points de $\mathcal{P}_s$ correspondants. Les paramètres $\underline{P}_L$, dépendant uniquement de $\underline{P}$, décrivent la configuration de cette surface dans le repère de la caméra.

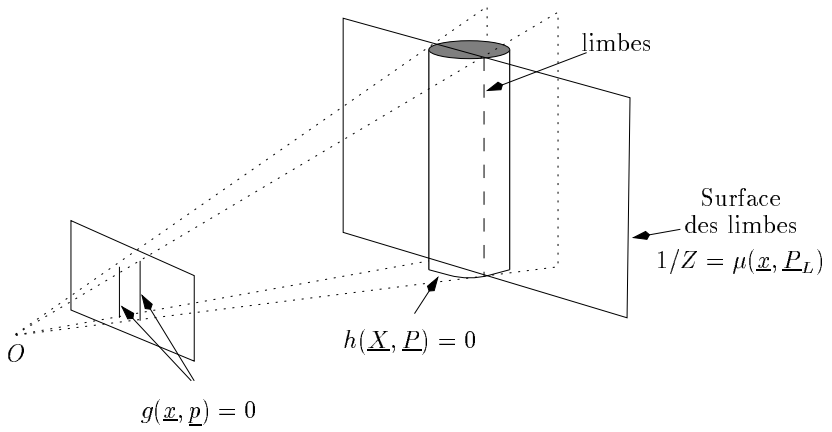


Figure 1. *Projection de la primitive (h) dans l'image (g) et surface des limbes (μ) dans le cas d'un cylindre*

Soit $T_c = (V(O), \Omega)$ le torseur cinématique de la caméra où $V(O) = (V_X, V_Y, V_Z)$ représente la vitesse de translation de la caméra et $\Omega = (\Omega_X, \Omega_Y, \Omega_Z)$ sa vitesse de rotation. La variation de $\underline{P}$ qui relie le mouvement projeté de la primitive dans l'image au mouvement de la caméra T_c peut être calculée explicitement et s'exprime par :

$$\dot{\underline{p}} = L_{\underline{p}}^T(\underline{p}, \underline{P}_L)\, T_c \quad [6]$$

où $L_{\underline{p}}^T(\underline{p}, \underline{P}_L)$, appelée matrice d'interaction associée à $\underline{p}$, caractérise les interactions entre le capteur et la primitive considérée [CHA 90], [ESP 92].

L'estimation des paramètres $\underline{P}$ s'effectue en deux étapes [BOU 93], [CHA 96] : tout d'abord, les paramètres $\underline{P}_L$ caractérisant la surface des limbes μ sont obtenus à partir de l'équation (6) en utilisant la mesure de $T_c, \underline{p}$ et $\dot{\underline{p}}$:

$$\underline{P}_L = \underline{P}_L(T_c, \underline{p}, \dot{\underline{p}}) \qquad [7]$$

Grâce aux contraintes géométriques sur le type de la primitive à reconstruire (droite, cylindre,...) caractérisées par l'équation $h(\underline{X}, \underline{P}) = 0$, il est ensuite possible de remonter, par intersection de la surface des limbes avec le cône de centre O et de génératrice $g(\underline{x}, \underline{p}) = 0$, aux paramètres $\underline{P}$ qui caractérisent la primitive étudiée :

$$\underline{P} = \underline{P}(\underline{P}_L, \underline{p}). \qquad [8]$$

2.2. *Exemple : le cas de la droite*

Une droite peut être représentée par l'intersection de deux plans orthogonaux :

$$h(\underline{X}, \underline{P}) = \begin{cases} A_1X + B_1Y + C_1Z = 0 \\ A_2X + B_2Y + C_2Z + D_2 = 0 \end{cases} \quad \text{avec} \quad \begin{cases} A_1^2 + B_1^2 + C_1^2 = 1 \\ A_2^2 + B_2^2 + C_2^2 = 1 \\ A_1A_2 + B_1B_2 + C_1C_2 = 0. \end{cases} \qquad [9]$$

Une représentation minimale et complète de la droite 2D correspondante est donnée par :

$$g(\underline{x}, \underline{p}) = x\cos\theta + y\sin\theta - \rho = 0 \text{ avec } \begin{cases} \cos\theta &= A_1/\sqrt{A_1^2 + B_1^2} \\ \sin\theta &= B_1/\sqrt{A_1^2 + B_1^2} \\ \rho &= -C_1/\sqrt{A_1^2 + B_1^2}. \end{cases} \qquad [10]$$

De plus, la fonction μ est aisément obtenue à partir de (9) :

$$1/Z = \mu(\underline{x}, \underline{P}_L) = Ax + By + C \text{ avec } \begin{cases} A &= -A_2/D_2 \\ B &= -B_2/D_2 \\ C &= -C_2/D_2 \end{cases} \qquad [11]$$

Enfin, la relation entre le mouvement de la droite dans l'image (défini par $(\dot{\rho}, \dot{\theta})$) et la vitesse de la caméra T_c est donnée par la matrice d'interaction associée aux paramètres (ρ, θ) [ESP 92] :

$$\begin{pmatrix} \dot{\rho} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \lambda_\rho\cos\theta & \lambda_\rho\sin\theta & -\lambda_\rho\rho & (1+\rho^2)\sin\theta & -(1+\rho^2)\cos\theta & 0 \\ \lambda_\theta\cos\theta & \lambda_\theta\sin\theta & -\lambda_\theta\rho & -\rho\cos\theta & -\rho\sin\theta & -1 \end{pmatrix} T_c \qquad [12]$$

avec $\lambda_\rho = -A\rho\cos\theta - B\rho\sin\theta - C$, et $\lambda_\theta = B\cos\theta - A\sin\theta$.

À partir des valeurs mesurées de $\rho, \theta, \dot{\rho}, \dot{\theta}$ et T_c, il faut maintenant estimer les paramètres des deux plans qui définissent la droite considérée.

Les paramètres A_1, B_1 et C_1 sont immédiatement déduit de ρ et θ. Ensuite, comme décrit précédemment, les paramètres décrivant la fonction μ sont déterminés en utilisant la mesure de la vitesse de la caméra et la vitesse apparente de la droite dans l'image qui en résulte. Ainsi, les paramètres A, B, C sont donnés par la résolution du système linéaire suivant :

$$\begin{cases} -A\rho\cos\theta - B\rho\sin\theta - C &= \lambda_\rho \\ -A\sin\theta + B\cos\theta &= \lambda_\theta \\ A\cos\theta + B\sin\theta - C\rho &= 0 \end{cases} \quad [13]$$

où λ_ρ et λ_θ sont obtenus à partir de (12):

$$\begin{cases} \lambda_\rho &= \dfrac{\dot{\rho} + (1+\rho^2)(\Omega_Y\cos\theta - \Omega_X\sin\theta)}{V_X\cos\theta + V_Y\sin\theta - \rho V_Z} \\ \lambda_\theta &= \dfrac{\dot{\theta} + \rho(\Omega_X\cos\theta + \Omega_Y\sin\theta) + \Omega_Z}{V_X\cos\theta + V_Y\sin\theta - \rho V_Z}. \end{cases} \quad [14]$$

On obtient finalement $D_2 = 1/\sqrt{A^2+B^2+C^2}$, $A_2 = -AD_2$, $B_2 = -BD_2$ et $C_2 = -CD_2$.

Des résultats similaires peuvent être obtenus pour d'autres primitives géométriques paramétrables. Le cas des points, cylindres, cercles et sphères est par exemple décrit dans [BOU 93], [CHA 96].

Signalons simplement que dans le cas d'un cylindre on a :

$$\begin{aligned} h(\underline{X}, \underline{P}) &= (X - X_0)^2 + (Y - Y_0)^2 + (Z - Z_0)^2 \\ &\quad -(UX + VY + WZ)^2 - R^2 = 0 \end{aligned} \quad [15]$$

où R est le rayon du cylindre, (U, V, W) représente le vecteur directeur de l'axe du cylindre et (X_0, Y_0, Z_0) sont les coordonnées du point de l'axe du cylindre qui est le plus proche du centre optique O de la caméra. La fonction relative à la surface des limbes μ est donnée par :

$$1/Z = \mu(\underline{x}, \underline{P}_L) = A\,x + B\,y + C \text{ avec } \begin{cases} A = X_0/(X_0^2 + Y_0^2 + Z_0^2 - R^2) \\ B = Y_0/(X_0^2 + Y_0^2 + Z_0^2 - R^2) \\ C = Z_0/(X_0^2 + Y_0^2 + Z_0^2 - R^2) \end{cases} \quad [16]$$

où A, B et C sont les composantes du vecteur normal à la surface des limbes du cylindre. L'image d'un cylindre est constituée de deux droites dont les paramètres ρ et θ peuvent s'exprimer en fonction des paramètres $\underline{P}$. La matrice d'interaction associée à chacune de ces droites est donnée par l'équation (12), la seule différence par rapport au cas précédent portant sur la valeur des paramètres $\underline{P}_L$ intervenant dans λ_ρ et λ_θ. Les paramètres $\underline{P}_L$ sont estimés par résolution d'un système linéaire construit à partir de la matrice d'interaction associée aux deux droites. Il est ensuite possible de remonter aux paramètres $\underline{P}$. Notons que le cylindre peut aussi être reconstruit en utilisant la projection d'un seul de ses limbes [BOU 93].

2.3. *Reconstruction 3D par vision active*

Les résultats obtenus en utilisant la méthode décrite précédemment sont généralement assez médiocres. En effet, la qualité de l'estimation est très sensible à la nature des mouvements de la caméra [ESP 87]. Une solution efficace pour résoudre ce problème est d'utiliser le formalisme de la vision active [ALO 87], [BAJ 88]. Contrairement à la vision dynamique où l'on se contente d'observer le mouvement de la caméra, dans le cas de la vision active, et par analogie avec la vision humaine, le mouvement est commandé. L'asservissement du capteur aux informations présentes dans les images fait de la vision active un processus de perception intelligente des données. Le principal apport de la vision active est l'inclusion de l'asservissement dans le système ainsi que la collecte des données nécessaires à cet asservissement.

On peut montrer [BOU 93], [CHA 96] que, pour obtenir une estimation robuste et non biaisée, il suffit que la caméra réalise une tâche de fixation et une tâche de focalisation. Plus précisément, la primitive doit rester immobile dans l'image ($\underline{\dot{P}} = 0$), et la surface des limbes doit rester immobile dans le repère (mobile) de la caméra ($\underline{\dot{P}}_L = 0$) afin d'assurer une estimation non biaisée (fixation). De plus, la primitive doit apparaître à des positions particulières dans l'image pour assurer une estimation minimisant les effets des erreurs de mesure (focalisation).

Ainsi, dans le cas d'une droite, on obtient une estimation optimale si celle-ci apparait centrée et verticale (ou horizontale) dans l'image, et si, de plus, $V_Y = V_Z = \Omega_X = 0$ (ou $V_X = V_Z = \Omega_Y = 0$) (voir figure 2.a). Enfin, un cylindre doit se projeter en deux droites symétriques parallèles verticales (ou horizontales) dans l'image avec en outre $V_Y = 0$ (ou $V_X = 0$). La trajectoire optimale de la caméra correspond alors à un cercle concentrique au cylindre et perpendiculaire à son axe (voir figure 2.b).

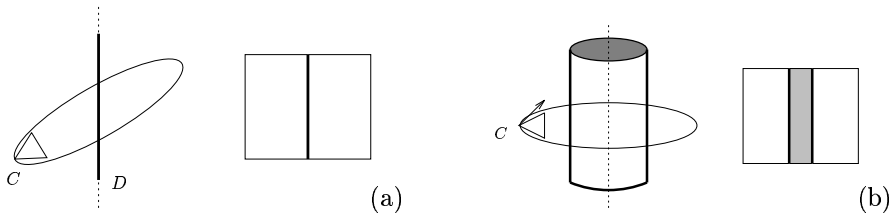


Figure 2. *Mouvements optimaux de la caméra pour l'estimation d'une droite (a) et d'un cylindre (b)*

Le processus de reconstruction décrit précédemment considère que les primitives ont une longueur infinie. De manière à déterminer précisément la position 3D des extrémités de la primitive, celles-ci doivent être successivement observées au centre de l'image, ce qui nécessite un mouvement de la caméra (voir

figure 3). Quand la caméra a atteint la position désirée, la position du point 3D correspondant est simplement calculée par l'intersection de l'axe de la primitive avec l'axe optique de la caméra.

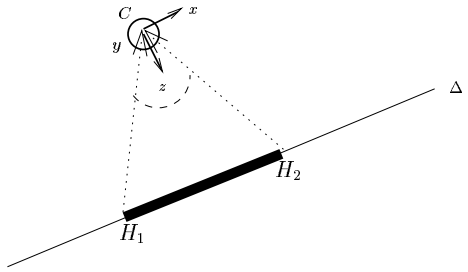


Figure 3. *Calcul de la longueur d'un segment par asservissement visuel*

Les techniques d'asservissement visuel [CHA 90], [ESP 92], [HAS 93] permettent d'effectuer automatiquement les différents mouvements satisfaisant les contraintes nécessaires à une estimation optimale (tâches de fixation, de focalisation, et observation des extrémités des primitives au centre de l'image). L'asservissement visuel consiste en effet à introduire directement et en boucle fermée les informations extraites de l'image dans une boucle de commande. Dans notre cas, les tâches s'expriment comme la régulation à zéro d'une fonction de tâche combinant une tâche primaire (telle que la primitive apparaisse dans l'image à sa position désirée) et une tâche secondaire (ici un suivi de trajectoire prenant en compte l'évitement des butées articulaires du robot). Cette tâche secondaire est construite afin de permettre des mouvements de la caméra tout en assurant la réalisation de la tâche primaire (voir [ESP 92], [BOU 93], [CHA 96] pour de plus amples détails). Signalons finalement que l'utilisation de l'asservissement visuel pour la reconstruction 3D a aussi été récemment étudiée par Smith [SMI 94].

2.4. *Reconnaissance de primitives*

La seule hypothèse effectuée sur la scène, outre les dimensions d'un volume l'englobant, porte sur le fait qu'elle est constituée uniquement de segments et de cylindres. La méthode de reconstruction présentée ci-dessus implique une connaissance *a priori* sur la nature de la primitive observée (segment ou cylindre). Un processus de reconnaissance est donc nécessaire. D'un point de vue géométrique, on peut facilement distinguer laquelle de ces deux primitives est observée en fonction du mouvement projeté dans l'image. En effet, ce mouvement est différent puisque les paramètres de la surface définie par μ (équation (11) dans le cas de la droite, équation (16) dans le cas du cylindre) apparaissent dans la matrice d'interaction associée à ces primitives. Pour obtenir un critère robuste, nous supposons dans un premier temps que le segment

considéré correspond à l'image d'un limbe d'un cylindre. Une reconstruction, basée sur ce segment 2D, est alors effectuée à l'aide d'une suite de mouvements asservis de la caméra. Cette reconstruction fournit N estimations successives des paramètres 3D de la primitive (en particulier de son rayon $R_i, i = 1 \ldots N$). Quand ces estimations sont effectuées, deux hypothèses sont considérées :

$\boldsymbol{H_1}$: la primitive observée était en fait un segment. Idéalement, le rayon estimé doit alors être nul.

$\boldsymbol{H_2}$: la primitive observée était effectivement un cylindre de rayon R ;

Un test au maximum de vraisemblance est alors utilisé pour déterminer laquelle de ces deux hypothèses est correcte. Notons L_0 et L_1 les fonctions de vraisemblance associées aux hypothèses H_1 et H_2. En supposant que les estimations successives R_i du rayon du cylindre suivent une loi Gaussienne de moyenne R et de variance σ^2, nous obtenons :

$$L_0 = (\frac{1}{2\pi\sigma^2})^{\frac{N}{2}} e^{-\frac{\sum_{i=1}^{N} {R_i}^2}{2\sigma^2}} , \qquad L_1 = (\frac{1}{2\pi\sigma^2})^{\frac{N}{2}} e^{-\frac{\sum_{i=1}^{N} (R_i - R)^2}{2\sigma^2}} \qquad [17]$$

Le choix de l'hypothèse H_1 contre l'hypothèse H_0 est basé sur l'optimisation du rapport de vraisemblance ξ donné par $\xi = \log \frac{L_1}{L_0}$. En substituant dans cette équation les expressions données dans (17), on arrive à :

$$\xi = -\frac{1}{2\sigma^2}(\sum_{i=1}^{N} (R_i - R)^2 - \sum_{i=1}^{N} {R_i}^2) \qquad [18]$$

Le critère pour déterminer la nature de la primitive peut alors être défini par :

$$\max_R \xi \geq \lambda$$

où λ est un seuil prédéterminé.

Le paramètre optimal $\widehat{R}$ doit satisfaire la relation $\frac{\partial \xi}{\partial R} = 0$, ce qui amène à $\widehat{R} = R$, soit la moyenne des estimations R_i. En utilisant cette relation, ξ peut finalement s'exprimer de la manière suivante :

$$\xi = \frac{N R^2}{2\sigma^2} \qquad [19]$$

L'hypothèse H_1 (cylindre) est donc retenue contre l'hypothèse H_0 (segment) si la valeur obtenue du rapport de vraisemblance ξ est plus grande que λ. En fait, quand la primitive est un segment, le processus de reconstruction d'un cylindre donne une valeur très faible du rayon, avec un écart-type assez élevé (ce qui implique une valeur faible de ξ). Par contre, quand la primitive est réellement un cylindre, la valeur estimée du rayon est très proche de sa valeur réelle avec une faible variance (ce qui implique une valeur importante de ξ). Cette méthode s'est avérée être plus robuste en pratique que le critère proposé dans [VAI 90].

Si la primitive est reconnue comme étant un cylindre, une estimation plus précise de ses paramètres, basée sur les deux limbes du cylindre, est ensuite réalisée. Avec l'estimation obtenue par la méthode basée sur un seul limbe, il est en effet possible de déterminer la position dans l'image du segment correspondant au deuxième limbe du cylindre.

3. Stratégies de perception

Le problème qui nous intéresse maintenant est celui de la reconstruction complète d'une scène contenant plusieurs objets. En effet, si l'approche retenue pour la reconstruction permet une estimation très précise et robuste des paramètres des primitives concernées, elle ne permet cependant de reconstruire qu'une seule primitive à la fois. L'objectif de cette partie est donc de définir des stratégies de perception capables de fournir une représentation 3D précise et complète de la scène. De manière schématique, l'approche utilisée consiste à sélectionner automatiquement les informations images pertinentes puis à focaliser successivement la caméra sur les différents objets de la scène afin de les reconnaître et ensuite de les reconstruire. Des phases d'exploration sont notamment nécessaires afin d'assurer la complétude de la reconstruction.

3.1. *Positionnement automatique de caméra pour l'observation de scène*

La plupart des travaux antérieurs traitant des problèmes de calcul de points de vue et de l'observation/exploration de scènes supposent connu un modèle complet de la scène [COW 88], [TAR 95b]. Cowan et Kovesi [COW 88] se proposent de sélectionner automatiquement les points de vue permettant d'observer une ou plusieurs faces d'un objet. Ils définissent une série de contraintes sur les paramètres de la caméra qui permettent la réalisation de la tâche visuelle. Le problème traité par Tarabanis, Tsai et Allen dans [TAR 95b] repose sur le même principe mais le système optique est plus général et permet d'obtenir un point de vue généralisé dans un espace à 8 dimensions. Une fonction d'optimisation permet de caractériser la qualité de la solution calculée en se basant sur les mesures liées aux contraintes. Dans le cas d'un environnement partiellement connu ou dynamique, Al Chami et Laugier [Al 94] ont proposé une méthode pour déterminer automatiquement les positions d'une caméra embarquée sur un robot, afin de réaliser un ensemble d'opérations sensorielles qui seront utilisées par la suite pour construire des représentations de l'environnement. Cette méthode consiste à rechercher un positionnement de la caméra qui minimise une fonction d'énergie dont les paramètres représentent les contraintes imposées par la tâche et par l'environnement. L'optimisation de la fonction d'énergie est réalisée par un processus de recuit simulé. Cette méthode est reprise par Triggs et Laugier dans [TRI 95] qui ont proposé d'optimiser la fonction d'énergie à l'aide d'une méthode de recherche probabiliste. Les contraintes prises en compte sont de type optique, géométrique et mécanique (accessibilité).

Le problème est plus complexe si les informations sur la scène sont incomplètes ou nulles (c'est-à-dire si le capteur évolue dans un environnement inconnu). En effet, il faut alors effectuer une tâche d'exploration autonome. De nombreux articles [CON 85], [WIX 94], [MAV 93], [WHA 94] traitent ce problème sous des angles différents. Connoly [CON 85] se propose, en utilisant un capteur laser, de déterminer un modèle complet de la scène à partir d'un ensemble de points de vue. Il décrit l'algorithme dit du *planétarium* qui utilise une décomposition en *oct-tree* de la scène. La caméra se déplace sur la surface d'une sphère régulièrement échantillonnée et circonscrite à la scène. Le point de vue permettant de dévoiler la zone non vue la plus importante est sélectionné. Maver et Bajcsy [MAV 93] utilisent les occlusions pour déterminer les différents points de vue nécessaires pour acquérir les informations 3D des parties cachées de l'environnement. Le capteur est composé d'une caméra et d'un plan laser. Dans [WIX 94], Wixson décrit des stratégies pour rechercher un objet connu dans une zone encombrée. Il explore différentes stratégies pour l'exploration d'un monde 2D à partir d'un capteur 1D. Deux variantes à l'algorithme du planétarium sont proposées : la première basée sur la recherche du point de vue offrant le maximum de visibilité des zones encore inconnues, la deuxième favorisant le point de vue qui minimise le déplacement du capteur. Une autre stratégie étudiée est basée sur l'utilisation des arêtes d'occlusion. L'intérêt principal de cette méthode repose sur l'évaluation des coûts et des bénéfices des opérations de perception effectuées.

Contrairement aux approches précédentes qui utilisent des hypothèses sur la structure de la scène ou qui sont basées sur l'utilisation d'autres types de capteurs qu'une caméra (laser), notre objectif est de reconstruire tous les objets de la scène sans aucune connaissance sur leur nombre, leur position et leur dimension.

3.2. *Reconstruction de scène : principe général*

En sus des hypothèses générales de base sur les dimensions maximales de la scène et la nature des objets (segments et cylindres), les seules données dont l'on dispose sont celles fournies par les images 2D acquises par la caméra. Les stratégies de perception proposées sont principalement basées sur l'utilisation de ces informations 2D. L'une des étapes fondamentales de notre algorithme est donc la création de bases de données répertoriant ces informations. Les bases de données principalement employées, notées ω_{ϕ_t} (où ϕ_t représente la position de la caméra), contiennent la liste des segments correspondant à la projection des objets de la scène visibles dans l'image depuis la position ϕ_t. À chaque segment, on associe une série d'informations indiquant sa position dans l'image et s'il a déjà été traité ou non.

Les algorithmes de traitements d'image que nous utilisons pendant les phases d'asservissement visuel et de reconstruction n'autorisent le suivi en temps réel que d'un petit nombre de segments [BOU 93]. De ce fait, et afin de respecter la contrainte temps réel, ces bases de données ne peuvent être créées à chaque

itération du processus d'estimation, mais seulement à la fin de la reconstruction d'une primitive (*i.e.* quand une précision spécifiée ou un nombre maximal d'itérations ont été atteints).

D'autre part, une seconde base de données, notée Ω_{Φ_t}, est également utilisée. Elle regroupe l'ensemble des bases de données $\omega_{\phi_i}, i = 1$ à t. Plus précisément, Ω_{Φ_t} contient l'ensemble des segments qui n'ont pas encore fait l'objet d'une reconstruction à l'instant t, ainsi que la position de la caméra depuis laquelle ils ont été observés.

Depuis un point de vue ϕ_t et en utilisant les informations 3D déjà recueillies, il est possible, en utilisant un algorithme de lancer de rayons, de calculer la zone observée $V(\phi_t)$. Notons $\mathcal{V}(\Phi)$ la zone de l'espace observé par la caméra depuis le début du processus de reconstruction (*i.e.* les primitives 3D et l'espace libre connus). On a :

$$\mathcal{V}(\Phi_t) = \bigcup_{i=1}^{t} V(\phi_i), \text{ avec } \quad \Phi_t = \bigcup_{i=1}^{t} \phi_t \qquad [20]$$

La reconstruction de la scène s'achèvera quand :

$$\forall \phi_{t+1} \in \mathcal{E}, \mathcal{V}(\Phi_t) \cup V(\phi_{t+1}) = \mathcal{V}(\Phi_t). \qquad [21]$$

où $\mathcal{E}$ définit l'ensemble des points de vue pouvant être atteint par la caméra pendant le processus d'exploration. Cela signifie que l'exploration est aussi complète que possible si, pour tous les points de vue appartenant à $\mathcal{E}$, la caméra observe une zone déjà connue. En utilisant les informations observées (*i.e.* la base de données Ω_{Φ_t} et la carte partielle de l'environnement), nous pouvons définir des stratégies de positionnement de la caméra qui assureront une reconstruction aussi complète que possible de la scène. Ce processus d'exploration, permettant de calculer la position ϕ_{t+1}, est constitué de deux niveaux distincts :

— une **exploration locale** est réalisée si un segment correspondant à une nouvelle primitive apparaît dans le champ de vision de la caméra ou si un tel segment a été précédemment observé depuis une autre position de la caméra. Dans ce cas, un calcul explicite de point de vue n'est pas nécessaire.

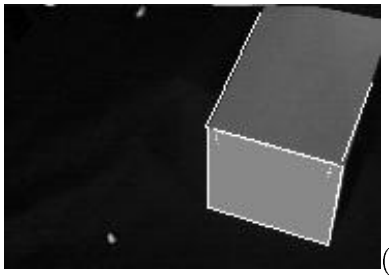
— par contre, dans les autres cas, quand tous les segments précédemment observés ont été reconstruits, une stratégie plus complexe doit être mise en œuvre de manière à focaliser la caméra sur des zones de la scène n'ayant pas encore été observées. Nous parlerons alors d'**exploration globale**.

3.3. *Stratégie locale de focalisation*

Afin de minimiser le déplacement de la caméra, il apparaît souhaitable de mettre en œuvre une stratégie *locale* utilisant de façon explicite la connaissance courante sur la scène [MAR 96b]. Cette stratégie est pilotée par des événements détectés dans l'image et par les informations 3D précédemment acquises.

La démarche s'appuie sur les hypothèses suivantes : la scène est constituée d'objets 3D liés par des relations topologiques ; la projection des objets 3D dans

l'image peut être représentée par un graphe où les nœuds sont les jonctions multiples et les arcs les contours. Chaque arc de ce graphe correspond à un segment 2D de la base de données courante ω_{ϕ_t} (figure 4.b). Ces arcs sont valués en fonction de leur position dans l'image et de la connaissance éventuelle que l'on a sur la primitive 3D correspondante (*i.e.* ils correspondent ou non à la projection dans l'image d'une primitive déjà reconstruite). Si plusieurs arcs de ce graphe se révèlent être la projection de primitives non reconstruites, un choix est effectué pour savoir lequel sélectionner. Ce choix qui permet de minimiser la distance parcourue par la caméra est réalisé en utilisant le graphe et la base de données courante. On recherche un segment non traité connexe à celui qui vient d'être reconstruit. Si un tel segment existe et est unique, il est retenu et reconstruit. Dans le cas ou un tel segment n'existe pas ou n'est pas unique (cas d'une jonction multiple) le segment dont la position dans l'image est la plus proche d'une position optimale pour la reconstruction est alors retenu.


(a)

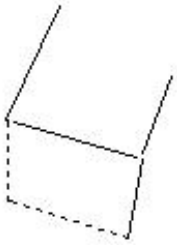

(b)

Figure 4. *(a) Image acquise (b) Base de données 2D (les segments en pointillés ont déjà été reconstruits)*

Dans le cas où tous les segments de ω_{ϕ_t} correspondent à des primitives qui ont déjà été reconstruites, nous recherchons dans la base de données globale Ω_{Φ_t} un segment non traité. La caméra se déplace alors jusqu'à la position depuis laquelle il avait été observé (phase de retour arrière), et une reconstruction de la primitive associée à ce segment est effectuée.

La stratégie *locale* que nous avons développée assure une reconstruction efficace de toute primitive ayant été observée lors du processus de reconstruction. Elle ne fait pas appel à un calcul explicite de nouveaux points de vue et minimise localement le déplacement de la caméra. Cependant, elle ne donne pas l'assurance d'une reconstruction complète de la scène. Pour résoudre ce problème, une stratégie *globale* doit être mise en œuvre.

3.4. *Complétude de la reconstruction : Sélection de points de vue*

Quand toutes les primitives observées pendant les phases d'exploration locale ont été reconstruites (la base de données Ω_{Φ_t} est vide), nous devons déter-

miner les points de vue de la caméra permettant de découvrir éventuellement de nouveaux objets à reconstruire et ainsi d'assurer la complétude de la reconstruction. De tels points de vue sont calculés en utilisant la connaissance courante sur la géométrie spatiale de la scène et une représentation des zones de l'espace déjà observées.

3.4.1. *Calcul de points de vue*

La recherche du point de vue s'effectue en minimisant une fonction d'objectif $\mathcal{F}(\phi)$ qui représente la qualité d'un point de vue ϕ. Nous avons retenu quatre critères qui sont intégrés dans cette fonction :

- le gain en volume découvert apporté par la nouvelle position : l'objectif de l'exploration étant l'acquisition d'informations supplémentaires sur la scène, il est impératif de modéliser ce gain dans la fonction d'objectif.
- le coût de déplacement vers la nouvelle position : un tel critère se justifie par le fait que nous souhaitons avoir une trajectoire de longueur minimale en évitant des mouvements de la caméra d'amplitude importante.
- l'accessibilité : le nouveau point de vue doit, bien évidemment, être accessible à la caméra.
- l'éloignement des butées articulaires du manipulateur : en cas de découverte d'un objet, la caméra pourra alors effectuer les mouvements nécessaires à sa reconstruction.

À chacun de ces critères est associée une mesure à valeur dans $[0,1] \cup \infty$. Une valeur de la mesure proche de 0 signifie que la satisfaction apportée par ce point de vue vis-à-vis du critère associé est maximale. Une valeur proche de 1 signifie au contraire que la position n'a aucun intérêt vis-à-vis du critère associé. Enfin, une valeur infinie signifie que la position est automatiquement à rejeter.

- Le gain apporté par une nouvelle position ϕ_{t+1} est défini par le volume de la zone inobservée qui apparaît dans le cône de vision de la caméra quand celle-ci se déplace de ϕ_t à ϕ_{t+1}. La zone découverte depuis cette position correspond à la zone $\mathcal{G}(\phi_{t+1})$ définie par (voir figure 5.a) :

$$\mathcal{G}(\phi_{t+1}) = \mathcal{V}(\phi_{t+1}) - \mathcal{V}(\phi_{t+1}) \cap \mathcal{V}(\Phi_t) \qquad [22]$$

La mesure du gain associé à la position ϕ_{t+1} est alors donnée par :

$$g(\phi_{t+1}) = 1 - \frac{volume(\mathcal{G}(\phi_{t+1}))}{volume(\mathcal{V}(\phi_{t+1}))} \qquad [23]$$

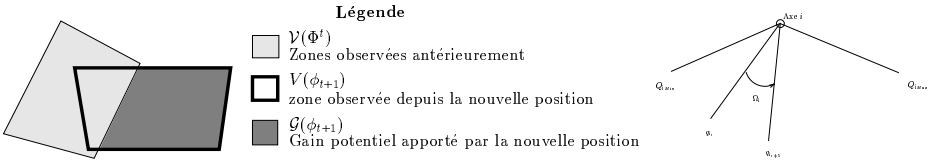
Remarque : La zone $\mathcal{G}(\phi_t)$ correspond en fait à une zone potentiellement découverte. En effet, dans le cas où un nouvel objet apparaît dans le champ de vision, les occlusions dûes à cet objet font que la zone réellement observée $\mathcal{G}'(\phi_t)$ est de taille inférieure ($\mathcal{G}'(\phi_t) \subseteq \mathcal{G}(\phi_t)$).

— La mesure du coût du déplacement entre deux positions ϕ_t et ϕ_{t+1} est définie en calculant la distance entre ϕ_t et ϕ_{t+1}. Plus précisément, ce coût est donné par (voir figure 5.b) :

$$\mathcal{C}(\phi_t, \phi_{t+1}) = \frac{1}{N_{ddl}} \sum_{i=1}^{N_{ddl}} \beta_i \frac{\mid q_{i_t} - q_{i_{t+1}} \mid}{\mid Q_{i_{Max}} - Q_{i_{Min}} \mid} \quad [24]$$

où

– N_{ddl} représente le nombre de degrés de liberté du robot ;
– q_i représente la coordonnée articulaire de l'axe i ($\phi = (q_0, q_1, \ldots, q_{N_{ddl}})$).
– $\mid Q_{i_{Max}} - Q_{i_{Min}} \mid$ est la distance entre les deux butées articulaires de l'axe i.
– les coefficients β_i sont des poids qui permettent éventuellement de favoriser les déplacements sur certains axes du robot (les mouvements de rotation de la caméra peuvent, par exemple, être préférés à des mouvements de translation).



(a) Mesure du gain (projection 2d) (b) Mesure du coût

Figure 5. *Calcul de la fonction de gain et de coût*

De plus, des contraintes supplémentaires sont associées à certaines positions du robot. Ces contraintes tendent à :

— éviter les positions inatteignables pour la caméra en raison des butées articulaires du robot. Une pénalité infinie est donnée à une position si celle-ci n'est pas accessible :

$$\mathcal{A}(\phi) = \begin{cases} 0 & \text{si } \phi \text{ est accessible} \\ \infty & \text{sinon} \end{cases} \quad [25]$$

— éviter des positions trop proches des butées articulaires afin que les mouvements nécessaires à la reconstruction d'une primitive apparaissant depuis ϕ_{t+1} soit possibles :

$$\mathcal{B}(\phi) = \frac{1}{N_{ddl}} \sum_{i=1}^{N_{ddl}} \frac{4(q_i - \frac{Q_{i_{Max}} + Q_{i_{Min}}}{2})^2}{(Q_{i_{Max}} - Q_{i_{Min}})^2} \quad [26]$$

La mesure associée à cette contrainte sera optimale (égale à 0) si les coordonnées articulaires q_i se trouvent au milieu des butées du robots.

La fonction de coût $\mathcal{F}(\phi_{t+1})$ est définie par la somme pondérée de ces différentes mesures :

$$\mathcal{F}(\phi_{t+1}) = \mathcal{A}(\phi) + \alpha_1 g(\phi_{t+1}) + \alpha_2 \mathcal{C}(\phi_t, \phi_{t+1}) + \alpha_3 \mathcal{B}(\phi) \quad [27]$$

La détermination des coefficients α_i dans un problème d'optimisation de ce type est un problème non trivial. Nous nous sommes contentés de choisir ces coefficients de manière empirique. Cependant, leur valeur fixe l'ordre de priorité associé à chacun des critères. L'accessibilité est bien sûr prioritaire (le caractère "binaire" de son résultat rend inutile une quelconque pondération). De plus, la découverte de nouvelles zones à explorer étant notre objectif, nous avons choisi $\alpha_1 > \alpha_2 > \alpha_3$.

3.4.2. *Optimisation*

Chaque position ϕ peut *a priori* être solution de ce problème d'optimisation. Cependant, de manière à contraindre le problème, nous autorisons la caméra à se déplacer sur la surface d'une sphère circonscrite à la scène. La position de la caméra peut alors être décrite par un vecteur à cinq paramètres $(\theta, \varphi, \alpha, \beta, \gamma)$ où θ et φ représentent la latitude et la longitude de la caméra sur la sphère et où α, β et γ représentent l'orientation de la caméra.

Pour minimiser $\mathcal{F}(\phi)$, nous avons choisi d'utiliser une méthode déterministe classique de type gradients conjugués avec un pas de descente à plusieurs niveaux : nous utilisons tout d'abord des incréments importants afin de déterminer la région de l'espace des paramètres où l'optimum de la fonction $\mathcal{F}(\phi)$ est probablement situé. Puis, nous itérons le processus depuis cette nouvelle position avec un incrément plus faible. Contrairement aux méthodes stochastiques de type recuit simulé, nous ne pouvons assurer que la convergence s'effectue vers le minimum global de la fonction. Cependant, le gain en temps de calcul est très important et les expériences ont montré qu'un optimum correct est toujours atteint en un faible nombre d'itérations. De plus, l'intérêt de trouver un minimum global de la fonction d'énergie ne nous a pas paru fondamental dans la mesure où une position apportant un complément d'informations important est trouvée.

4. Résultats expérimentaux

Les expérimentations présentées dans cet article ont été réalisées sur la cellule de vision robotique de l'IRISA composée d'une caméra CCD montée sur l'effecteur d'un robot à 6 degrés de liberté (voir figure 6).

La création des bases de données ω_{ϕ_t}, incluant l'extraction de contours (filtre de Shen-Castan et seuillage) et une approximation polygonale des contours, est réalisée sur une station de travail SUN SPARC Station 20 en 1 seconde environ. Pour les phases de reconstruction 3D par vision active, le traitement d'images est exécuté sur une carte spécialisée (EDIXIA IA 1000). Le traitement consiste à

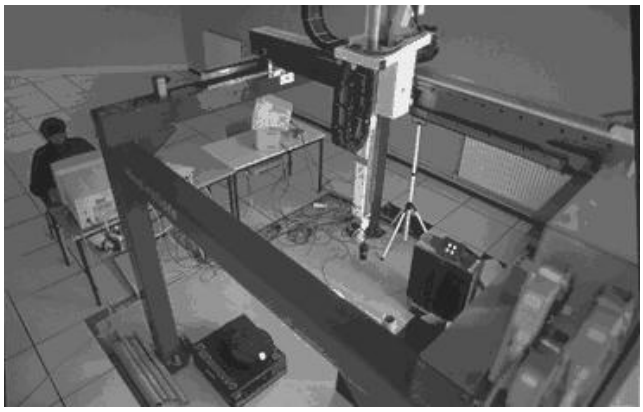


Figure 6. *Cellule robotique de l'IRISA*

suivre sur la séquence d'images acquises le segment 2D sélectionné (ou les deux segments correspondant à la projection d'un cylindre lors de la seconde phase de sa reconstruction) et à en déterminer les paramètres (ρ, θ) décrivant sa position dans l'image. L'extraction et le suivi du segment (en fait une liste de points de contour) est réalisé en 80 ms. La méthode utilisée est décrite dans [BOU 93]. Elle est basée sur une mise en correspondance locale et robuste des éléments de contour en mouvement constituant la droite sélectionnée. L'estimation des paramètres 3D des primitives ainsi que le calcul de la commande sont aussi réalisés sur la station SUN à une cadence de 10 Hz. Le processus de calibration utilisé repose sur la méthode proposée dans [CHA 89]. Finalement, le calcul d'un nouveau point de vue, résultant de la minimisation de la fonctionelle (27), est réalisé en environ 1 seconde avec une précision d'un cm et en environ 10 secondes avec une précision de l'ordre du millimètre.

Un séquencement de tâches d'asservissement visuel est nécessaire pour passer de la reconstruction d'une primitive à une autre, et ainsi obtenir, par construction incrémentale, une représentation complète de la scène. Nous avons donc conçu un séquenceur donnant alternativement à une tâche le contrôle du système de vision. Ce type de séquenceur peut souvent être réduit à un simple automate (ce qui est généralement le cas dans les systèmes réactifs). Pour notre part, nous nous sommes intéressés à la conception d'un automate hiérarchique capable de gérer l'ensemble du processus de reconstruction d'une scène. Chaque tâche est contrainte sur un intervalle de temps (dont les bornes sont définies par la réalisation d'événements particuliers (voir figure 7)). Ce formalisme autorise par exemple la combinaison de comportement, la préemption et le séquencement de tâches. Cet automate sélectionne et gère les actions à effectuer en fonction des événements 2D perçus dans l'image, des connaissances acquises et répertoriées au fur et à mesure de la reconstruction, ainsi que de la détection de la fin d'une tâche d'asservissement visuel.

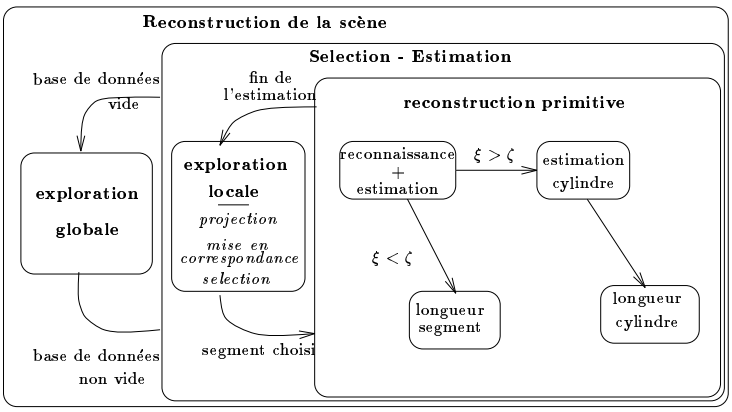


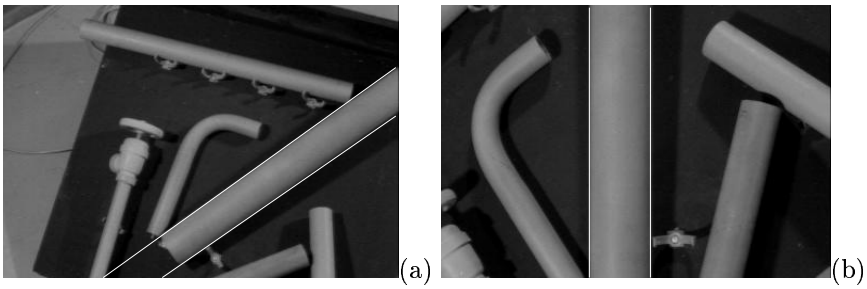
Figure 7. *Automate hiérarchique*

L'ensemble du processus a, dans un premier temps, été mis en œuvre avec le langage C++. Puis, nous avons utilisé le langage synchrone SIGNAL pour la spécification et la mise en œuvre de notre système de vision active. Grâce à ce langage, nous disposons d'une méthodologie de programmation unifiée permettant d'intégrer l'aspect continu des algorithmes de commande et d'estimation avec l'aspect discret des stratégies de perception au sein d'un langage temps réel de haut niveau [MAR 95].

4.1. *Reconstruction de primitive 3D : le cas du cylindre*

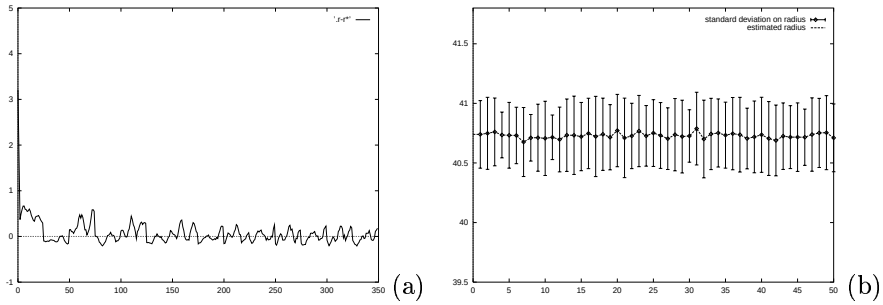
Nous présentons tout d'abord des résultats portant sur la reconstruction d'un cylindre afin de montrer l'intérêt de cette méthode de reconstruction par vision active. Des résultats plus complets sont décrits dans [BOU 93]. De manière à obtenir une estimation robuste et non biaisée de ses paramètres, le cylindre doit apparaître centré et vertical ou horizontal dans l'image (voir figure 8) pendant le mouvement de la caméra qui consiste en une rotation à distance constante de l'axe du cylindre. Les techniques d'asservissement visuel ont donc été employées pour réaliser cette tâche en temps réel : une itération de la boucle de commande et une estimation sont simultanément effectuées en 100 ms. La figure 9.a montre l'erreur entre la valeur estimée du rayon et sa valeur réelle (*i.e.*, $R_i - R^*$) obtenue en utilisant une estimation basée sur les deux limbes du cylindre.

Des tests de stabilité ont également été réalisés. Les paramètres du cylindre ont été estimés 50 fois en partant de positions initiales différentes. Les résultats reportés figure 9.b montrent pour chacune des 50 estimations le rayon $\widehat{R}$ estimé ainsi que l'écart-type $\sigma_{\widehat{R}}$ sur cette estimation. Pour chaque estimation, l'erreur entre la valeur estimée et la valeur réelle est inférieure à 0.1 mm et l'écart-type sur l'ensemble des moyennes est inférieur à 0.02 mm. Ceci démontre que l'algorithme de reconstruction est robuste, stable et précis.



(a) (b)

Figure 8. *Cylindre à reconstruire avant et après la tâche de focalisation*



(a) (b)

Figure 9. *(a) estimation du rayon $R_i - R^*$ (en mm) (b) tests de statibilité*

4.2. *Stratégies de perception*

La scène considérée est composée d'un cylindre et de plusieurs polygones disposés dans des plans différents. L'image 10 montre une vue extérieure de la scène et des différents objets qui la composent. Notre approche ne permet pas actuellement d'éviter d'éventuels obstacles. Aussi, pour éviter toute collision entre le robot et la scène, les objets ont été disposés dans une zone inatteignable par le robot.

4.2.1. *Exploration locale*

La figure 11 représente les images acquises avant chaque reconstruction optimale. À chacune de ces images est associée la base de données 2D correspondante. Les lignes pleines montrent les éléments de la base de données qui n'ont pas encore été traités.

Les lignes en pointillés représentent les segments correspondant à des primitives 3D déjà reconstruites.

La figure 11.a montre l'image acquise depuis la position ϕ_0 de la caméra. Aucune reconstruction n'a encore été effectuée et trois segments seulement sont visibles depuis cette position. On peut noter que la scène complète n'est pas

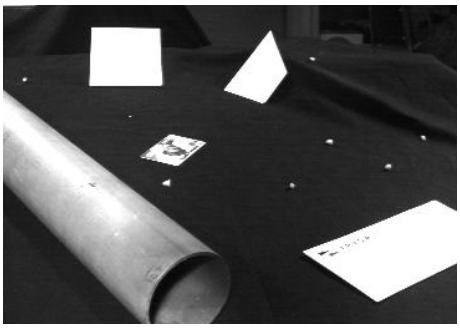
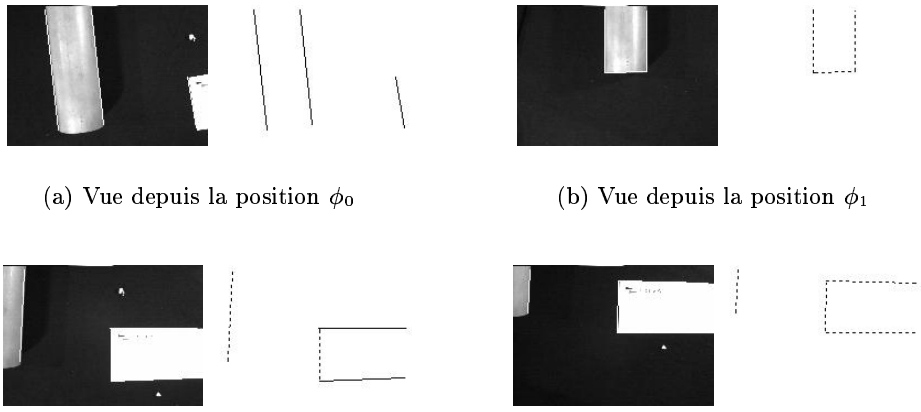


Figure 10. *Vue extérieure de la scène*

visible depuis cette position de la caméra. Le segment extrait de la base de données ω_{ϕ_0} est celui correspondant au limbe de droite du cylindre. Après la phase de reconnaissance et de reconstruction du cylindre, la caméra est positionnée en ϕ_1 (image 11.b). Les segments de la base de données ω_{ϕ_1} ont tous été traités. Après consultation de la base de données globale Ω_Φ, on constate qu'un segment a été observé depuis la position ϕ_0 et n'a pas encore été reconstruit. La caméra se déplace donc en ϕ_0 et se focalise sur le segment retenu. Après l'estimation de cette primitive, la caméra est positionnée en ϕ_2 (image 11.c). Deux segments correspondant à des primitives non reconstruites apparaissent dans la base de données ω_{ϕ_2}. Le segment le plus proche du centre de l'image est sélectionné et reconstruit. Ce processus est renouvelé jusqu'à ce que toutes les primitives observées pendant cette phase d'exploration locale aient été estimées (image 11.d correspondant à la position ϕ_6 de la caméra). Notons que des primitives qui n'apparaissaient pas dans le champ de vision initial de la caméra ont été découvertes et reconstruites. La scène estimée à ce stade du processus de reconstruction est présentée sur la figure 12.

4.2.2. *Exploration globale*

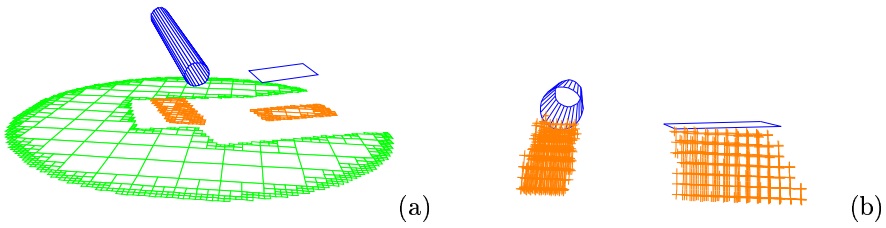
La figure 13 montre les différentes étapes nécessaires à l'exploration complète de la scène. Chacune des figures décrit la scène reconstruite, la trajectoire de la caméra jusqu'à sa position courante, et une visualisation de la zone non observée. La figure 13.a correspond à la position ϕ_6 de la caméra obtenue à l'issue de l'exploration locale décrite dans le paragraphe précédent. Les premiers déplacements de la caméra (voir figure 13.b et 13.c) permettent de diminuer la zone de la scène non encore explorée. À la position ϕ_{13} représentée figure 13.d une nouvelle primitive est détectée marquant le début d'une seconde phase d'exploration locale qui se termine à l'étape 24 (figure 13.f). Les deux polygones en haut de la scène sont alors reconstruits. Une nouvelle exploration globale amène la caméra en ϕ_{25} (figure 13.g) où un segment appartenant au dernière objet (une carte de téléphone) apparaît. Après une dernière exploration locale,



(a) Vue depuis la position ϕ_0 (b) Vue depuis la position ϕ_1

(c) Vue depuis la position ϕ_2 (d) Vue depuis la position ϕ_6

Figure 11. *Exploration locale de la scène*



(a) (b)

Figure 12. *Résultat de l'exploration locale de la scène considérée : (a) scène reconstruite et projection sur un plan virtuel de la zone non observée (b) vue de coté représentant la scène reconstruite et les zones en occlusion*

permettant la reconstruction des 4 cotés de la carte, la caméra est en position ϕ_{30} (figure 13.h). À ce stade, 97% de l'espace a été observé assurant ainsi une reconstruction complète de la scène. La figure 14 montre une visualisation 3D de la scène telle qu'elle a été reconstruite, ainsi que la trajectoire de la caméra pendant l'exploration globale.

Nous analysons à présent l'influence des coefficients α_i de l'équation (27) sur le parcours de la caméra. En particulier, nous mettons en évidence l'importance de la prise en compte de la distance entre deux points de vue successifs dans la fonction d'énergie. La scène est constituée d'un cylindre et d'un polygone qui ont été reconstruits lors d'une phase d'exploration locale. La première stratégie ne prend pas en compte la distance parcourue par la caméra et est donc principalement basée sur une maximisation de la zone découverte pour chaque point de vue (le coefficient α_2 dans l'équation (27) est nul). La deuxième stra-

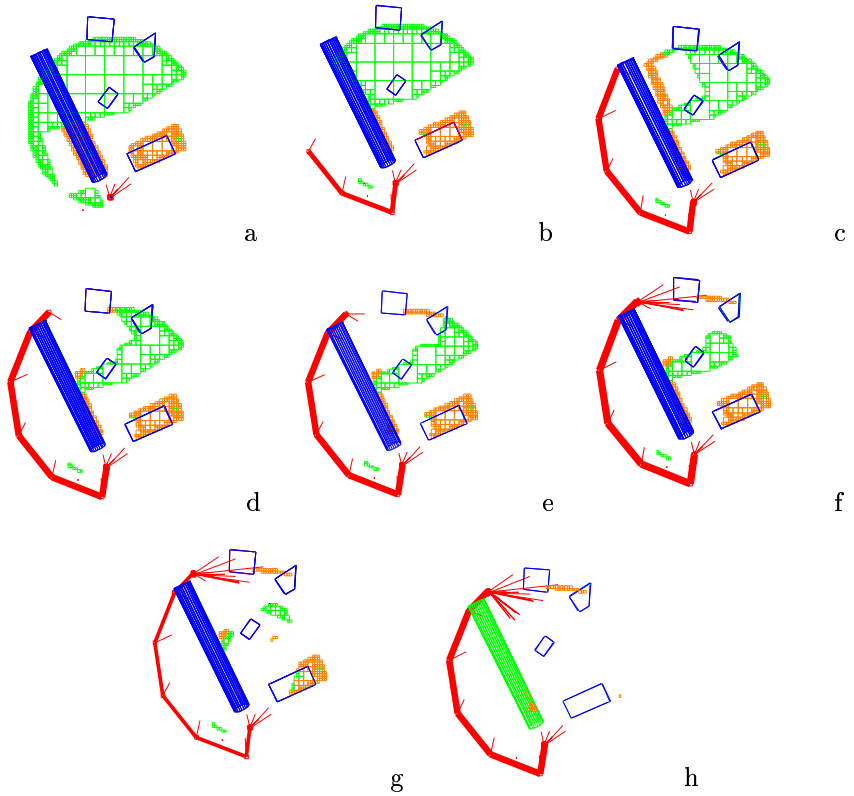


Figure 13. *Différentes étapes de l'exploration globale (trajectoire de la caméra, scène reconstruite, et projection sur le sol de la zone non observée)*

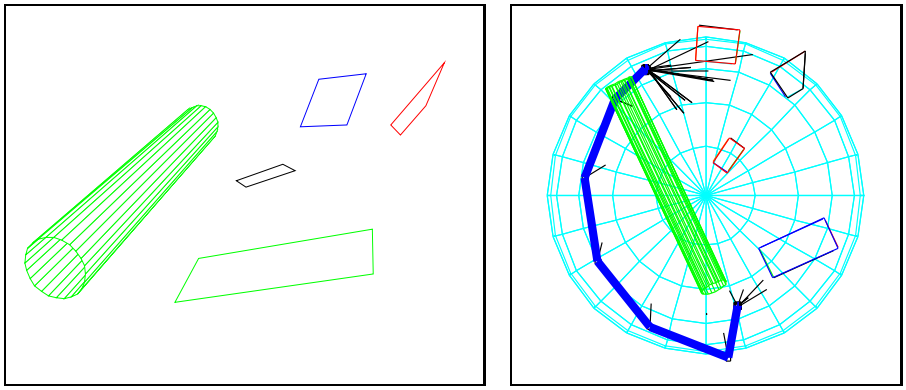
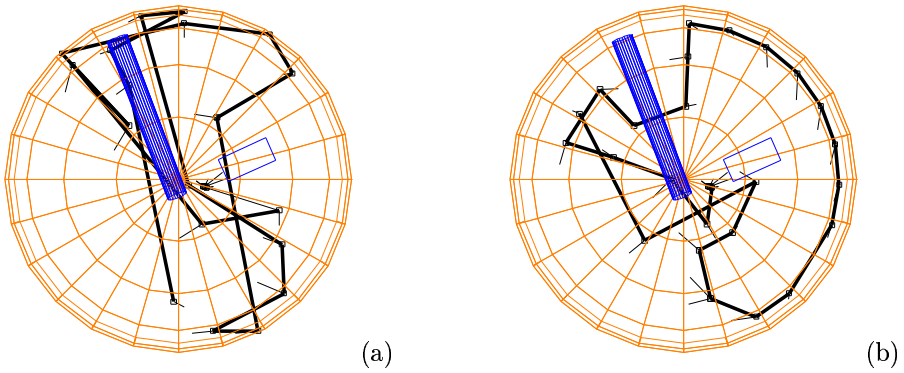
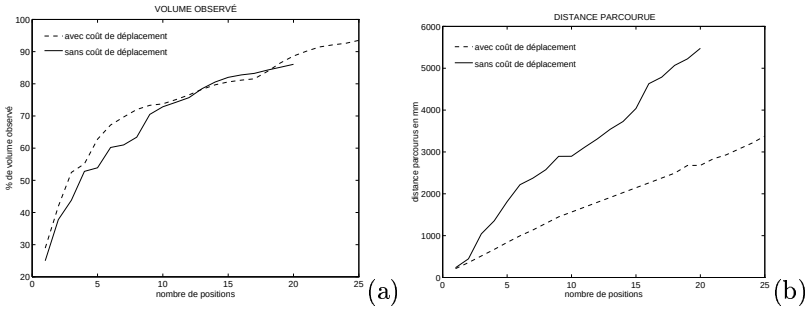


Figure 14. *Visualisation de la scène reconstruite et vue polaire de la trajectoire finale de la caméra pendant l'exploration globale*



(a) (b)

Figure 15. *Exploration globale de la scène - Trajectoire du robot sans tenir compte du coût de déplacement ($\alpha_2 = 0$) (a) et en tenant compte du coût de déplacement ($\alpha_2 > 0$) (b)*



(a) (b)

Figure 16. *Pourcentage de la scène observée (a) et distance parcourue par la caméra (b) en fonction du nombre de points de vue calculés*

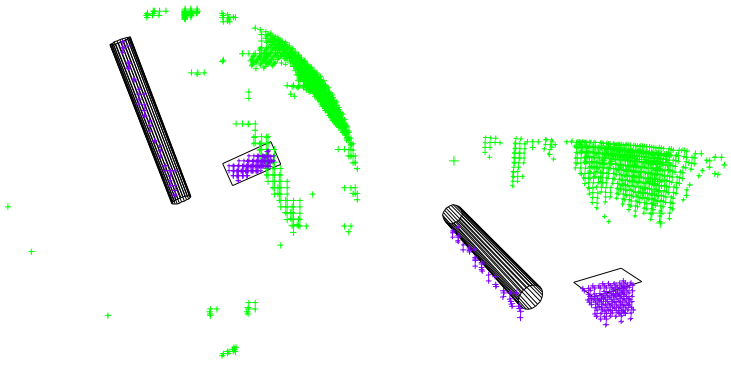


Figure 17. *Zones de la scène restant à observer (vue de dessus et vue de face)*

tégie tient compte de cette distance et tend donc à diminuer la distance totale parcourue par la caméra. Les figures 15.b et 15.c montrent les différentes trajectoires effectuées menant à une exploration totale de la scène. La figure 16.a montre le pourcentage de volume de la scène observée après chaque nouvelle position de la caméra. La figure 16.b montre la distance cumulée parcourue par la caméra pour les deux stratégies. On notera que si la distance parcourue n'est pas prise en compte, la caméra effectue une trajectoire en "vol d'abeille" alors que de tels mouvements ne se produisent plus si la distance est introduite dans la fonction d'énergie. D'autre part, on peut noter que la deuxième stratégie a tendance à favoriser les mouvements suivants les trois axes de rotation par rapport aux mouvements suivant les trois axes de translation. Ceci est dû au fait que les rotations étant moins coûteuses, elles sont moins pénalisées que les translations (les coefficients β_i dans l'équation (26) ont été fixés à 0.4 pour les degrés de libertés en rotation et 0.6 pour ceux en translation).

Notons enfin que dans le cas présent, les zones résiduelles qui n'ont pas été observées (moins de 3% du volume total) correspondent à des zones d'occlusion situées sous le cylindre et sous le rectangle (voir figure 17) ou à des zones situées en périphérie de la scène. Trois perspectives semblent se dessiner pour résoudre ce problème des zones résiduelles non observées :

— en premier lieu, il est possible de ne plus se limiter à des déplacements seulement sur la surface de la sphère, mais de pouvoir pénétrer à l'intérieur de celle-ci. Il convient dans cette optique de vérifier que le point de vue est accessible à la caméra (*i.e.* qu'il appartient à une zone déjà observée qui ne contient pas d'objet).
— on peut ensuite utiliser le même algorithme d'exploration globale que précédemment mais en se restreignant à chaque petite zone circonscrite aux zones résiduelles non observées. Le gain apporté par un nouveau point de vue ne sera alors plus négligeable (voir figure 18).
— par ailleurs, pour résoudre les problèmes d'occlusion, il semble également prometteur de se focaliser sur les segments d'occlusion et d'aller observer les zones occultées par ces segments.

Ces perspectives sont traitées (en partie pour la dernière) dans [MAR 96a].

5. Conclusion

Dans cet article, nous avons proposé une méthode permettant la reconstruction d'un environnement 3D en utilisant une séquence d'images acquises par une caméra mobile. Nous avons décrit un processus de reconstruction permettant une estimation précise et robuste des paramètres d'une primitive géométrique. Cette méthode étant basée sur des mouvements particuliers de la caméra, des stratégies perceptives permettant d'effectuer une succession de reconstructions et assurant la complétude de la reconstruction ont été proposées. Des expérimentations menées sur une cellule robotique ont démontré la validité de notre approche mais aussi ses limitations : les contraintes sur les mouvements de la

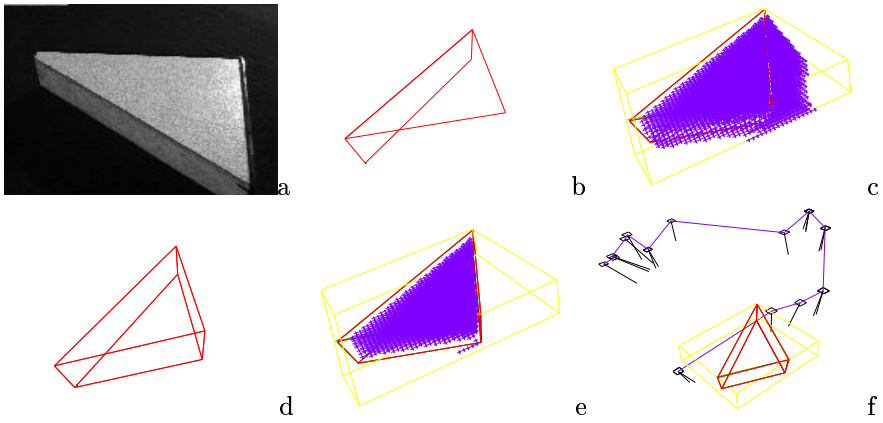


Figure 18. *Reconstruction d'un polyèdre (a) image de la scène, (b) deux polygones ont été reconstruits après une phase d'exploration locale (c) zones de la scène inobservées après cette première phase résultant d'occlusions (d) Modèle de la scène à la fin de la reconstruction (e) zones demeurant inobservées à la fin de l'exploration globale (les zones inobservées sont situées à l'intérieur du polygone) (f) trajectoire de la caméra*

caméra impliquent un séquencement fort des tâches de reconstruction. C'est pourquoi il apparaît utile en perspective de ce travail de déterminer les mouvements optimaux de la caméra nécessaires à la reconstruction simultanée de plusieurs primitives, et ce sans dégradation notable de la qualité de la reconstruction. La reconstruction d'objets plus complexes est également envisagée. Il ne s'agira alors plus de considérer l'objet globalement comme étant une primitive géométrique paramétrable, mais de réaliser une reconstruction locale de la surface de l'objet en utilisant là encore une approche active. Des stratégies de perception devront être mises en œuvre pour assurer une reconstruction complète et tenir compte des problèmes liés aux occlusions ou aux changements de topologie de l'objet.

Remerciements

Cette étude a été effectuée dans le cadre du projet inter-PRC VIA (Vision Intentionnelle et Action) et a reçu le soutien du MESR (Ministère de l'enseignement supérieur et de la recherche).

6. Bibliographie

[ADI 89] 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.

[Al 94] O. Al Chami et C. Laugier. « Stratégie perceptive pour positionner une caméra ». *Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle, RFIA'94*, vol 1, pages 617–622, Paris, France, January 1994.

[ALO 87] Y. Aloimonos, I. Weiss, et A. Bandopadhay. « Active Vision ». *International Journal of Computer Vision*, 1(4):333–356, January 1987.

[ALO 90] Y. Aloimonos. « Purposive and qualitative active vision ». *IAPR Int. Conf. on Pattern Recognition, ICPR'90*, vol 1, pages 346–360, Atlantic City, New Jersey, June 1990.

[BAJ 88] R. Bajcsy. « Active Perception ». *Proc. of the IEEE*, 76(8):996–1005, August 1988.

[BOU 93] S. Boukir. « *Reconstruction 3D d'un environnement statique par vision active* ». Thèse de doctorat, Université de Rennes 1, IRISA, October 1993.

[CHA 89] F. Chaumette et P. Rives. « Modélisation et calibration d'une caméra ». *7ème congrès AFCET Reconnaissance des formes et intelligence artificielle, RFIA'89*, vol 1, pages 527–536, Paris, December 1989.

[CHA 90] F. Chaumette. « *La relation vision-commande: théorie et application à des tâches robotiques* ». Thèse de doctorat, Université de Rennes I, IRISA, July 1990.

[CHA 96] F. Chaumette, S. Boukir, P. Bouthemy, et D. Juvin. « Structure from controlled motion ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(5):492–504, May 1996.

[CHI 89] C. Chien et 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.

[CON 85] C. Connoly. « The Determination of Next Best Views ». *IEEE Int. Conf. on Robotics and Automation*, pages 432–435, St Louis, Missouri, March 1985.

[COW 88] C.K. Cowan et P.D. Kovesi. « Automatic Sensor Placement from Vision task Requirements ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(3):407–416, May 1988.

[ESP 87] B. Espiau et P. Rives. « Closed-Loop Recursive Estimation of 3D Features for a Mobile Vision System ». *IEEE Int. Conf. on Robotics and Automation*, vol 3, pages 1436–1443, Raleigh, Caroline du Nord, April 1987.

[ESP 92] B. Espiau, F. Chaumette, et P. Rives. « A new approach to visual servoing in robotics ». *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.

[FAU 93] O. Faugeras. *Three-Dimensionnal Computer Vision: a Geometric Viewpoint*. MIT Press, Cambridge, Massachusetts, 1993.

[HAS 93] K. Hashimoto, éditeur. *Visual Servoing : Real Time Control of Robot Manipulators Based on Visual Sensory Feedback*. World Scientific Series in Robotics and Automated Systems, Vol 7, World Scientific Press, Singapour, 1993.

[KUT 94] K.N. Kutulakos et C.R. Dyer. « Recovering Shape by Purposive Viewpoint Adjustment ». *International Journal of Computer Vision*, 12(2):113–136, February 1994.

[MAR 95] E. Marchand, F. Chaumette, et E. Rutten. « Real Time Active Visual Reconstruction using the Synchronous Paradigm ». *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'95*, vol 1, pages 96–102, Pittsburgh, Pennsylvanie, August 1995.

[MAR 96a] E. Marchand. « *Stratégies de perception par vision active pour la reconstruction et l'exploration de scènes statiques* ». Thèse de doctorat, Université de Rennes 1, IRISA, No 1589, June 1996.

[MAR 96b] E. MARCHAND et F. CHAUMETTE. « Controlled Camera Motions for Scene Reconstruction and Exploration ». *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'96*, pages 169–176, San Francisco, Californie, June 1996.

[MAV 93] J. MAVER et 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.

[SMI 94] C.E. SMITH et N.P. PAPANIKOLOPOULOS. « Computation of Shape Through Controlled Active Exploration ». *IEEE Int. Conf. on Robotics and Automation*, vol 3, pages 2516–2521, San Diego, Californie, May 1994.

[SWA 93] M.J. SWAIN et M.A. STRICKER. « Promising Direction in Active Vision ». *International Journal of Computer Vision*, 11(2):109–127, October 1993.

[TAR 95a] K. TARABANIS, P.K. ALLEN, et R. TSAI. « A Survey of Sensor Planning in Computer Vision ». *IEEE Trans. on Robotics and Automation*, 11(1):86–104, February 1995.

[TAR 95b] K. TARABANIS, R. TSAI, et P.K. ALLEN. « The MVP Sensor Planning System for Robotic Vision Tasks ». *IEEE Trans. on Robotics and Automation*, 11(1):72–85, February 1995.

[TRI 95] B. TRIGGS et C. LAUGIER. « Automatic camera placement for robot vision ». *IEEE Int. Conf. on Robotics and Automation*, vol 2, pages 1732–1738, Nagoya, Japon, May 1995.

[VAI 90] R. VAILLANT. « Using Occluding Contours for 3D Objects Modeling ». *1st European Conf. on Computer Vision, ECCV'90*, pages 454–464, Antibes, France, April 1990.

[WAX 87] A.M. WAXMAN, B.K. PARSI, et M. SUBBARAO. « Closed-form Solutions to Image Flow Equations for 3D Structure and Motion ». *International Journal of Computer Vision*, 1(3):239–258, October 1987.

[WEN 90] J. WENG, T.S. HUANG, et N. AHUJA. « Estimation and Structure from Line Matches: Performance Obtained and Beyond ». *IAPR Int. Conf. on Pattern Recognition, ICPR'90*, vol 1, pages 168–172, Atlantic City, New Jersey, June 1990.

[WHA 94] P. WHAITE et F. FERRIE. « Autonomous exploration: Driven by uncertainty ». *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pages 339–346, Seattle, Washington, June 1994.

[WIX 94] L.E. WIXSON. « Viewpoint Selection for Visual Search ». *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pages 800–805, Seattle, Washington, June 1994.

[XIE 89] M. XIE et P. RIVES. « Toward dynamic vision ». *IEEE Workshop on Interpretation of 3D Scenes*, pages 91–99, Austin, Texas, November 1989.

[ZHA 95] Z. ZHANG. « Estimating Motion and Structure from Correspondences of Line Segments between Two Perspective Images ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(12):1129–1139, December 1995.

Article reçu le 2 octobre 1995.
Version révisée le 11 avril 1996.
Rédacteur responsable : Marc Richetin.

Éric Marchand *est né en janvier 1970 à Rennes. Après une maîtrise et un DEA en informatique obtenu en 1993 à l'Ifsic (Université de Rennes 1), il a soutenu en juin 1996 une thèse de doctorat en informatique à l'Université de Rennes 1. Cette thèse a été effectuée à l'Irisa (Institut de Recherche en Informatique et Systèmes Aléatoires) au sein du projet Temis. Il effectue actuellement un séjour Post-Doctoral dans le laboratoire de vision artificielle et robotique de l'Université de Yale. Ses travaux de recherche concernent les stratégies de perception par vision active et principalement la coopérations entre la perception et l'action notamment dans le domaine de la reconstruction 3D et de la coopérations multi-capteurs. Il s'intéresse d'autre part aux aspects de mise en œuvre des systèmes de vision robotique au travers des langages synchrones.*

François Chaumette *est né à Nantes en décembre 1963. Après une maîtrise de mathématiques appliquées, obtenue en 1985, et un diplome d'ingénieur en automatique de l'Ecole Nationale Supérieure de Mécanique de Nantes, obtenu en 1987, il a effectué un doctorat en informatique à l'Université de Rennes 1 dans le laboratoire de l'Irisa. Sa thèse, soutenue en 1990 et portant sur les techniques d'asservissement visuel en robotique lui a valu de recevoir le prix annuel AFCET / CNRS de la meilleure thèse en automatique. Chargé de recherche Inria à l'unité Irisa/Inria Rennes depuis la fin de son doctorat, ses travaux de recherche concernent principalement le couplage entre la robotique et la vision par ordinateur, notamment dans les domaines de l'asservissement visuel, de la vision active et dans l'élaboration de stratégies automatiques de perception/action. En collaboration avec la société rennaise Edixia, il participe également au transfert technologique des techniques d'asservissement visuel, notamment dans le domaine de la robotique en milieu nucléaire.*