

# Stratégies de reconstruction par vision active : une approche par réseaux Bayesiens

## Active reconstruction strategies : a Bayes nets-based approach

Éric Marchand

François Chaumette

IRISA / INRIA Rennes

Campus de Beaulieu

F-35042 Rennes Cedex

*Email* : {marchand,chaumett}@irisa.fr

### Résumé

*Cet article traite du problème de la reconstruction et de l'exploration de scènes polyédriques dans un contexte de vision active. Les stratégies de perception que nous avons élaborées permettent de reconstruire de manière incrémentale l'ensemble des primitives qui constituent la scène. La phase d'exploration que nous décrivons dans cet article repose sur une approche de prédiction/vérification d'hypothèses gérées à l'aide de réseaux bayesiens. Cette approche fournit une représentation de plus haut niveau des objets considérés, tout en traitant les problèmes locaux d'occultation. Les méthodes que nous avons développées ont été mises en œuvre sur la cellule de vision robotique de l'IRISA. Elles permettent de reconstruire de façon précise, robuste et totalement autonome, un environnement 3D composé de plusieurs objets.*

### Mots Clef

Vision active, reconstruction 3D, réseaux Bayesiens, prédiction/vérification d'hypothèses.

### Abstract

*We propose in this paper an active vision approach for performing the 3-D reconstruction of polyhedral scenes. The perceptual visual strategies we have defined allow the system to incrementally build a complete 3-D map of the scene. The exploration algorithm we describe in this paper is based on a prediction/verification scheme managed using decision theory and Bayes nets. It provides a more complete high level description of the scene and partially solves occlusions problems. Experiments carried out on a robotic cell demonstrate the validity of our approach.*

### Keywords :

Active vision, 3D reconstruction, Bayes nets, prediction/verification

## 1 Introduction

De nombreux travaux menés en vision artificielle se sont fixés pour objectif la réalisation de systèmes puissants capables d'accéder à la géométrie spatiale d'une

scène à partir de son observation par une caméra mobile. Ces systèmes doivent fournir une description géométrique claire et complète de la scène à partir d'une séquence d'images souvent bruitées et difficilement exploitables. L'étude proposée ici tente d'apporter sa contribution au problème de la reconstruction d'environnements tridimensionnels assez restreints, avec comme ambition la reconstruction de scènes de plus en plus complexes.

Dans une optique d'analyse de scènes, les approches de vision inspirées du paradigme de Marr considèrent un capteur statique ou mobile, mais non contrôlé. Cette approche s'avère cependant insuffisante pour résoudre un grand nombre de problèmes où une modification pertinente des paramètres intrinsèques et/ou extrinsèques du capteur est nécessaire. C'est pourquoi Aloimonos, Bajcsy ou encore Ballard ont proposé de modifier cet état de fait en élaborant le concept de vision active (voir [14] pour une tentative de classification des différents courants en vision active selon les motivations des auteurs). D'un point de vue méthodologique, la vision active, où les informations perçues sont utilisées au sein d'une boucle de rétroaction, tente surtout d'améliorer la qualité de la perception par rapport à l'approche passive classique, où l'on se restreint à observer, mesurer et interpréter les données issues du capteur. La vision active peut être définie comme un processus d'acquisition "intelligent" des données afin de résoudre les problèmes soulevés lors de la conception d'un système de vision par ordinateur, à savoir principalement leur sensibilité au bruit, leur faible précision, et surtout leur manque de réactivité.

Dans le cadre de ce travail, nous utilisons le principe de la vision active pour la reconstruction 3D et l'exploration de scènes à l'aide d'un système de vision robotique calibré. Plus précisément, notre objectif est d'obtenir une reconstruction complète et aussi précise que possible de scènes statiques composées d'objets polyédriques, sans connaissances *a priori* sur le nombre, la dimension ou la position spatiale de ces objets. Pour cela, nous nous plaçons à deux niveaux différents : à un niveau local, les mouvements de la caméra sont contraints de manière à optimiser la qualité de la re-

adéquats de la caméra permettent d'assurer l'exploration complète de la scène. Sur le premier point, nous avons utilisé une approche continue dont l'originalité consiste à générer automatiquement les mouvements de la caméra dans le but d'obtenir une reconstruction 3D précise et robuste de primitives géométriques paramétrables [6]. Cette méthode repose sur la mesure du mouvement de la caméra et sur l'observation du mouvement de la primitive dans la séquence d'images. Elle a été utilisée pour la reconstruction des primitives les plus classiques (*i.e.*, point, droite, cercle, sphère et cylindre) [6]. Quand les mouvements de la caméra sont quelconques, d'importantes erreurs sur l'estimation des paramètres 3D de la primitive apparaissent. La qualité de l'estimation est en effet très sensible à la nature des mouvements de la caméra. Une méthode reposant sur le concept de la vision active a donc été développée afin de minimiser les effets des erreurs de mesure et les erreurs de discrétisation. Il est montré dans [6] que deux tâches visuelles (appelées tâche de fixation et tâche de focalisation) permettent d'obtenir une estimation robuste et non biaisée. Les mouvements effectués pour réaliser ces deux tâches sont automatiquement générés par asservissement visuel [9]. Concernant le cas du segment, il doit apparaître vertical (ou horizontal) et centré dans la séquence d'images pendant le mouvement de la caméra (voir Figure 1). L'utilisation de la vision active à ce niveau est cependant assez contraignante puisque les mouvements de la caméra sont spécifiques à chaque primitive de la scène. La reconstruction d'une scène composée de plusieurs objets polyédriques nécessite donc des focalisations successives sur les différents segments la constituant.

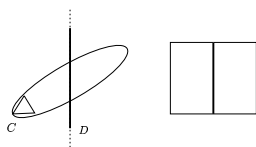


FIG. 1 – *Mouvements de la caméra pour la reconstruction optimale d'une droite et image correspondante*

Pour cela, nous nous plaçons à un second niveau, plus global, en développant des **stratégies de perception** qui permettent l'acquisition d'une carte complète de la scène. Nous proposons donc une approche originale au problème de l'exploration/reconstruction 3D. Les mouvements de la caméra sont explicitement commandés de manière à observer l'ensemble de la scène tout en étant focalisé sur des zones particulières. Cet aspect "stratégies de perception" est divisé en deux étapes distinctes : la première, décrite dans cet article, se propose de reconstruire de manière incrémentale l'ensemble des objets qui apparaissent dans le champ de vision de la caméra. La seconde, déjà décrite dans [14, 15], consiste à focaliser la caméra sur les zones de la scène n'ayant pas encore été traitées, aboutissant ainsi à la complétude de la reconstruction.

efficace, qui permet de façon incrémentale l'ensemble des segments observés par la caméra. La modélisation obtenue (en l'occurrence une liste de segments 3D) reste cependant de bas niveau et est parfois incomplète. Nous proposons donc une approche permettant d'obtenir une représentation de la scène de haut niveau tout en traitant les problèmes locaux d'occultations. La méthode que nous avons développée pour y parvenir vient se greffer sur l'algorithme de reconstruction incrémentale et repose sur des techniques de **prédiction/vérification d'hypothèses**. Du fait des incertitudes dans les mesures et les observations, nous avons utilisé une approche probabiliste reposant sur l'utilisation de réseaux Bayésiens.

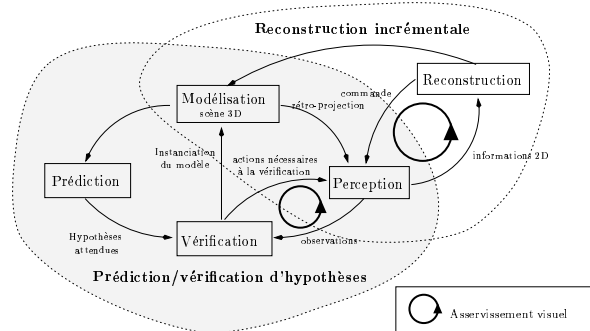


FIG. 2 – *Principe général*

De manière plus précise, la reconstruction de la scène peut se diviser en deux parties fortement interdépendantes. La stratégie retenue est représentée sur la Figure 2. La première partie gère la reconstruction incrémentale elle-même et contient, outre un cycle interne de perception-action (asservissement visuel) assurant la reconstruction de chaque segment, un second cycle permettant la détection, la focalisation puis la reconstruction de l'ensemble des segments observés. La deuxième partie gère le processus de prédiction/vérification d'hypothèses et est composée des trois étapes successives suivantes :

- **Prédiction et génération d'hypothèses.** Plusieurs hypothèses sont émises en parallèle ; à chacune d'entre elles est associée la confiance qui lui est accordée en fonction des connaissances relatives à la position 3D des segments considérés et des connaissances introduites dans le réseau Bayésien gérant la génération d'hypothèses.
- **Vérification des hypothèses émises.** Chacune des hypothèses est ensuite vérifiée ou infirmée en utilisant des informations 2D. Si cette information est présente dans la séquence d'images déjà acquises, la décision peut se prendre sans acquisition d'observations supplémentaires. Dans le cas contraire, il est nécessaire d'acquérir explicitement cette information par des mouvements adéquats de la caméra, ceux-ci étant ici encore générés par asservissement visuel.

fin, la modélisation locale de la scène est intégrée dans une modélisation de plus haut niveau. La vérification d'une hypothèse permet en effet d'inférer l'existence d'objets plus complexes (chaîne de segments coplanaires, polygones, faces,...) qu'une simple liste de segments 3D.

## 2 Reconstruction incrémentale de la scène

Nous présentons ici succinctement l'algorithme permettant de reconstruire incrémentalement l'ensemble des primitives observées par la caméra (voir [14, 15] pour plus de détails). Cet algorithme est composé de trois étapes principales : le choix du prochain segment à reconstruire, la reconstruction d'une primitive 3D, et la mise à jour des informations 2D disponibles. Ce processus est itéré jusqu'à ce que toutes les primitives observées aient été reconstruites.

Comme nous l'avons déjà dit, la scène est supposée composée uniquement d'objets polyédriques. De ce fait, les contours de ces objets se projettent dans le plan image sous la forme de segments. La première étape du processus de reconstruction incrémentale consiste donc à extraire ces segments de l'image. Seuls les segments ayant une taille minimale et correspondant à la projection de primitives non reconstruites sont retenus. Parmi les segments 2D visibles ou ayant été observés depuis les positions précédentes de la caméra, seul un sous-ensemble  $\Omega$  correspond à la projection de segments 3D non reconstruits. Il convient donc d'extraire de ce sous-ensemble un des segments et à procéder à la reconstruction de la primitive associée. Cette phase de reconstruction inclut le processus d'estimation active décrit dans [6] et rappelé brièvement en introduction, ainsi que l'estimation de la longueur du segment. Les caractéristiques 3D de la primitive sont ensuite introduites dans la carte de la scène. C'est alors qu'intervient le processus de prédiction/vérification d'hypothèses. Finalement, il est nécessaire de remettre à jour l'ensemble  $\Omega$  et de choisir une nouvelle primitive à reconstruire. Ces différentes étapes (estimation, création des ensembles de segments, et sélection) sont itérées tant qu'il existe un segment non reconstruit dans l'ensemble  $\Omega$ , ce qui permet de traiter l'ensemble des primitives observées par la caméra.

L'étape de prédiction/vérification d'hypothèses est l'un des points clés de notre méthode. En effet, en ne la considérant pas, le processus de reconstruction incrémentale présente les deux forts inconvénients suivants. Tou d'abord, la description de la scène dont on dispose est une description de bas niveau constituée d'une liste de segments 3D. Il paraît cependant indispensable de disposer d'une description de plus haut niveau en terme d'objets (jonctions, polygones, faces, etc.). De plus, la reconstruction de la scène n'est pas complète pour les deux raisons suivantes :

- La projection dans l'image de certaines primitives a une taille trop petite pour être retenue dans les

ne sont pas reconstruites.

- Ne traitant que les segments observés, cet algorithme se contente d'une vision locale de la scène. Certains objets peuvent donc ne pas apparaître dans le champ de vision de la caméra, soit en raison d'occultations, soit parce qu'ils sont situés dans des zones totalement inconnues. Ces segments non vus ne peuvent évidemment pas faire l'objet d'une reconstruction.

A titre d'exemple, considérons le cas de la scène "polyèdre" présentée sur la Figure 3. Le modèle obtenu en utilisant une simple reconstruction incrémentale est donné sur la Figure 3.c. Il présente l'ensemble des inconvénients décrits ci-dessus : le modèle 3D est constitué de cinq segments *a priori* disjoints, plusieurs segments n'ont pas été pris en compte en raison de leur faible longueur, et un segment de taille importante n'a pas été reconstruit (car occulté en permanence). La méthode que nous décrivons à présent permet de compléter grandement ce modèle, comme le montrent les résultats donnés sur la Figure 3.d

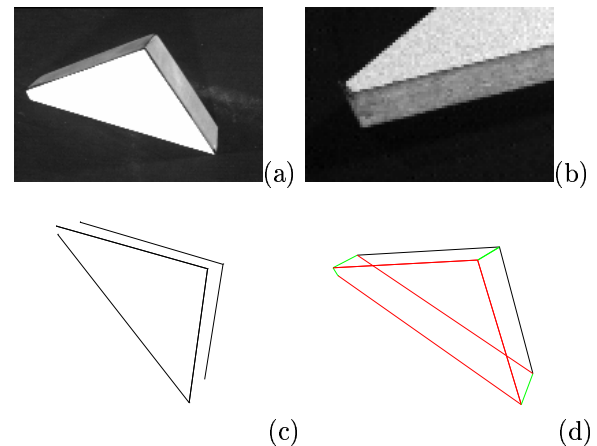


FIG. 3 – Scène "polyèdre" : (a) Vue de la scène à reconstruire, (b) présence d'un "petit" segment, (c) modèle de la scène acquis en se servant uniquement du module de reconstruction incrémentale, (d) modèle reconstruit en se servant du module de prédiction/vérification d'hypothèses

## 3 Génération / vérification d'hypothèses

Les techniques de prédiction/vérification d'hypothèses visent généralement à résoudre le problème de l'adéquation des données aux modèles. Elle ont principalement été utilisées pour la reconnaissance d'objets 2D [2] ou 3D [4], mais aussi pour de nombreuses autres applications telles, par exemple, la mise en correspondance de segments dans un couple d'images stéréoscopiques [3], ou l'acquisition de modèles 3D à partir d'un ensemble d'images aériennes [11]. Ces techniques consistent à émettre des hypothèses sur la présence

suite que des appareillages sont compatibles avec ces hypothèses.

Dans notre cas, l'objectif n'est pas de reconnaître l'objet considéré mais d'obtenir une modélisation la plus complète possible d'un objet à partir de données tridimensionnelles incomplètes et d'informations visuelles disponibles ou à acquérir. La manipulation de l'incertitude inhérente à la manipulation de données réelles ainsi que les stratégies de perception nécessaires à l'accomplissement de cette tâche sont gérées en utilisant des techniques issues de la théorie de la décision. La théorie de la décision a pour objectif l'étude des techniques mathématiques utilisées pour prendre une décision en présence de plusieurs alternatives. Différents cadres mathématiques sont envisageables pour traiter ce problème. Citons par exemple la théorie de Dempster Shafer (utilisée en vision par Hutchinson et Kak [12] pour la reconnaissance d'objets), la théorie des possibilités (utilisée, par exemple, par Giraudon et Sandakly dans le système MESSIE pour l'interprétation de scènes 3D [10]), les chaînes de Markov cachées (utilisées par Rimey et Brown [18] pour simuler le mouvement des yeux) ou les réseaux Bayésiens [17]. Nous avons choisi d'utiliser le cadre Bayésien car il nous paraît bien adapté à notre problème. En effet, l'utilisation des réseaux Bayésiens permet de modéliser à travers la structure du réseau les connaissances d'un "expert". De plus, ils se prêtent bien à l'exécution d'actions au cours des raisonnements, ce qui permet d'introduire les stratégies de perception au sein même du processus d'interprétation de la scène. Les réseaux Bayésiens ont déjà été de nombreuses fois utilisés en analyse mono-image : pour établir des groupements perceptuels [20], la segmentation en niveaux de gris [13], la reconnaissance d'objets [1], etc. L'utilisation de ces réseaux en vision active est plus récente. Les travaux les plus discriminants ont été proposés par Rimey et Brown [19] avec le système TEA-1, Buxton et Gong [5], ou encore Djian et Rives [8]. Les finalités de ces systèmes sont différentes, ils ont cependant un point commun : la réalisation de la tâche nominale nécessite l'exécution d'actions de perception qui sont générées en utilisant les connaissances présentes dans le ou les réseaux Bayésiens utilisés.

### 3.1 Réseaux Bayésiens

Les réseaux Bayésiens permettent de représenter les distributions de probabilités jointes d'un ensemble de variables en utilisant un ensemble de connaissances *a priori* sur les relations entre ces variables. Un réseau Bayésien est un graphe orienté acyclique où les nœuds représentent une proposition (ou une variable) associée, dans le cas des variables aléatoires discrètes, à un ensemble discret de valeurs. Les liens entre les nœuds (arcs orientés) indiquent les relations de dépendance (causalité) entre variables. Une quantification de cette dépendance est donnée par des tables définissant la probabilité conditionnelle qu'un nœud ait telle valeur sachant que son prédécesseur a telle autre va-

une connaissance disponible sur un domaine particulier. La structure du graphe ainsi que les connaissances *a priori* qu'il contient (sous la forme de table de probabilité conditionnelle) doivent être introduites par le concepteur de l'application.

Les réseaux Bayésiens ont l'avantage de permettre la représentation des connaissances que l'on a *a priori* de l'application. Cette connaissance est reflétée à deux niveaux différents :

- dans la structure même du réseau à travers la nature et le nombre des nœuds (variables) qui le composent, les différents états que peuvent prendre ces variables et enfin les relations d'indépendance entre les variables (caractérisées par les arcs qui relient les différents nœuds entre eux).
- dans les tables de probabilités conditionnelles associées aux différentes variables du réseau et qui reflètent les choix de l'expert. Ajoutons que ces tables permettent aussi de modéliser l'incertitude associée aux observations.

Enfin, les mécanismes de propagation associés aux réseaux Bayésiens permettent de prendre en compte toute nouvelle observation. Son influence est propagée aux autres variables du réseau compte tenu des relations de dépendances fixées. Décrire ce processus complexe n'est pas l'objectif de cet article (voir [16]).

### 3.2 Approche générale retenue

L'objectif de notre approche est de déterminer les relations existant entre les objets déjà reconstruits et les nouveaux objets introduits dans le modèle de la scène, ainsi que d'inférer la présence de nouveaux segments et/ou de nouveaux objets. La reconstruction se faisant de façon incrémentale, ce module doit déterminer les conséquences induites par l'intégration de chaque nouveau segment  $S_t$ . Plus précisément, pour chaque couple de segments  $(S_{t'}, S_t)$ ,  $t' \in [0, t - 1]$ , le système détermine les informations 3D supplémentaires que ce couple de segments peut apporter. Puis, il vérifie si les hypothèses émises sont conformes à la réalité. Finalement, le système propose un nouveau modèle de la scène résultant de l'intégration de ce nouveau segment.

### 3.3 Prédiction : génération des hypothèses

Considérant deux segments 3D  $S_{t'}$  et  $S_t$ , il est possible de réaliser les actions suivantes : fusionner les segments, créer une jonction, ou ajouter un ou deux segments entre leurs extrémités. L'objectif de cette première étape est donc d'émettre une ou plusieurs hypothèses menant à la réalisation d'une ou plusieurs de ces actions, à savoir :

- $H_1$  : il existe une jonction entre  $S_{t'}$  et  $S_t$  ;
- $H_2$  : il existe un ou deux segments entre  $S_{t'}$  et  $S_t$ , ce qui implique qu'il existe une chaîne  $(S_{t'}, S_k, S_t)$  ou  $(S_t, S_j, S_{t'})$  ou  $(S_{t'}, S_k, S_t, S_j)$  (voir Figure 4) ;

FIG. 4 – Hypothèse  $H_2$  : ajout de segments

- $H_3$  :  $S_{t'}$  et  $S_t$  sont le même segment ;
- $H_4$  : il n'existe pas ou il existe d'autres relations entre  $S_{t'}$  et  $S_t$ .

Deux séries d'hypothèses seront émises, l'une concernant les relations entre les deux extrémités les plus proches des segments, et l'autre concernant les relations entre les deux extrémités les plus éloignées. Il faut ensuite associer à chacune de ces hypothèses la confiance que nous avons en elle, en fonction des observations et des connaissances déjà acquises sur la scène. Pour cela, nous recherchons des relations topologiques simples (voisinage, coplanarité, colinéarité) entre les deux segments  $S_{t'}$  et  $S_t$ , et en déduisons la classe à laquelle peuvent appartenir ces segments. Enfin, à partir de la classe retenue, une hypothèse est émise sur la nature de la relation associée aux segments. Ce schéma de raisonnement est codé dans la structure du réseau Bayésien de la Figure 5.

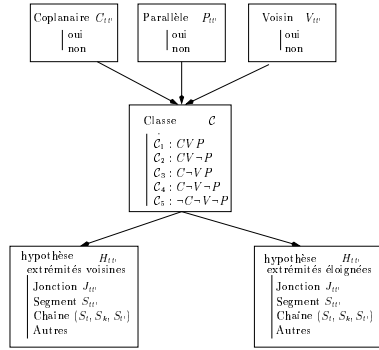


FIG. 5 – Réseaux de génération d'hypothèses

Plus précisément, compte tenu de la confiance que nous accordons aux trois relations (coplanarité  $p(C_{tt'})$ , voisinage  $p(V_{tt'})$  et parallélisme  $p(P_{tt'})$ ), il est possible de classer un couple de segments en cinq classes (voir la première colonne du tableau de la Figure 6) :

- $C_1$  : les segments sont coplanaires, voisins et parallèles ;
- $C_2$  : les segments sont coplanaires, voisins et non parallèles ;
- $C_3$  : les segments sont coplanaires, non voisins et parallèles ;
- $C_4$  : les segments sont coplanaires, non voisins et non parallèles ;
- $C_5$  : les segments ne sont ni coplanaires, ni voisins, ni parallèles.

Il convient ensuite de déterminer l'hypothèse à adopter sur la nature des relations entre les deux segments compte tenu des connaissances disponibles. Ces stratégies de décision sont codées dans les tables de probabilités conditionnelles  $P(\mathcal{H}|C)$  où  $\mathcal{H}$  est l'hypothèse et  $C$  la classe.

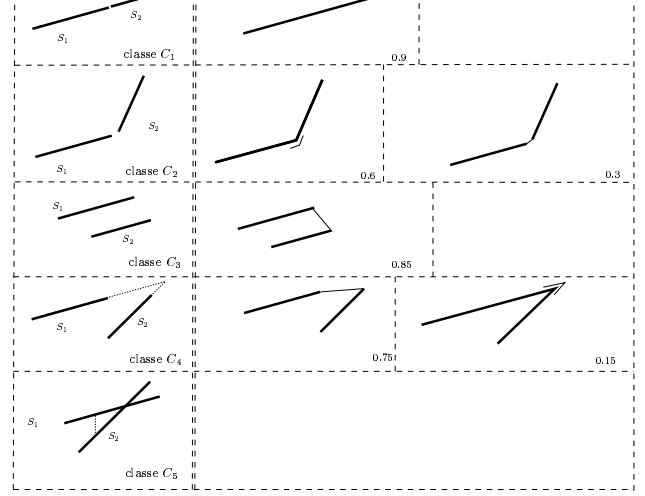


FIG. 6 – Classes élémentaires et hypothèses associées : extrémités proches

Afin d'émettre des hypothèses censées, il apparaît nécessaire de proposer quelques considérations élémentaires sur les relations topologiques que l'on retrouve généralement dans un groupe de segments. Ces considérations partent du constat que, dans la plupart des cas, il y a une certaine organisation qui régit la modélisation des objets polyédriques. Ces considérations sont les suivantes :

- il est rare qu'un segment se retrouve isolé, car il appartient généralement à un objet plus complexe ;
- il est courant que deux segments voisins forment une jonction et appartiennent, par conséquent, au même objet ;
- de même, il est courant de voir un ou deux segments reliant les extrémités de deux segments coplanaires, formant ainsi un polygone ;
- par contre, il est rare de voir deux segments n'ayant aucune relation de voisinage, coplanarité ou de parallélisme entre eux appartenir au même objet.

Ces considérations reflètent, dans la plupart des cas, la réalité et peuvent donc servir de base pour la définition des hypothèses que nous souhaitons pouvoir émettre. Pour illustrer ce point, prenons l'exemple de deux segments voisins, coplanaires et non parallèles (ces deux segments appartiennent donc à la classe  $C_2$  et les hypothèses associées sont décrites sur la deuxième ligne de la Figure 6). L'hypothèse la plus vraisemblable dans ce cas est de supposer la présence d'une jonction entre ces deux segments. Cependant, compte tenu de l'incertitude associée aux calculs des paramètres des deux segments (et en particulier de leur longueur), il est aussi possible de supposer qu'il existe un segment reliant leur extrémité voisine. Les autres hypothèses (fusion des segments, etc) sont envisageables mais très peu probables. Ce type de raisonnement nous permet de définir la table de probabilités conditionnelles associées à cette classe. Les tableaux 1 et 2 montrent les

hypothèses pour les extrémités proches et éloignées des segments considérés. Ces tables sont définies de manière empirique. Cependant, fixer précisément les valeurs de chaque probabilité n'est pas fondamental. Elles doivent juste refléter de manière grossière la connaissance que l'on souhaite transmettre au système. Comme nous l'avons déjà indiqué, l'hypothèse la plus vraisemblable n'est pas toujours celle qui est correcte. C'est pourquoi nous retenons, dans chacune des deux séries d'hypothèses émises, les deux hypothèses auxquelles les plus fortes confiances ont été accordées. Une approche similaire dans un contexte de reconnaissance d'objets est proposée dans [8]. Les deux hypothèses sélectionnées font l'objet d'une vérification, étape que nous décrivons à présent.

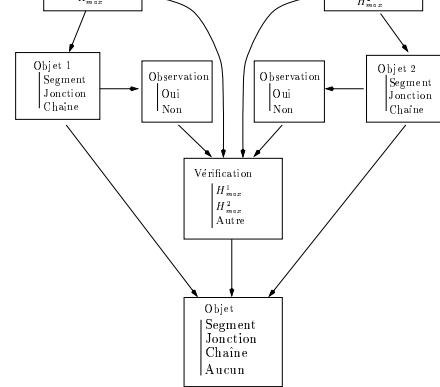


FIG. 7 – Réseau de vérification d'hypothèses

hypothèse	Classe				
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$H_1$ Segment	0.90	0.025	0.05	0.05	0.025
$H_2$ Jonction	0.025	0.60	0.05	0.15	0.025
$H_3$ Chaîne	0.025	0.30	0.85	0.75	0.025
$H_4$ Autres	0.05	0.075	0.05	0.05	0.925

TAB. 1 – Table de probabilité conditionnelle  $P(\text{hypothèse} \mid \text{classe})$  pour les extrémités proches (Figure 6)

hypothèse	Classe				
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$H_1$ Segment	0.025	0.025	0.025	0.025	0.025
$H_2$ Jonction	0.025	0.025	0.025	0.025	0.025
$H_3$ Chaîne	0.025	0.5	0.7	0.5	0.025
$H_4$ Autres	0.925	0.45	0.25	0.45	0.925

TAB. 2 – Table de probabilité conditionnelle  $P(\text{hypothèse} \mid \text{classe})$  pour les extrémités éloignées

### 3.4 Vérification des hypothèses émises

Afin de vérifier la validité des hypothèses émises, nous utilisons le réseau Bayésien présenté sur la Figure 7. Deux réseaux similaires sont en fait utilisés, chacun étant associé à l'un des jeux d'hypothèses émises (*i.e.*, correspondant l'un aux extrémités voisines des segments et l'autre aux extrémités éloignées). Pour chaque hypothèse, la nature et la position de l'objet qui lui est associé sont tout d'abord définies. Puis, le système vérifie par des observations déjà acquises ou à réaliser si cet objet existe ou non. Ensuite, connaissant la confiance accordée à chacune des deux hypothèses retenues, ainsi que la vraisemblance des observations associées, il est possible de déterminer laquelle des deux hypothèses était valide (ou éventuellement de les rejeter toutes deux). Enfin, compte tenu du résultat de la validation, on peut en déduire l'objet associé au couple de segments considéré.

**Observations - Actions de vérification :** La partie la plus importante du réseau correspond aux nœuds d'*observation*, car elle fait directement intervenir les stratégies de perception. En effet, si dans certains cas une hypothèse peut être directement confirmée ou infirmée en se basant uniquement sur les informations

déjà acquises, dans d'autres cas, cette information peut s'avérer insuffisante. Il est alors nécessaire de déplacer la caméra afin d'acquérir (ou non) le complément d'information attendu.

À chaque hypothèse émise, le réseau Bayésien associe un objet 3D à détecter dans la scène. Cet objet 3D est soit une jonction entre deux segments, soit un autre segment 3D. La vérification s'effectue dans un premier temps en utilisant l'information 3D déjà disponible et les observations 2D déjà réalisées. Il convient tout d'abord de calculer, pour toutes les positions précédemment occupées par la caméra, la position dans le plan image de l'objet considéré. Pour cela, on procède par rétro-projection du modèle 3D supposé dans le plan image. Nous cherchons ensuite à mettre en correspondance les segments 2D observés dans plusieurs images (afin d'éviter de faux appariements) avec la projection de l'objet 3D. Pour chaque appariement possible, une mesure de la confiance accordée à cet appariement est calculée.

Si l'objet recherché est une jonction entre deux segments, et dans l'hypothèse où cette jonction existe, elle a forcément déjà été observée. Le cas où l'objet supposé est un segment est plus intéressant. Dans une chaîne composée de trois segments, l'existence de deux d'entre eux a déjà été confirmée (ce sont eux qui ont servi à prédire la présence de l'autre). Ce dernier, par contre, n'a pas encore été reconstruit et sa présence doit être validée. Dans le cas où aucun appariement n'est trouvé en utilisant les images déjà acquises, il convient d'essayer de déterminer la raison pour laquelle le segment attendu n'a pas été observé. Il existe deux raisons possibles : la première est bien entendu que l'objet considéré n'existe pas, et la seconde est qu'il ait été occulté pour l'ensemble des points de vue précédents. Il convient alors de déplacer la caméra afin de permettre l'observation de l'objet attendu.

Plusieurs techniques sont envisageables pour calculer ce nouveau point de vue (citons par exemple les techniques mises au point par Cowan et Kovesi [7] ou Tarabani [21]). Dans notre cas, plutôt que de calculer explicitement ce point de vue, puis de rechercher *off-line* le segment supposé dans l'image, nous avons dé-

ments appartenant à la fois au polygone d'occultation et au plan dans lequel se situe le segment recherché (voir Figure 11). Pendant ce mouvement de rotation, généré par asservissement visuel, nous recherchons *on-line* l'apparition dans l'image du segment supposé. Pour cela, nous disposons un "observateur" à une position adéquate dans l'image. En l'occurrence, nous cherchons à détecter un segment ayant une certaine orientation et une certaine position, déterminées ici encore par rétro-projection. La position de l'observateur dépend donc de ces paramètres. La détection éventuelle du segment, s'il existe, se fera en recherchant un contour en mouvement ayant les caractéristiques attendues. L'exemple décrit sur la Figure 11 illustre la stratégie retenue.

### 3.5 Modélisation de la scène

La représentation de la scène dont nous disposons est, à ce stade de la reconstruction, une modélisation en termes de segments et jonctions 3D. L'étape suivante consiste, dans la mesure du possible, à passer de cette représentation à une représentation en termes de polygones 3D. Pour cela, nous utilisons les informations obtenues sur les jonctions 3D ainsi que les informations de coplanarité déjà utilisées lors de la génération d'hypothèses.

Le réseau Bayésien présenté sur la Figure 8 sert à représenter les connaissances acquises sur la scène et à guider le raisonnement permettant d'aboutir à la conclusion. Notre objectif est de savoir si deux objets  $O_1$  et  $O_2$  font en fait partie du même objet (et si oui quel est cet objet).  $O_1$  et  $O_2$  sont soit des segments, soit des chaînes de segments précédemment construites en utilisant cette même approche. Le nœud central du réseau permet, le cas échéant, de chaîner deux objets connaissant à la fois la probabilité d'existence de ces objets, la probabilité qu'ils soient coplanaires et la probabilité qu'il y ait une jonction entre ces deux objets<sup>1</sup>. Ensuite, connaissant la confiance que l'on accorde au chaînage des deux objets et la probabilité de savoir si la chaîne est fermée (fermeture définie par la jonction  $(S_i, S_n)$ ), il est possible de déduire la probabilité que l'objet final soit formé de deux chaînes indépendantes, une seule chaîne, ou un polygone.

## 4 Résultats expérimentaux

Nous présentons ici les résultats de la reconstruction d'un polyèdre (voir Figure 3). Ce polyèdre est formé de six polygones non coplanaires. Cet objet permet d'illustrer les intérêts de la méthode proposée. La reconstruction de cet objet pose en effet un grand nombre de difficultés : segments trop petits pour être reconstruits, présence d'occultations, caractéristiques géométriques peu visibles (par exemple, l'angle aigu est coupé (voir Figure 3.b), ce qui induit la présence d'un

1. Dans le réseau de la Figure 8, une seule jonction  $(S_j, S_k)$  est représentée. En fait, si l'objet considéré est une chaîne, il existe plusieurs autres jonctions possibles (*i.e.*,  $(S_i, S_k)$ ,  $(S_j, S_n)$ ,  $(S_i, S_n)$ , etc).

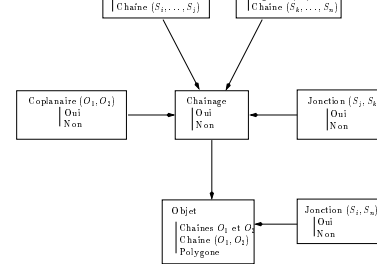


FIG. 8 – Réseau de modélisation

segment d'une longueur de 1 cm environ).

La Figure 9.a montre la première image de cet objet et les Figures 9.b à 9.f décrivent les itérations suivantes de la reconstruction incrémentale. La reconstruction de chacun de ces segments est effectuée en utilisant la méthode décrite brièvement au début de cet article, ce qui explique que la projection de la dernière primitive reconstruite apparaisse horizontale ou verticale centrée dans l'image. Une numérotation des segments dans l'ordre de leur reconstruction est donnée sur la Figure 10.

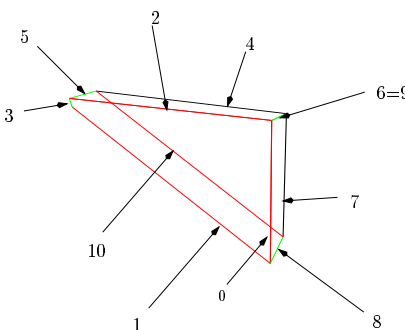


FIG. 10 – Scène reconstruite et numérotation des segments dans l'ordre de leur introduction dans la carte de l'environnement

Pour chaque segment introduit dans la carte de l'environnement, le processus de prédiction/vérification d'hypothèses est utilisé pour déterminer les relations de ce segment avec ceux précédemment reconstruits. Nous allons maintenant détailler ce processus.

Quand les deux premiers segments  $S_0$  et  $S_1$  ont été reconstruits, les observations réalisées sont les suivantes : la probabilité que  $S_0$  et  $S_1$  soient coplanaires est de 99%, voisins de 91%, et parallèles de 1%. La classe d'observation la plus probable à laquelle appartiennent ces deux segments est donc la classe  $\mathcal{C}_2$ .

L'utilisation du réseau Bayésien générateur d'hypothèses pour les extrémités proches permet d'émettre les hypothèses suivantes : la probabilité que  $S_0$  et  $S_1$  soient identiques est de 3% ; qu'il y ait une jonction entre  $S_0$  et  $S_1$  de 55% ; qu'il y ait un segment reliant  $S_0$  et  $S_1$  entre leurs deux plus proches extrémités de 35% ; et, enfin, la probabilité qu'il n'y ait aucune relation entre  $S_0$  et  $S_1$  est de 7%. Ces valeurs sont différentes de celles que l'on trouve dans le Tableau 1 car elles dépendent bien sûr de la probabilité pour  $S_0$  et

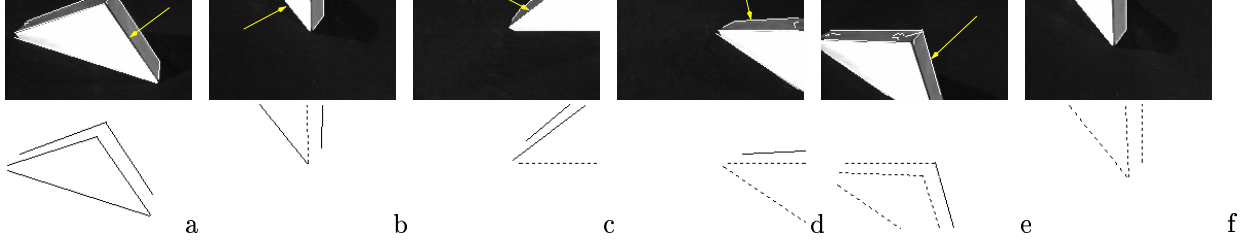


FIG. 9 – *Reconstruction incrémentale de la scène polyèdre. Les lignes en pointillés correspondent aux segments déjà reconstruits; les autres correspondent à des segments non encore traités. La flèche pointe sur le prochain segment à reconstruire*

$S_1$  d'appartenir à telle ou telle classe, et donc des probabilités  $p(C_{01})$ ,  $p(V_{01})$  et  $p(P_{01})$  obtenues à partir de la position 3D des segments reconstruits.

Conformément à la stratégie codée dans les tables de probabilités conditionnelles, l'hypothèse d'une jonction est favorisée de même que, mais dans une moindre mesure, l'hypothèse de la présence d'un segment reliant  $S_0$  et  $S_1$ . Les images précédemment acquises permettent de valider l'hypothèse principale "jonction" avec une probabilité de 98%, alors que l'hypothèse secondaire de la présence d'un segment est rejetée, celui-ci n'ayant pas été observé. On en déduit la probabilité de la présence d'un objet jonction à 85%. Au niveau de la modélisation de la scène, la présence d'une jonction entre deux segments coplanaires permet ensuite de créer un chaîne de segments ayant une probabilité d'existence de 98%.

Un segment est également supposé exister entre les extrémités éloignées de  $S_0$  et  $S_1$ . Cette hypothèse est vérifiée par l'observation de  $S_2$ . La longueur de  $S_2$  est suffisamment importante pour permettre sa reconstruction. Une estimation optimale des paramètres de ce segment est donc réalisée. Il convient ensuite de considérer les relations entre le segment  $S_0$  et le segment  $S_2$ , mais aussi entre les segments  $S_1$  et  $S_2$ . L'hypothèse la plus probable associée au couple  $(S_0, S_2)$  est, comme dans le cas précédent, une jonction; cette hypothèse est très facilement vérifiée en utilisant les informations présentes dans l'image. Le comportement du système est similaire à celui présenté pour le couple  $(S_0, S_1)$ . Nous ne le détaillerons donc pas. Précisons simplement que l'ajout de cette jonction permet, au niveau de la modélisation de la scène, la création d'une chaîne de segments coplanaires  $(S_2, S_0, S_1)$  ayant une probabilité d'existence de 97%.

Considérons à présent le cas du couple  $(S_1, S_2)$ . Ces deux segments sont dans l'espace 3D très proches l'un de l'autre (environ 1cm sépare leur extrémité voisine). Les probabilités qu'ils soient voisins et coplanaires sont respectivement de 61% et 99%. La classe la plus probable à laquelle le système associe ces deux segments est, comme dans le cas précédent, la classe  $C_2$ . Les hypothèses les plus probables générées sont donc dans ce cas qu'il y ait une jonction entre  $S_1$  et  $S_2$  (46%) et qu'il y ait un segment reliant les deux segments considérés (41%). Les 13% restants sont à partager entre

les deux autres hypothèses.

Lors de la vérification de ces hypothèses, la première (jonction) est vérifiée avec une probabilité de 60%. Cette probabilité relativement élevée résulte du fait que les projections des extrémités de  $S_1$  et  $S_2$  sont proches l'une de l'autre dans l'espace image (environ 5 pixels), ce qui a tendance à renforcer (avec une certaine "prudence") la première hypothèse. Cependant, la seconde hypothèse s'avère vérifiée avec une probabilité de 95%. Un segment 2D, correspondant à la projection du segment prédit, apparaît en effet à la position attendue dans plusieurs images. Finalement, compte tenu de la confiance accordée aux deux hypothèses et de celle accordée aux observations, un nouveau segment  $S_3$  est introduit dans la carte de la scène (avec une confiance de 53%, alors que la confiance accordée à l'ajout d'une jonction n'est que de 37%). On constate ici tout l'intérêt de considérer plusieurs hypothèses en parallèle puisqu'une technique classique de prédiction/vérification aurait vraisemblablement validée la première hypothèse.

Le quatrième segment  $S_3$  est donc introduit dans la carte de la scène. Contrairement aux trois précédents, celui-ci ne résulte pas d'une reconstruction explicite. Il convient cependant de regarder quelles sont les relations entre ce segment et les segments précédemment reconstruits. Ce cas est assez simple, et les jonctions  $(S_1, S_3)$  et  $(S_3, S_2)$  sont détectées avec une très forte probabilité, ce qui permet de créer un polygone  $(S_3, S_1, S_0, S_2)$  avec une confiance élevée de 98%.

La reconstruction des arêtes suivantes du polyèdre s'effectue de manière similaire et aboutit à la reconstruction de deux polygones supplémentaires. On peut noter que la reconstruction du segment  $S_4$  a eu pour conséquence la création automatique des segments  $S_5$  et  $S_6$  (après vérification de leur présence). De manière similaire, la reconstruction du segment  $S_7$  entraîne la génération d'une hypothèse portant sur l'ajout de deux segments:  $S_8$  et un éventuel segment " $S_9$ " dont la position est identique à celle du segment  $S_6$ . La présence de ces deux segments est vérifiée et ils sont donc introduits dans la carte de l'environnement. La recherche des relations entre  $S_9$  et les huit autres segments amène le système à émettre l'hypothèse que les segments  $S_6$  et  $S_9$  sont identiques et doivent être fu-



le segment  $S_9$  est supprimé de la carte. La reconstruction du segment  $S_7$  a aussi pour conséquence la création d'une jonction avec  $S_4$ . De plus, une seconde hypothèse correspondant à la présence d'un segment reliant les extrémités éloignées de  $S_7$  et  $S_4$  est aussi émise. Ce segment prédit n'apparaît dans aucune image précédemment acquise par la caméra. Cependant, connaissant la position des polygones reconstruits ainsi que les positions successivement occupées par la caméra, il est possible d'affirmer que ce segment n'a jamais pu être observé, les faces du polyèdre ne permettant pas son observation. Afin de vérifier l'hypothèse, un mouvement du capteur est donc réalisé. Comme nous l'avons indiqué précédemment, la caméra se focalise sur  $S_7$  et effectue un mouvement de rotation autour de ce segment. Pendant ce mouvement, généré par asservissement visuel, les observateurs recherchent un contour en mouvement à la position attendue du segment dans l'image (voir Figure 11.ab). Le segment découvert est ensuite reconstruit (voir Figure 11.c) et inséré dans la carte de l'environnement.

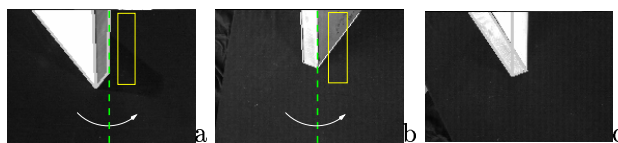


FIG. 11 – *Vérification d'une hypothèse: (a) rotation autour de  $S_7$ , (b)  $S_{10}$  est découvert et (c) reconstruit*

On peut finalement noter l'absence d'un segment reliant  $S_{10}$  et  $S_1$ . L'hypothèse de sa présence a été faite avec une probabilité de 83% lors de l'ajout de  $S_{10}$ , mais sa présence n'a pu être vérifiée en raison de sa position par rapport aux autres polygones et en raison des contraintes mécaniques du manipulateur, empêchant ainsi son observation par la caméra.

Signalons que les six images présentées sur la Figure 9 ont été suffisantes pour vérifier l'ensemble des autres hypothèses émises. Sur ces images sont effectuées une extraction des contours par un filtrage de Shen-Castan et une approximation polygonale des contours afin d'obtenir l'ensemble  $\Omega$  de segments. La reconstruction d'un segment est elle effectuée à une cadence de 10 Hz et requiert l'acquisition et le traitement d'environ 200 images, soit le temps nécessaire à la caméra pour réaliser un mouvement d'amplitude suffisante permettant d'obtenir une précision de reconstruction de l'ordre du millimètre. De plus, compte tenu des calculs nécessaires au processus de propagation dans les réseaux Bayésiens, il faut donc environ 5 minutes pour reconstruire en ligne une scène de ce type.

D'autres scènes ont également été reconstruites en utilisant la même approche (voir par exemple les résultats présentés sur la Figure 12). Notons que la méthode de reconstruction 3D permet aussi la reconstruction de cylindres. L'exploration de cette scène en utilisant le seul module de reconstruction incrémentale n'est cependant pas complète si l'image initiale n'en fournit

des objets disjoints de ceux reconstruits n'ont jamais été observés. En greffant à la reconstruction incrémentale le module d'exploration globale décrit dans [15], on aboutit à la complétude de la reconstruction (voir Figure 12.d).

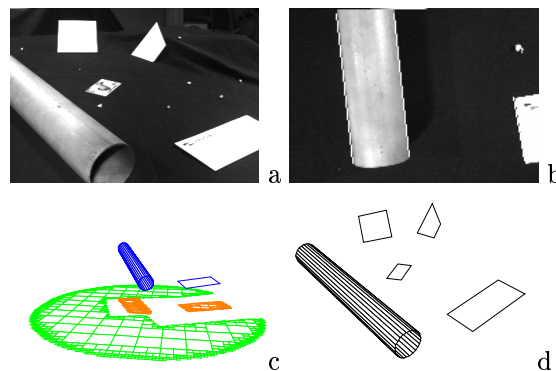


FIG. 12 – *Autre scène reconstruite: (a) vue extérieure de la scène, (b) image initiale, (c) scène reconstruite après la première phase de reconstruction incrémentale et projection sur un plan virtuel des zones inconnues, (d) scène reconstruite après l'exploration globale*

Concernant les aspects de prédiction et de vérification d'hypothèses, on peut signaler qu'ils ont permis la détection des jonctions entre segments et la création des quatre polygones présents dans la scène. Cette scène nous permet également d'illustrer un point important concernant la combinatoire de notre système. En théorie, celle-ci peut en effet augmenter assez rapidement puisqu'il est nécessaire de considérer les relations entre chaque nouveau segment introduit dans le modèle de la scène avec ceux déjà reconstruits. Cependant, en pratique, on ne génère des hypothèses qu'entre segments qui ne sont pas trop distants. Cette simple heuristique permet de traiter de manière indépendante des objets nettement disjoints et ne comporte guère de conséquences facheuses: si une relation existe entre deux segments éloignés, elle sera très probablement obtenue par l'observation de segments, soit lors de la reconstruction incrémentale, soit lors de l'exploration globale. De plus, seules les hypothèses émises avec une probabilité supérieure à un seuil donné sont considérées (typiquement 10%), ce qui permet d'éviter de vérifier des hypothèses très improbables. Finalement, on peut rappeler que pour les segments considérés comme non coplanaires, les connaissances introduites dans le réseau Bayésien de génération d'hypothèses impliquent qu'il n'y a pas de relation (ou une relation complexe) entre ces segments, et qu'il n'y a donc pas de vérification de cette relation à effectuer. Ces différentes considérations font que, même pour des objets complexes ou des scènes composées de plusieurs objets, le nombre de relations effectives rapporté au nombre de segments reconstruits décroît rapidement en même temps que le nombre de vérifications à réaliser.

Dans cet article, nous avons présenté des stratégies de perception par **vision active** permettant la reconstruction d'objets polyédriques. Un algorithme, simple et efficace, permet de reconstruire de façon incrémentale toutes les primitives qui ont été observées par la caméra. Afin de passer d'une modélisation locale incomplète de la scène à une modélisation plus complète et de haut niveau (en termes d'objets), nous avons couplé à la reconstruction incrémentale un algorithme reposant sur les techniques de prédiction/vérification d'hypothèses. Les stratégies retenues ont été codées dans des réseaux Bayésiens. Elles permettent en outre de gérer et résoudre les problèmes locaux d'occlusion. L'approche développée a finalement fait l'objet de validations expérimentales sur diverses scènes réelles. Il faut bien reconnaître que les images que nous avons considérées sont très peu bruitées (voire même quasiment binaires). Cette simplicité est à ce jour imposée par le fait que les mouvements de la caméra pendant les phases de reconstruction et de vérification sont calculées en boucle fermée à partir des informations extraites des images. De fait, le traitement d'images doit avoir lieu à la cadence vidéo et il existe aujourd'hui peu de systèmes fiables capables de suivre des objets même simples en temps réel dans un environnement bruité.

Dans le problème que nous venons de présenter, les connaissances *a priori* sur la scène étaient minimales, excepté bien sûr le fait que nous n'avons considéré que des objets polyédriques. L'introduction d'hypothèses supplémentaires aurait pu être envisagée (comme par exemple : "une jonction a une forte probabilité d'être triple"). Cependant, cette introduction ne nous a pas paru primordiale, voire justifiée, puisque les hypothèses rajoutées doivent être choisies en fonction de l'intérêt qu'elles apportent au traitement de l'application particulière considérée.

La reconstruction de scènes non polyédriques est bien entendu une des perspectives majeures de ce travail. De plus, l'utilisation des techniques de prédiction/vérification dans le cadre de la vision active n'est assurément pas limitée au problème de la reconstruction 3D. Ces méthodes peuvent en effet s'appliquer à de multiples problèmes d'analyse de scènes en environnement inconnu ou partiellement connu. L'introduction de l'action au sein du processus codant les stratégies de perception (et donc le raisonnement à suivre) permet d'envisager, de manière générale, une intégration très forte du cycle perception-action en vision intentionnelle.

## Remerciements

Cette étude a été effectuée dans le cadre du projet inter-PRC VIA (Perception et raisonnement géométrique pour le mouvement dans le monde réel / PRC-IA et PRC-CHM). Les auteurs tiennent à remercier David Djian et Patrick Rives pour les discussions qu'ils ont eues avec eux à propos de l'utilisation des réseaux Bayésiens en vision active.

- [1] J.M. Agosta. The structure of bayes nets for vision recognition. In *Proc. of 4th Wk on Uncertainty in Artificial Intelligence*, Minneapolis, Août 1988.
- [2] N. Ayache, O. Faugeras. Hyper: A new approach for the recognition and positioning of two-dimensional objects. *IEEE Trans. on PAMI*, 8(2):44–54, Janvier 1986.
- [3] N. Ayache, B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *IJCV*, 1(2):107–132, 1987.
- [4] R.C. Bolles, R. Horaud. 3DPO: A three-dimensional part orientation system. *Int. J. of Robotics Research*, 5(3):3–26, Juin 1986.
- [5] H. Buxton, S. Gong. Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, 78(1-2):431–459, Octobre 1995.
- [6] F. Chaumette, S. Boukir, P. Bouthemy, D. Juvin. Structure from controlled motion. *IEEE Trans. on PAMI*, 18(5):492–504, Mai 1996.
- [7] C.K. Cowan, P.D. Kovesi. Automatic sensor placement from vision task requirements. *IEEE Trans. on PAMI*, 10(3):407–416, Mai 1988.
- [8] D. Djian, P. Probert, P. Rives. Reconnaissance de modèles géométriques simples à l'aide de réseaux bayésiens. In *RFIA '96*, pp. 396–404, Rennes, Janvier 1996.
- [9] B. Espiau, F. Chaumette, P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, Juin 1992.
- [10] G. Giraudon, F. Sandakly. Interprétation de scènes d'intérieur pour un robot mobile. *TSI*, 16(1):69–100, Janvier 1997.
- [11] M. Herman, T. Kanade. Incremental reconstruction of 3D scenes from multiple, complex images. *Artificial Intelligence*, 30(3):289–341, Décembre 1986.
- [12] S.A. Hutchinson, A.C. Kak. Planning sensing strategies in a robot work cell with multi-sensor capabilities. *IEEE Trans. on Robotics and Automation*, 5(6):765–783, Décembre 1989.
- [13] V.P. Kumar, U.B. Desai. Image interpretation using bayesian networks. *IEEE Trans. on PAMI*, 18(1):74–77, Janvier 1996.
- [14] E. Marchand. *Stratégies de perception par vision active pour la reconstruction et l'exploration de scènes statiques*. Thèse Université de Rennes 1, IRISA, Juin 1996.
- [15] E. Marchand, F. Chaumette. Controlled camera motions for scene reconstruction and exploration. In *CVPR '96*, pp. 169–176, San Francisco, Juin 1996.
- [16] R. Neapolitan. *Probabilistic Reasoning in Expert Systems*. John Wiley, Chichester, 1990.
- [17] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, San Mateo, 1988.
- [18] R.D. Rimey, C. Brown. Controlling eye movements with hidden markov models. *IJCV*, 7(1):47–65, Janvier 1991.
- [19] R.D. Rimey, C. Brown. Control of selective perception using bayes nets and decision theory. *IJCV*, 12(2/3):173–207, Avril 1994.
- [20] S. Sarkar, K. Boyer. Integration, inference, and management of spatial information using bayesian networks: perceptual organization. *IEEE Trans. on PAMI*, 15(3):256–274, Mars 1993.
- [21] K. Tarabanis, R. Tsai, P.K. Allen. The MVP sensor planning system for robotic vision tasks. *IEEE Trans. on Robotics and Automation*, 11(1):72–85, Février 1995.