
Stratégies de perception par vision active : Une approche par réseaux bayesiens

Éric Marchand et François Chaumette

*IRISA / INRIA Rennes
Campus de Beaulieu
35042 Rennes cedex, France*

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'occlusion. 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.

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.

MOTS CLÉS: Vision active, reconstruction 3D, réseaux bayesiens, prédiction/vérification d'hypothèses.

KEY WORDS: 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 [ALO 87, ALO 90], Bajcsy [BAJ 88] ou encore Ballard [BAL 91] ont proposé de modifier cet état de fait en élaborant le concept de vision active [SWA 93]. Les techniques de vision active tirent leur origine d'une tentative de simulation du système visuel humain en essayant de recréer ses facultés d'adaptation. 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 consiste en effet à élaborer des stratégies de perception intelligentes, en contrôlant les paramètres du capteur (position, vitesse, mise au point, etc.). Elle 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 de segments et 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 reconstruction et, à un niveau global, des déplacements 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 [BOU 93, CHA 96]. 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.*, le point, la droite, le

cercle, la sphère et le cylindre) [BOU 93]. 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 [BOU 93, CHA 96] que la réalisation de 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 [ESP 92]. 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.

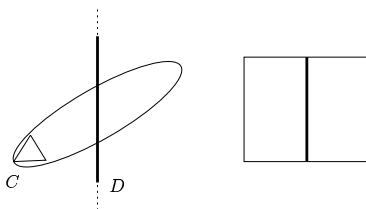


Figure 1. *Mouvements de la caméra pour la reconstruction optimale d'une droite et image correspondante*

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. 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 de "stratégies de perception" est divisé en deux étapes distinctes.

- La première étape que nous allons décrire dans cet article se propose de reconstruire de manière incrémentale l'ensemble des segments qui apparaissent dans le champ de vision de la caméra. Cette phase est locale dans le sens où seules les informations disponibles sont utilisées afin de déterminer sur quelle primitive la caméra doit se focaliser.
- Puis, quand toutes les primitives précédemment observées ont été reconstruites, une stratégie plus globale est mise en œuvre de manière à focaliser la caméra sur les zones de la scène n'ayant pas encore été traitées, aboutissant ainsi à la complétude de la reconstruction. Cette étape ne sera pas décrite dans cet article, mais seulement évoquée en section 4. Une

description complète de l'exploration globale est en effet donnée dans [MAR 96b, MAR 96a] et [MAR 97b].

Nous décrivons tout d'abord un algorithme simple et efficace, qui permet de reconstruire 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'occlusion. 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 bayesiens. Ces réseaux permettent d'émettre des hypothèses sur l'existence et la localisation de nouveaux segments, puis de proposer l'exécution d'une action conduisant à vérifier ou à infirmer cette hypothèse, et enfin, en fonction du résultat de l'étape de vérification, de compléter le modèle 3D de la scène.

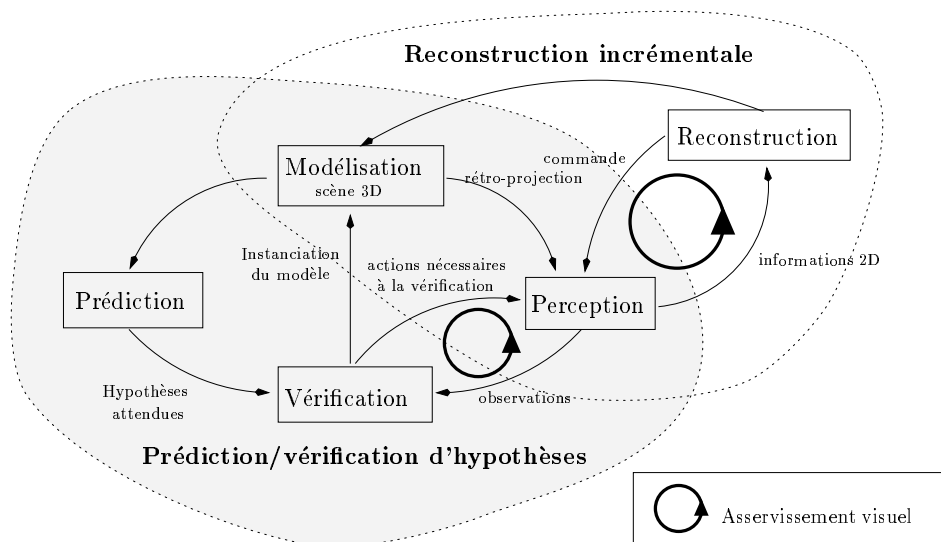


Figure 2. *Principe général*

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.
- **Instanciation des modèles de la scène.** Enfin, 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. Les conclusions sont finalement propagées au reste du système.

L'utilisation des réseaux bayésiens implique une connaissance parfaite de la tâche à effectuer. La définition du réseau permet de coder les connaissances *a priori* dont on dispose ainsi que les stratégies d'interprétation et d'analyse de la scène. Par conséquent, l'utilisation de réseaux bayésiens implique que l'on se place résolument dans le cadre de la **vision intentionnelle** [ALO 90]. Dans notre cas, nous avons tenté de réduire cette connaissance *a priori* au minimum afin de ne pas s'opposer aux hypothèses initiales (*i.e.*, aucune connaissance sur le nombre, les dimensions et la position des objets considérés).

Dans la suite de cet article, nous présentons tout d'abord l'algorithme de reconstruction incrémentale. Puis, nous détaillons l'étape fondamentale de cet algorithme, à savoir la génération et la vérification d'hypothèses. Nous présentons finalement des résultats issus d'expériences réalisées sur la cellule de vision robotique de l'IRISA qui valident notre approche.

2. Reconstruction incrémentale de la scène

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 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. Nous appellerons par la suite cet ensemble de

segments $\omega_{\phi_t} = \{\mathcal{S}_i, i = 1 \dots n_t\}$ où ϕ_t est la position de la caméra depuis laquelle les n_t segments \mathcal{S}_i sont visibles.

Un autre ensemble, noté $\Omega_{\mathcal{T}_{t_1}^{t_2}}$, est également utilisé. Il contient l'ensemble des segments non reconstruits qui ont été précédemment observés, ainsi que les positions de la caméra depuis lesquelles ils ont été observés. Plus précisément, nous avons :

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

où

- $\mathcal{T}_{t_1}^{t_2} = \{\phi_{t_1}, \phi_{t_1+1}, \dots, \phi_{t_2}\}$ représente l'ensemble des positions occupées par la caméra entre les instants t_1 et t_2 ¹. Initialement, $\mathcal{T}_0^0 = \{\phi_0\}$ représente la première position de la caméra.
- \mathcal{S}_i représente un segment 2D non reconstruit et ϕ_k la position depuis laquelle il a été vu. ϕ_k appartient à $\mathcal{T}_{t_1}^{t_2}$
- N représente le nombre de segments présents dans cet ensemble.

L'algorithme de traitement d'image que nous utilisons pour l'extraction et le suivi des segments pendant la phase de reconstruction 3D ne permet de suivre en temps réel qu'un faible nombre de primitives simultanément. Par conséquent, les ensembles ω_{ϕ_t} et $\Omega_{\mathcal{T}_{t_1}^{t_2}}$ ne sont pas construits à chaque itération du processus de reconstruction des primitives, mais seulement à l'initialisation de l'algorithme et à la fin de la reconstruction de chaque segment.

2.1. Algorithme de reconstruction incrémentale

Ces définitions étant posées, nous présentons à présent l'algorithme permettant de reconstruire incrémentalement l'ensemble des primitives observées par la caméra. Cet algorithme est composé de trois étapes principales : la reconstruction d'une primitive 3D, la mise à jour des informations 2D disponibles, et le choix du prochain segment à reconstruire.

- Étape 0 : Initialisation

Nous considérons que la caméra est positionnée en ϕ_0 . Depuis cette position, la caméra observe l'ensemble de segments ω_{ϕ_0} et nous avons $\Omega_{\mathcal{T}_0^0} = \{(\mathcal{S}_i, \phi_0), i = 1 \dots n_0\}$.

Un des segments \mathcal{S}_i est ensuite extrait de l'ensemble ω_{ϕ_0} afin d'estimer la primitive 3D qui lui est associée. Comme nous l'avons déjà dit, la primitive doit apparaître verticale ou horizontale centrée dans l'image. Afin d'accélérer l'étape de reconstruction, nous choisissons donc le segment qui est le plus proche de l'une de ces deux positions.

1. Par la suite, nous appellerons abusivement $\mathcal{T}_{t_1}^{t_2}$ la "trajectoire" de la caméra entre t_1 et t_2 .

– **Étape 1 : Reconstruction du segment sélectionné et mise à jour de la carte**

Considérons que la caméra est positionnée en ϕ_t ($\phi_t = \phi_0$ à la première itération de la reconstruction).

Une reconstruction de la primitive correspondant au segment \mathcal{S}_i sélectionné est réalisée. Cette phase de reconstruction inclut le processus d'estimation active décrit dans [BOU 93] 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 ici qu'intervient le processus de prédiction/vérification d'hypothèses qui sera décrit dans la section suivante. Nous reviendrons à la fin de cette section sur les motivations qui nous ont poussés à introduire cette étape dans le processus de reconstruction de la scène.

Signalons finalement qu'il convient aussi de retirer de la liste des segments $\Omega_{\mathcal{T}_0^t}$ l'ensemble des segments 2D correspondant à la primitive reconstruite. En effet, cette primitive a pu être observée depuis d'autres positions précédemment occupées par la caméra. La suppression s'effectue par une simple mise en correspondance entre la rétro-projection du segment reconstruit et les segments de $\Omega_{\mathcal{T}_0^t}$.

– **Étape 2 : Création des listes locales et globales de segments 2D**

À l'issue de la reconstruction de la primitive, la caméra est située en ϕ_{t+1} . Une nouvelle liste de segments $\omega_{\phi_{t+1}}$, correspondant à cette position, est créée. Cette liste est ensuite fusionnée avec l'ensemble $\Omega_{\mathcal{T}_0^t}$:

$$\Omega_{\mathcal{T}_0^{t+1}} = \Omega_{\mathcal{T}_0^t} \cup \{(\mathcal{S}_i, \phi_{t+1}), i = 1 \dots n_{t+1}\}$$

– **Étape 3 : Focalisation sur le segment suivant**

Trois cas de figures peuvent alors se présenter :

1. Dans le cas général, il y a au moins un segment dans l'ensemble $\omega_{\phi_{t+1}}$ qui correspond à une primitive non reconstruite (*i.e.*, cet ensemble n'est pas vide, voir figure 3.a). Un segment \mathcal{S}_i est alors extrait de cet ensemble. Une estimation des paramètres de la primitive associée est alors réalisée (étape 1). Ces différentes étapes (estimation, création des ensembles de segments, et sélection) sont itérées tant qu'il existe un segment non reconstruit dans l'ensemble $\omega_{\phi_{t+1}}$.
2. Le deuxième cas de figure concerne le cas où tous les segments observés à la position courante ϕ_{t+1} correspondent à des primitives déjà reconstruites (voir figure 3.b). Dans ce cas, il se peut que des segments précédemment observés (et qui sont donc dans $\Omega_{\mathcal{T}_0^{t+1}}$) n'aient pas encore été traités.

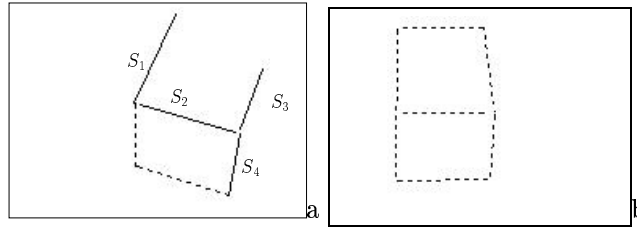


Figure 3. Ensemble des segments perçus dans l'image, les traits pleins correspondent à l'ensemble ω_{ϕ_t} (a) cas 1 : $\omega_{\phi_{t+1}} \neq \emptyset$, (b) cas 2 : $\omega_{\phi_{t+1}} = \emptyset$

Une phase de retour arrière (*backtracking*) est alors effectuée en sélectionnant dans l'ensemble $\Omega_{\mathcal{T}_0^{t+1}}$ le couple (\mathcal{S}_i, ϕ_k) pour lequel la distance entre la position courante de la caméra ϕ_{t+1} et ϕ_k est la plus faible (nous cherchons à réduire les déplacements de la caméra). La caméra se replace alors à cette position ϕ_k ($\phi_{t+2} = \phi_k$) et une reconstruction du segment \mathcal{S}_i est ensuite réalisée (retour à l'étape 1).

3. Finalement, si l'ensemble $\Omega_{\mathcal{T}_0^{t+1}}$ est lui aussi vide (c'est-à-dire que l'ensemble des segments 2D observés entre la date 0 et la date $t + 1$ ont fait l'objet d'une reconstruction), le processus de reconstruction incrémentale de la scène prend fin.

2.2. Discussion

L'algorithme que nous venons de présenter permet de reconstruire toutes les primitives qui se sont trouvées dans le champ de vision de la caméra. Nous avons seulement évoqué dans la description de cet algorithme l'étape de prédiction/vérification d'hypothèses (étape 1). Cette étape est cependant l'un des points clés de notre algorithme. En effet, en ne la considérant pas, le processus de reconstruction incrémentale présente les inconvénients suivants :

- 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 termes d'objets (jonctions, polygones, faces, etc.).
- 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 listes de segments. Par conséquent, ces primitives 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'occlusion, 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 4. Le modèle obtenu en utilisant une simple reconstruction incrémentale est donné sur la figure 4.c. Il présente l'ensemble des inconvénients décrits ci-dessus : le modèle 3D est constitué de cinq segments 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 4.d.

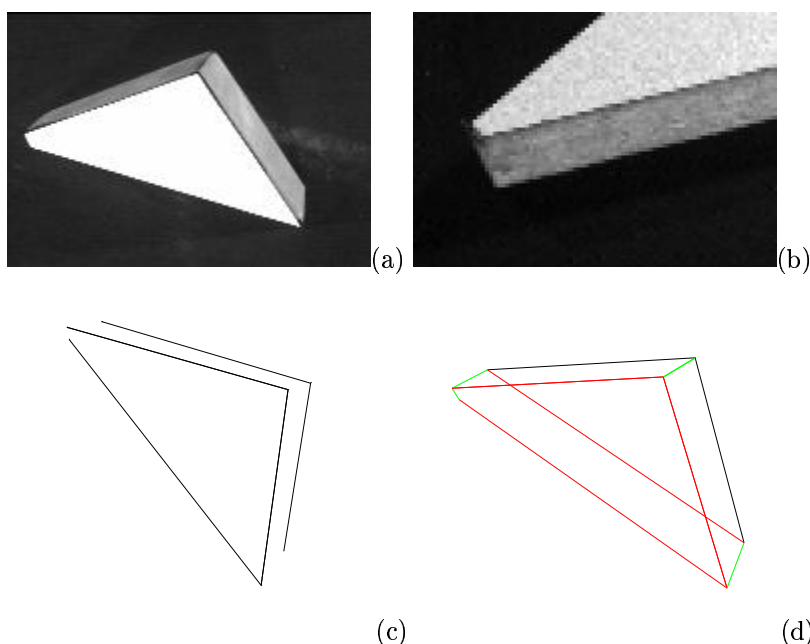


Figure 4. Scène "polyèdre" : (a) Vue de la scène à reconstruire, (b) présence d'un "petit" segment, (c) modèle de la scène acquis en se servant uniquement du module de reconstruction incrémentale, (d) modèle reconstruit en se servant du module de prédiction/vérification d'hypothèses

3. Prédiction/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 [AYA 86] ou

3D [BOL 86], mais aussi pour de nombreuses autres applications telles, par exemple, la mise en correspondance de segments dans un couple d'images stéréoscopiques [AYA 87], ou l'acquisition de modèles 3D à partir d'un ensemble d'images aériennes [HER 86]. Ces techniques consistent à émettre des hypothèses sur la présence et la position spatiale de primitives et à vérifier ensuite que des appariements 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 [SHA 76] visant à modéliser l'ignorance (et non pas l'incertitude) et qui a été utilisée en vision par Hutchinson et Kak [HUT 89] 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 [GIR 97]), les chaînes de Markov cachées (utilisées par Rimey et Brown [RIM 91] pour simuler le mouvement des yeux) ou les réseaux bayésiens [PEA 88]. 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 [SAR 93], la segmentation en niveaux de gris [KUM 96], la reconnaissance d'objets [AGO 88], etc. L'utilisation de ces réseaux en vision active est plus récente. Les travaux les plus discriminants ont été proposés par Rimey et Brown [RIM 94] avec le système TEA-1, Buxton et Gong [BUX 95], ou encore Djian et Rives [DJI 96a]. 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.

Rimey et Brown [RIM 94] ont développé le concept de vision sélective basée sur la collecte séquentielle d'informations pour répondre à des questions spécifiques. Ces questions concernent l'interprétation d'une scène de manière générale ou bien encore la détection de la présence d'un objet donné. Le système proposé n'exécute que les actions nécessaires pour pouvoir répondre à la question posée, et utilise une base de connaissance sur le modèle général de la scène (organisation) et sur les modèles particuliers des objets. L'ensemble de ces connaissances est regroupé dans des réseaux bayésiens. Certains de ces réseaux regroupent la connaissance sur un domaine spécifique, d'autres regroupent la

connaissance relative à une tâche précise. La méthode employée ci-dessus reste dédiée à des tâches très spécifiques qui nécessitent de définir précisément les *a priori* contenus dans les différents réseaux bayésiens. Une approche très similaire est proposée dans [JEN 91].

Buxton et Gong [BUX 95] cherchent à construire un système de surveillance du trafic routier. Ils ne souhaitent pas uniquement suivre les véhicules dans une séquence d'images mais désirent aussi analyser, interpréter et classifier leur comportement. Pour cela, il convient de disposer de connaissances *a priori* sur les comportements possibles des différents objets mis en jeu. De plus, il est nécessaire de concevoir un système permettant de définir les différentes actions de perception utiles à l'accomplissement de la tâche nominale. Les réseaux bayésiens fournissent un cadre unifié permettant de disposer à la fois de la connaissance *a priori* et de la modélisation du contrôle de tâche. Enfin, ils autorisent une gestion dynamique (grâce au processus de propagation) des dépendances entre les acteurs impliqués dans le processus d'interprétation visuelle. L'approche proposée est décomposée en trois points : fournir une interprétation possible (la plus probable) des données disponibles, puis mettre à jour cette interprétation en fonction des nouvelles données et de l'interprétation courante. Dans leur article, Buxton et Gong proposent un cadre méthodologique général intéressant pour la conception et la mise en œuvre de systèmes de vision intentionnelle.

Djian, Probert et Rives utilisent les réseaux bayésiens dans un cadre de reconnaissance d'objets [DJI 96a]. L'objectif est de reconnaître des modèles géométriques simples (des portes et des fenêtres) dans une séquence d'images. Pour cela, des observateurs dynamiques cherchent à détecter l'apparition de motifs élémentaires dans la scène. Les réseaux bayésiens servent ici à guider le raisonnement en prenant une décision sur l'action à exécuter (positionnement dans l'image des observateurs) et en émettant une hypothèse sur le motif que l'on doit détecter dans ces observateurs. Un deuxième réseau permet de vérifier si l'hypothèse émise est correcte.

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 (par exemple la 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 valeur. Un tel réseau peut être utilisé pour représenter 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. Des algorithmes de propagation ont été proposées (voir [KRA 93] pour un simple aperçu ou [NEA 90, SPI 90] pour une description plus détaillée).

3.2. Approche générale retenue

L'information sur la scène dont nous disposons est composée, initialement, uniquement d'un ensemble $\mathcal{S}(\mathcal{T}_0^{t-1})$ de segments 3D. $\mathcal{S}(\mathcal{T}_0^{t-1})$ est un sous-ensemble de l'ensemble $\mathcal{O}(\mathcal{T}_0^{t-1})$ qui représente l'ensemble des objets de la scène (et qui contient des segments 3D, des jonctions, des polygones, etc.). L'objectif 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 permettre de déterminer les conséquences induites par l'intégration de chaque nouveau segment S_t . Il est donc utilisé dès qu'un segment est rajouté à l'ensemble \mathcal{S} des segments 3D composant la scène.

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.

Les hypothèses à émettre sont directement reliées aux actions que l'on souhaite réaliser, 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 5) ;

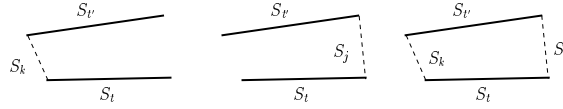


Figure 5. *Hypothèse H_2 : ajout de segments*

- 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. Notre objectif est d'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 procédons en plusieurs étapes : dans un premiers temps, nous recherchons des relations topologiques simples (voisinage, coplanarité, parallélisme) entre les deux segments $S_{t'}$ et S_t ainsi que la confiance que l'on accorde à cette relation (c'est-à-dire par exemple la probabilité que deux segments soient voisins). Les erreurs intervenant dans l'estimation des paramètres des segment 3D sont en général faibles. Elles ne sont cependant pas totalement négligeables. On considère donc ces mesures comme des variables aléatoires qui suivent une loi Gaussienne. Par exemple, dans le cas de la propriété de voisinage, $p(S_{t'} \text{ voisin de } S_t)$ doit être maximale quand $d(S_{t'}, S_t) = 0$ où d est la distance Euclidienne des deux points extrémités les plus proches. Cette probabilité doit décroître quand la distance augmente. On peut aussi modéliser la probabilité que deux segments soit parallèles (en fonction de la valeur du produit scalaire de leur vecteur directeur) ou que deux segments soit coplanaires (en fonction de la distance maximale d'un des points extrémités au plan formé par les trois autres).

L'étape suivante consiste à définir cinq classes distinctes auxquelles peuvent appartenir ces segments, chacune de ces classes représentant une combinaison particulière des différentes relations élémentaires précédemment définies. Enfin, à partir de ces classes, nous émettons une hypothèse sur la nature de la

relation associée aux segments. Ce schéma de raisonnement est codé dans la structure du réseau bayésien de la figure 6. Celui-ci, relativement simple, comporte six nœuds principaux, chacun étant associé à une étape du raisonnement. Un nœud est associé à chacune des relations topologiques 3D élémentaires : *voisinage*, *coplanarité* et *parallélisme*. Un autre nœud est associé à la classe à laquelle appartient le couple de segments. Enfin, deux nœuds sont associés aux hypothèses proposées (H_1, H_2, H_3, H_4). Les liens entre ces différents nœuds définissent les relations de causalité entre les différentes étapes du raisonnement et donc l'ordre dans lequel celui-ci doit s'effectuer.

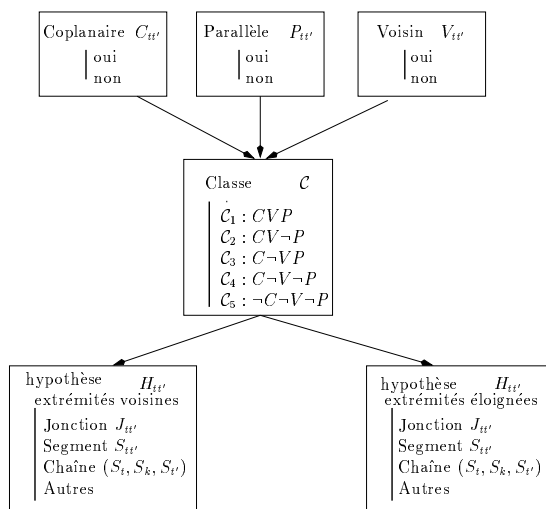


Figure 6. Réseaux de génération d'hypothèses

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 7) :

- \mathcal{C}_1 : les segments sont coplanaires, voisins et parallèles ;
- \mathcal{C}_2 : les segments sont coplanaires, voisins et non parallèles ;
- \mathcal{C}_3 : les segments sont coplanaires, non voisins et parallèles ;
- \mathcal{C}_4 : les segments sont coplanaires, non voisins et non parallèles ;
- \mathcal{C}_5 : les segments ne sont ni coplanaires, ni voisins, ni parallèles.

Enfin, à partir de la classe à laquelle appartient le couple de segments, il convient de déterminer l'hypothèse à adopter concernant la nature des relations entre ces deux segments. Nous allons donc définir des stratégies de décision qui

détermineront la meilleure hypothèse possible, compte tenu des connaissances disponibles. Ces stratégies sont codées dans les tables de probabilités conditionnelles $P(\mathcal{H}|\mathcal{C})$ où \mathcal{H} est l'hypothèse et \mathcal{C} la classe.

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. Notons qu'en aucun cas, elles ne s'opposent aux hypothèses qui ont été initialement réalisées sur le type de scènes que nous avons décidé d'étudier. 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. Cependant, elles ne sont, bien entendu, pas toujours vérifiées. Elles peuvent cependant 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 \mathcal{C}_2 et les hypothèses associées sont décrites sur la deuxième ligne de la figure 7). 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 :

$$\begin{aligned}
 P(H = \textit{fusion} | \mathcal{C}_2) &= 0.025 \\
 P(H = \textit{jonction} | \mathcal{C}_2) &= 0.6 \\
 P(H = \textit{chaîne} | \mathcal{C}_2) &= 0.3 \\
 P(H = \textit{autre} | \mathcal{C}_2) &= 0.075
 \end{aligned}$$

Ces tables de probabilités conditionnelles sont définies de manière empirique. Cependant, fixer précisément les valeurs de chaque probabilité conditionnelle n'est pas fondamental. Elles doivent juste refléter de manière grossière la connaissance que l'on souhaite transmettre au système. Il est cependant possible de les déterminer à partir d'une analyse statistique de données réelles.

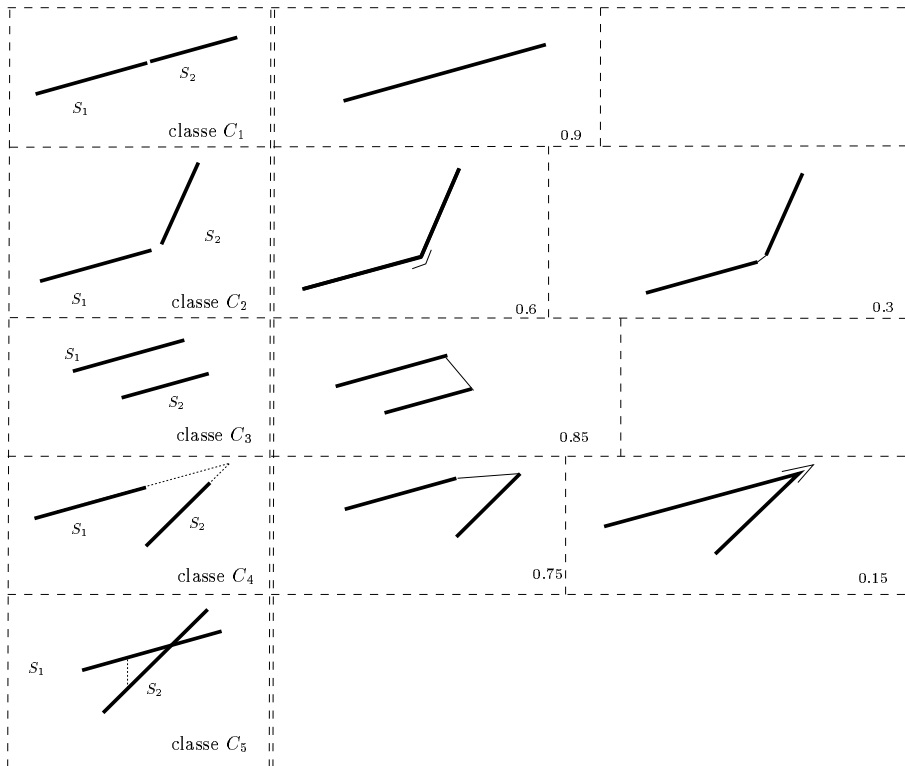


Figure 7. Classes élémentaires et hypothèses associées : extrémités proches

Les stratégies résultant de ce processus dépendront alors beaucoup plus du jeu d'apprentissage que des stratégies que l'on souhaite voir exécuter. Un ensemble d'apprentissage mal déterminé entraînera alors un mauvais comportement du système. Une telle technique d'apprentissage a cependant été utilisée avec succès dans [DJI 96b].

Les tableaux 1 et 2 montrent les différentes probabilités conditionnelles associées aux hypothèses pour les extrémités proches et éloignées des segments considérés. 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 par [DJI 96a]. Les deux hypothèses sélectionnées font l'objet d'une vérification, étape que nous décrivons à présent.

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

Tableau 1. Table de probabilité conditionnelle $P(\text{hypothèse} \mid \text{classe})$ pour les extrémités proches (figure 7)

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

Tableau 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 8. 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

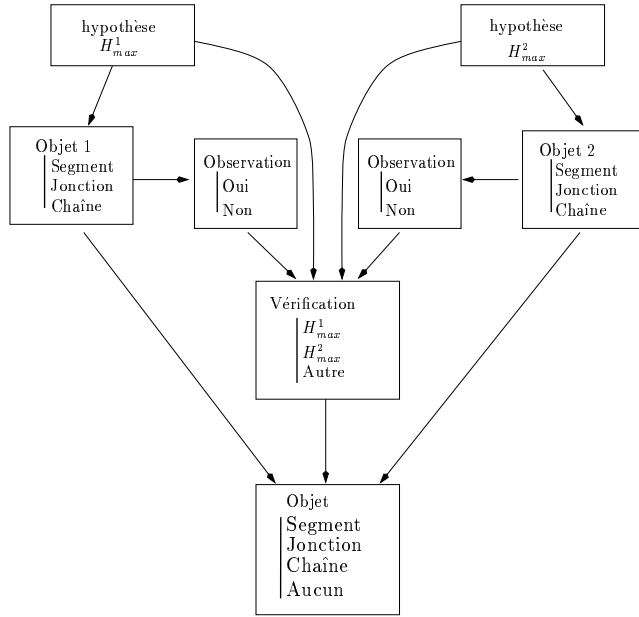


Figure 8. Réseau de vérification d'hypothèses

la caméra (c'est-à-dire appartenant à la trajectoire \mathcal{T}_0^t), 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. En effet, l'hypothèse de la présence d'une jonction découle inéluctablement de la présence de deux segments dont l'existence a déjà été validée. 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 de la trajectoire \mathcal{T}_0^t . Il convient alors de déplacer la caméra vers un nouveau point de vue 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 Ko-

vesi [COW 88], Tarabanis [TAR 95] ou Seales et Dyer [SEA 92]). La méthode que nous proposons est proche de celle décrite par Cowan [COW 88]. Celle-ci consiste à déterminer le lieu géométrique des points depuis lesquels la primitive est visible tout en respectant un certain nombre de contraintes. Cette méthode a l'avantage d'être simple, mais les régions de l'espace qui sont solution de ce problème de calcul de points de vue peuvent être très vastes et le choix d'un point de vue parmi l'ensemble des solutions possibles nécessite l'introduction de contraintes supplémentaires qui rendent le calcul de points de vue long et coûteux.

Plutôt que de calculer explicitement ce point de vue, puis de rechercher *off-line* le segment supposé dans l'image, nous avons décidé de faire pivoter la caméra autour d'un des segments appartenant à la fois au polygone d'occlusion et au plan dans lequel se situe le segment recherché (voir figure 9). Pendant ce mouvement de rotation, généré par asservissement visuel, nous recherchons *on-line* l'apparition dans l'image du segment supposé. Comme nous l'avons déjà dit précédemment, il n'est pas possible de réaliser une acquisition complète de l'ensemble des segments 2D présents dans l'image pendant les phases d'asservissement visuel en raison des contraintes de temps réel. La détection de l'apparition éventuelle de ce segment pendant le mouvement de la caméra est cependant impérative. Pour résoudre ce problème, nous disposons un observateur à une position adéquate dans l'image [DJI 96a]. 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. Pour cela, on utilise l'algorithme des ECM (éléments de contours en mouvement) [BOU 89].

L'exemple décrit sur la figure 9 illustre la stratégie retenue. On considère ici un objet polyédrique dont deux faces avec les arêtes associées ont déjà été reconstruites. Lors de la génération des hypothèses, la présence d'un segment a été prédite dans le plan formé par S_1 et S_2 . La stratégie retenue consiste à se focaliser sur l'un de ces deux segments puis à effectuer un mouvement de rotation autour de ce segment jusqu'à ce qu'un nouveau segment S apparaisse dans l'image. Si une butée articulaire du manipulateur est rencontrée, l'hypothèse ne sera pas validée. Ici, S_1 a été retenu car S_2 n'est pas assez long pour servir de base à la génération d'un mouvement de rotation par asservissement visuel.

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

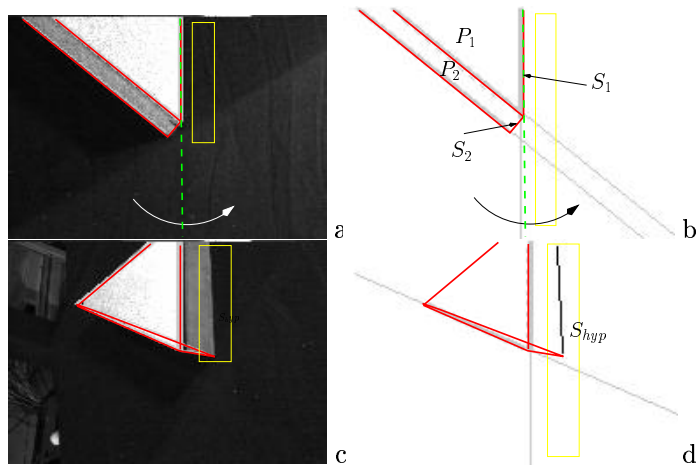


Figure 9. Vérification d'une hypothèse : (a-b) vue de la scène depuis la position initiale de la caméra, (c-d) vue de la scène depuis la position finale (détection du segment S_{hyp} après un mouvement de pivot autour de l'axe du segment S_1)

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 10 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 objets². 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.

Cette dernière étape de modélisation nous permet d'obtenir une représentation de plus haut niveau de la scène considérée. La représentation initiale était uniquement composée de segments 3D. La représentation finale est composée de ces mêmes segments, mais aussi d'informations complémentaires comme les jonctions 3D, les polygones et les faces des objets.

2. Dans le réseau de la figure 10, 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).

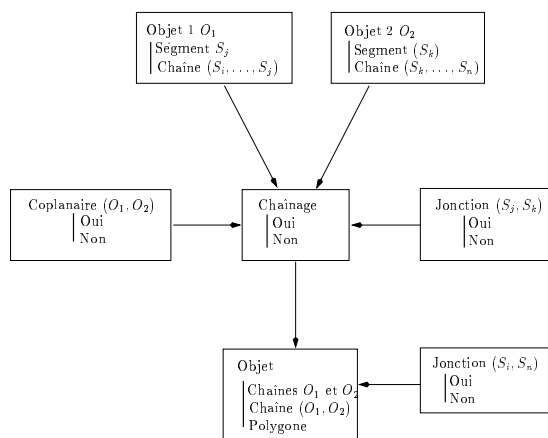


Figure 10. Réseau de modélisation

4. Exploration globale : calcul de points de vue

Si, après une phase de reconstruction incrémentale de la scène, il est possible d'assurer que toutes les primitives observées par la caméra ont été reconstruites, il est par contre impossible d'affirmer, à ce stade, que tous les objets composant la scène ont été traités. Compléter le modèle géométrique requiert des mouvements exploratoires permettant d'observer les parties non encore observées de la scène. Cette exploration passe par le calcul d'un certain nombre de points de vue qui permettront l'observation et ensuite la reconstruction de nouveaux objets, ou au contraire qui permettront d'assurer que telle ou telle partie de l'espace est vide. La description de ce processus n'est pas l'objet de cet article. Il nous semble cependant important d'en donner un rapide aperçu.

Le problème du calcul de points de vue est un problème difficile et peu étudié. Généralement, les solutions proposées reposent sur une connaissance *a priori* de la scène [TAR 95, COW 88, TRI 95], alors que, dans notre cas, ces connaissances sont obtenues au fur et à mesure de l'exploration. La stratégie de calcul de points de vue que nous avons mise en œuvre repose sur la modélisation et l'optimisation d'une fonction de coût codant au mieux la tâche que nous souhaitons effectuer. Nous avons retenu quatre critères associés au point de vue, qui sont intégrés dans cette fonction. En premier lieu, nous nous basons sur le gain apporté par une nouvelle position. Ce gain détermine le volume potentiel découvert calculé en utilisant les techniques de lancers de rayons. Un critère modélisant le coût du déplacement d'un point de vue au suivant est aussi justifié par le fait que nous souhaitons minimiser la distance totale parcourue par la caméra. Les contraintes mécaniques du robot conduisent également à l'introduction d'un critère éloignant le robot de ses butées articulaires. Enfin, les connaissances déjà acquises sur la scène permettent de définir un critère binaire représentant l'accessibilité d'un point de vue. La minimisation d'une

fonction intégrant ces différents critères permet le calcul d'un nouveau point de vue [MAR 97b, MAR 96b]. Cette minimisation est effectuée à l'aide d'un ICM multi-échelle préférable à l'emploi d'une méthode stochastique de type recuit simulé, notamment pour des raisons de temps de calcul. Initialement, la caméra se déplace sur une demi-sphère englobant la scène pour éviter les obstacles potentiels (encore inconnus). Dès que des zones accessibles sont reconstruites, la caméra peut se déplacer à l'intérieur de ces zones situées dans la demi-sphère. Un algorithme de focalisation sur les zones inconnues résiduelles a également été proposé [MAR 97a]; il permet de traiter de manière adéquate les parties occultées par des objets. Ajoutons finalement que nous avons proposé des techniques issues de la programmation dynamique qui permettent de minimiser (sous certaines hypothèses) le nombre de points de vue [MAR 97a].

L'exploration de la scène s'achève quand, quel que soit le point de vue choisi parmi tous les points de vue accessibles, il n'y a plus d'apport supplémentaire d'informations. Ceci signifie que la reconstruction est alors aussi complète que possible compte tenu des contraintes imposées par le manipulateur et/ou par la scène.

5. Résultats expérimentaux

5.1. *Un exemple détaillé de l'utilisation de l'algorithme de prédiction/vérification d'hypothèses*

Nous présentons ici les résultats de la reconstruction d'un polyèdre (voir figure 4). Celui-ci est constitué de six polygones non coplanaires. Cette scène 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'occlusions, caractéristiques géométriques peu visibles (par exemple, l'angle aigu est coupé (voir figure 4.b), ce qui induit la présence d'un segment d'une longueur de 1 cm environ).

La figure 11.a montre la première image de cet objet et les figures 11.b à 11.f décrivent les itérations suivantes de la reconstruction incrémentale. Les lignes en pointillés correspondent aux primitives déjà reconstruites; les autres correspondent à des segments non encore traités. Les flèches dans les images pointent sur le segment \mathcal{S}_i à reconstruire. À chaque fois, les paramètres de la primitive associée à \mathcal{S}_i sont estimés en utilisant la méthode de reconstruction 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 12.

Quelques cas simples. 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 pre-

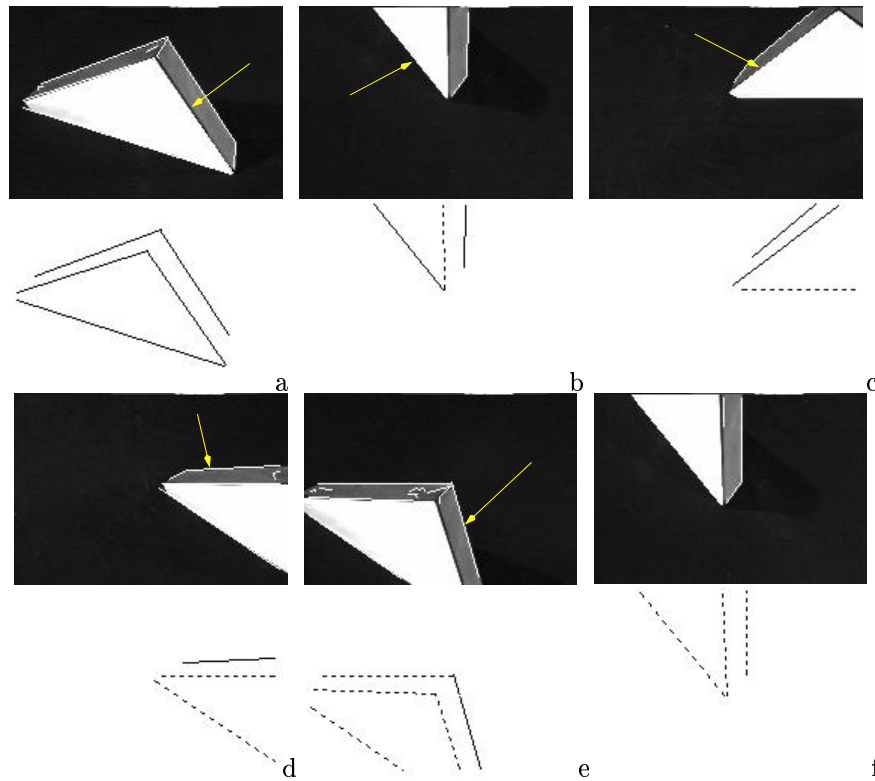


Figure 11. *Reconstruction d'un polyèdre. La flèche pointe sur le prochain segment à reconstruire*

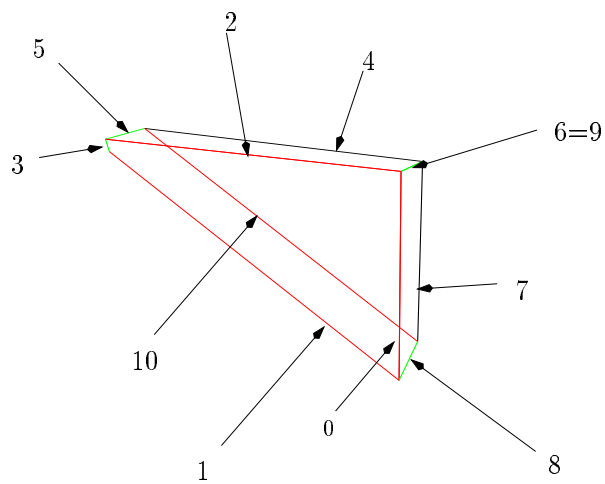


Figure 12. *Scène reconstruite et numérotation des segments dans l'ordre de leur introduction dans la carte de l'environnement*

miers 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 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 . Pour vérifier ces hypothèses, le deuxième réseau bayésien est utilisé.

Les observations faites dans la séquence d’images précédemment acquise 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 a pris soin de vérifier que ce segment, s’il doit exister, est observable depuis les positions déjà occupées par la caméra). Enfin, compte tenu de la probabilité de chacune des deux hypothèses réalisées et des observations, on en déduit la probabilité de la présence d’un objet jonction à 85 % et de la présence d’une chaîne de segments à 9 %. La probabilité de l’absence de toute relation est elle égale à 6 %. 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 %.

En accord avec les hypothèses émises et à partir des relations pouvant exister entre les extrémités éloignées de S_0 et S_1 , un segment est supposé exister entre eux. 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 %.

Intérêt des hypothèses multiples. 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 \mathcal{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 l'observation des différentes images montre 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.

Création de polygones. 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. Au niveau du réseau de modélisation, la probabilité d'un chaînage de ce segment avec la chaîne existante (S_1, S_0, S_2) est donc très forte (grâce à la jonction (S_3, S_1) et au fait que le segment S_3 soit coplanaire à la chaîne (S_1, S_0, S_2)). L'existence très probable d'une jonction entre S_3 et S_2 permet finalement 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 iden-

tiques et doivent être fusionnés. Après vérification, cette fusion est réalisée et le segment S_9 est supprimé de la carte.

Actions de vérification. La reconstruction du segment S_7 a aussi pour conséquence la création d'une jonction avec S_4 . De plus, dans le cas de l'observation d'un couple de segments similaires au couple (S_7, S_4) , une seconde hypothèse correspondant à la présence d'un segment reliant leurs extrémités éloignées est aussi émise (voir tableau 2). 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 13.ab). Le segment découvert est ensuite reconstruit (voir figure 13.c) et inséré dans la carte de l'environnement.

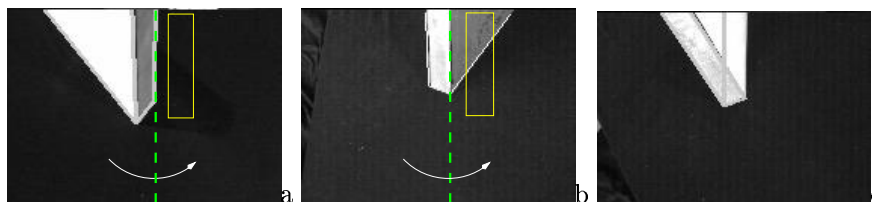


Figure 13. Vérification d'une hypothèse : (a) rotation autour de S_7 , (b) S_{10} est découvert et (c) reconstruit

Signalons que les six images présentées sur la figure 11 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 les listes de segments ω_ϕ et $\Omega_{\mathcal{T}}$. 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. Le traitement d'images consiste alors à suivre la position du segment sélectionné dans la séquence d'images par la méthode des ECM [BOU 89]. 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.

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.

5.2. Exploration: complétude de la reconstruction

L'objet de la scène précédente a pu être reconstruit entièrement sans faire appel à l'algorithme d'exploration succinctement décrit dans la Section 4. Cette étape est pourtant fondamentale afin d'assurer qu'il n'y a pas d'autres objets. Les résultats suivants (figure 14) illustrent ce cas de figure.

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 qu'une vue très partielle (voir figure 14.b), puisque des objets disjoints de ceux reconstruits n'ont jamais été observés. En greffant à la reconstruction incrémentale le module d'exploration globale évoqué en section 4, on aboutit alors à la complétude de la reconstruction (voir figure 14.d). La figure 15 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 15.a correspond à la situation obtenue à l'issue de la première reconstruction incrémentale. Les premiers déplacements de la caméra (voir figure 15.b) permettent ensuite de diminuer la zone de la scène non encore explorée. À la position représentée figure 15.c, un nouvel objet est détecté marquant le début d'une seconde phase de reconstruction incrémentale qui se termine à l'étape décrite figure 15.d. Une nouvelle exploration globale est finalement nécessaire pour observer un segment appartenant au dernier objet (une carte de téléphone). Après sa reconstruction, 97 % de l'espace a été observé, assurant ainsi une exploration complète de la scène.

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.

5.3. Discussion

Ces résultats nous permettent 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 et 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 fâcheuses : si une relation existe entre deux segments éloignés,

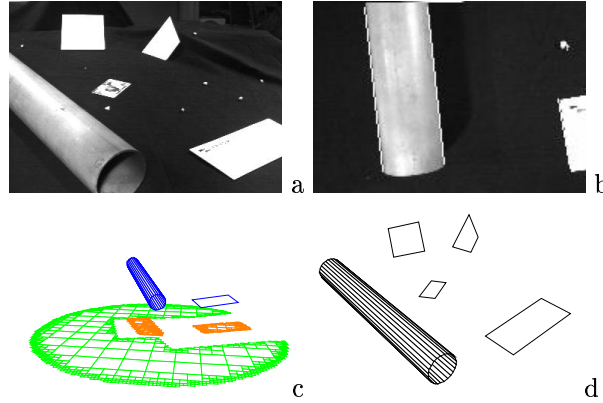


Figure 14. Autre scène reconstruite : (a) vue extérieure de la scène, (b) image initiale, (c) scène reconstruite après la première phase de reconstruction incrémentale et projection sur un plan virtuel des zones inconnues, (d) scène reconstruite après l'exploration globale

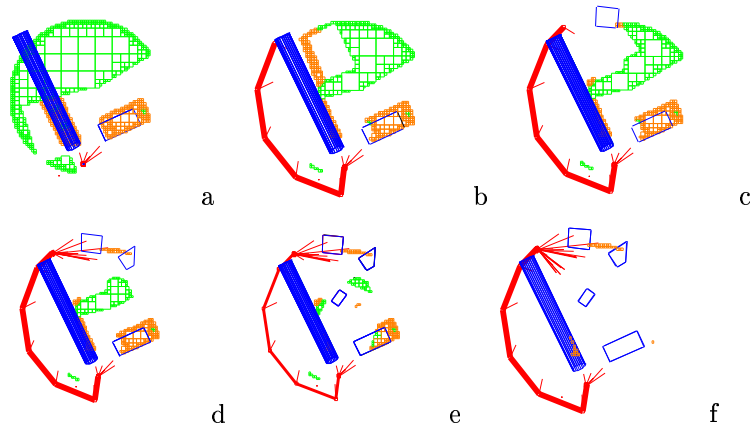


Figure 15. Différentes étapes de l'exploration globale : trajectoire de la caméra, modèle courant de la scène reconstruite et projection sur un plan virtuel des zones inconnues

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 en considérant 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.

Il faut par ailleurs 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. L'utilisation future d'algorithmes de traitement d'images plus robustes aux perturbations est bien sûr une de nos perspectives.

6. Conclusion

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.

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 prédic-

tion/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, projet commun aux PRC-IA et PRC-CHM) et a reçu le soutien du MESR (Ministère de l'enseignement supérieur et de la recherche). 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.

7. Bibliographie

- [AGO 88] J.M. AGOSTA. « The structure of Bayes nets for vision recognition ». *Proc. of 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, Minnesota, août 1988.
- [ALO 87] Y. ALOIMONOS, I. WEISS, et A. BANDOPADHAY. « Active Vision ». *International Journal of Computer Vision*, 1(4):333–356, janvier 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, juin 1990.
- [AYA 86] N. AYACHE et 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.
- [AYA 87] N. AYACHE et B. FAVERJON. « Efficient registration of stereo images by matching graph descriptions of edge segments. ». *International Journal of Computer Vision*, 1(2):107–132, 1987.
- [BAJ 88] R. BAJCSY. « Active Perception ». *Proc. of the IEEE*, 76(8):996–1005, août 1988.
- [BAL 91] D.H. BALLARD. « Animate vision ». *Artificial Intelligence*, 48(1):57–86, février 1991.
- [BOL 86] R.C. BOLLES et R. HORAUD. « 3DPO: A three-dimensional part orientation system ». *International Journal of Robotics Research*, 5(3):3–26, juin 1986.
- [BOU 89] P. BOUTHEMY. « A Maximum Likelihood Framework for Determining Moving Edges ». *IEEE Trans. on PAMI*, 11(5):499–511, mai 1989.

- [BOU 93] S. BOUKIR. Reconstruction 3D d'un environnement statique par vision active. thèse de doctorat, université de Rennes 1, IRISA, octobre 1993.
- [BUX 95] H. BUXTON et S. GONG. « Visual Surveillance in a dynamic and uncertain world ». *Artificial Intelligence*, 78(1-2):431–459, octobre 1995.
- [CHA 96] F. CHAUMETTE, S. BOUKIR, P. BOUTHEMY, et D. JUVIN. « Structure from controlled motion ». *IEEE Trans. on PAMI*, 18(5):492–504, mai 1996.
- [COW 88] C.K. COWAN et P.D. KOVESI. « Automatic Sensor Placement from Vision task Requirements ». *IEEE Trans. on PAMI*, 10(3):407–416, mai 1988.
- [DJI 96a] D. DJIAN, P. PROBERT, et P. RIVES. « Reconnaissance de modèles géométriques simples à l'aide de réseaux bayésiens ». *10ème Congrès AFCET/AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA'96*, 1:396–404, Rennes, France, janvier 1996.
- [DJI 96b] D. DJIAN, P. PROBERT, et P. RIVES. « Training Bayes Nets for Model-based Recognition ». *Int. Conf. on Automation, Robotics and Computer Vision, ICARCV'96*, Singapore, 1996.
- [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, juin 1992.
- [GIR 97] G. GIRAUDON et F. SANDAKLY. « Interprétation de scènes d'intérieur pour un robot mobile. ». *Technique et Science Informatiques, TSI*, 16(1):69–100, janvier 1997.
- [HER 86] M. HERMAN et T. KANADE. « Incremental Reconstruction of 3D Scenes from Multiple, Complex Images ». *Artificial Intelligence*, 30(3):289–341, décembre 1986.
- [HUT 89] S.A HUTCHINSON et 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.
- [JEN 91] F.V. JENSEN, H. CHRISTENSEN, et J. NIELSEN. « Bayesian methods for interpretation and control in multi-agent vision systems ». Rapport Technique LIA 91-5, Aalborg University, Suède, octobre 1991. *Artificial Intelligence X: Machine Vision and Robotics*, SPIE Proceedings Serie, No1708.
- [KRA 93] P. KRAUSE et D. CLARK. *Representing uncertain knowledge: artificial intelligence approach*. Kluwer Academic Publishers, Oxford, Royaume Uni, 1993.
- [KUM 96] V.P. KUMAR et U.B. DESAI. « Image Interpretation using bayesian networks ». *IEEE Trans. on PAMI*, 18(1):74–77, janvier 1996.
- [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, juin 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, juin 1996.

- [MAR 97a] E. MARCHAND et F. CHAUMETTE. « Active sensor placement for complete scene reconstruction and exploration ». *IEEE Int. Conf. on Robotics and Automation*, 1:743–751, Albuquerque, Nouveau Mexique, avril 1997.
- [MAR 97b] E. MARCHAND et F. CHAUMETTE. « Contrôle actif d'une caméra pour la reconstruction et l'exploration de scènes ». *Technique et Science Informatiques, TSI*, 16(1):101–129, janvier 1997.
- [NEA 90] R. NEAPOLITAN. *Probabilistic Reasoning in Expert Systems*. John Wiley, Chichester, 1990.
- [PEA 88] J. PEARL. *Probabilistic reasoning in intelligent systems : Networks of plausible inference*. Morgan Kaufmann Publisher Inc., San Mateo, Californie, 1988.
- [RIM 91] R.D. RIMEY et C. BROWN. « Controlling Eye Movements with Hidden Markov Models ». *International Journal of Computer Vision*, 7(1):47–65, janvier 1991.
- [RIM 94] R.D. RIMEY et C. BROWN. « Control of Selective Perception Using Bayes Nets and Decision Theory ». *International Journal of Computer Vision*, 12(2/3):173–207, avril 1994.
- [SAR 93] S. SARKAR et K. BOYER. « Integration, Inference, and Management of spatial information using Bayesian networks: perceptual organization ». *IEEE Trans. on PAMI*, 15(3):256–274, mars 1993.
- [SEA 92] W. SEALES et C. DYER. « Viewpoint from occluding contour ». *Computer Vision, Graphics, and Image Processing: Image Understanding*, 55(2):198–211, mars 1992.
- [SHA 76] G. SHAFER. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.
- [SPI 90] D. SPIEGELHALTER et S. LAURITZEN « Sequential updating of conditionnal probabilities on directed graphical structures ». *NETWORKS*, volume 20, pages 579–605, 1990.
- [SWA 93] M.J. SWAIN et M.A. STRICKER. « Promising Direction in Active Vision ». *International Journal of Computer Vision*, 11(2):109–127, octobre 1993.
- [TAR 95] 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, février 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, mai 1995.