

Fusion d'informations visuelles pour la localisation d'objets complexes

Grégory Flandin

François Chaumette

IRISA - INRIA Rennes

Campus de Beaulieu,

35042 Rennes Cedex

Gregory.Flandin@irisa.fr Francois.Chaumette@irisa.fr

Résumé

Les capteurs visuels fournissent une connaissance partielle et incertaine de la scène observée. Dans cet article, nous présentons un modèle de représentation de la connaissance qui permet la comparaison et la fusion de données incertaines et partielles. Il est basé sur un mélange de modèles stochastique et à erreur bornée. En considérant une hypothèse gaussienne des incertitudes et une approximation ellipsoïdale des supports de l'erreur, nous développons un processus d'estimation en ligne efficace intégrant les données visuelles. Enfin, nous élaborons un processus d'exploration optimale en temps réel basé sur la minimisation de l'incertitude de localisation de l'objet observé.

Mots-clés : vision active, fusion de données, exploration autonome

1 Introduction

Un défi majeur des systèmes visuels autonomes est leur capacité à modéliser l'environnement. Souvent la scène est représentée par des primitives géométriques ou au moyen d'une carte exhaustive des voxels la composant. Or une large classe d'applications peut se contenter d'une carte simplifiée. Nos travaux abordent le problème de l'estimation de modèles simplifiés ou *coarse models* [13, 5]. Nous nous intéressons en particulier à la définition de modèles simplifiés englobants. Cette notion étant fortement liée à la tâche assignée, nous devons introduire un nouveau concept concernant le type d'objets traité. Il n'est pas restrictif de considérer que toute scène peut se décomposer en groupes ou parties d'objets qui soient cohérents dans le sens où chaque groupe ou partie peut être décrit par un simple volume englobant de façon suffisamment significative pour la tâche considérée. Typiquement, un ensemble d'objets proches est cohérent s'il est observé d'une distance suffisamment grande mais devient incohérent s'il est proche de l'observateur auquel cas chaque objet doit être traité indépendamment. De même, un objet isolé peut nécessiter d'être découpé en parties cohérentes s'il apparaît trop fortement concave. Par abus de langage, nous appellerons "objet" tout groupe ou partie d'objets physiques qui soit cohérent. Cet article est dédié à la description, à l'estimation et au raffinement de volumes englobant les objets. Ces volumes sont définis par leur enveloppe ellipsoïdale. La phase de traitement d'images revient ainsi à extraire l'ellipse englobant le masque de l'objet dans l'image. Nos travaux peuvent

s'appliquer, en théorie, à tout type d'objet dans la mesure où l'on est capable de le segmenter. Pour n'imposer aucune contrainte concernant l'aspect des objets, nous faisons l'hypothèse que la frontière des objets est marquée par une discontinuité de profondeur. Ainsi un algorithme de segmentation au sens du mouvement nous donnera le masque des objets.

La première partie de nos travaux s'attache à décrire le modèle simplifié englobant choisi. Les méthodes que nous développons s'insèrent dans le cadre des techniques d'estimation. Celles-ci supposent généralement une description probabiliste des observations. La géométrie probabiliste [1, 3, 4] permet alors de les fusionner. Mais lorsque l'incertitude n'est pas intrinsèquement stochastique, une alternative intéressante est de considérer un modèle à erreur bornée [18, 15]. Fusionner les observations nécessite alors le calcul d'intersections de supports parfois complexes. Une méthode efficace consiste à approcher tous les supports par leur ellipsoïde englobant [11]. Dans cet article, nous tirons parti des deux modèles précédents au sein d'une représentation mixte des objets (voir [8] pour d'autres travaux concernant la combinaison de théories statistiques et ensemblistes).

Un autre problème concernant la modélisation de l'environnement est la génération autonome des mouvements exploratoires. Dans le contexte d'une scène initialement inconnue, les travaux ont adopté des points de vue différents. Dans [2], Connolly s'intéresse à la détermination des points de vue d'une image télémétrique qui maximisent le volume découvert. Whaitte et Ferries [17] modélisent les objets par des superquadriques. La stratégie d'exploration d'un télémètre laser est définie par minimisation de l'incertitude associée. Kutulakos, Dyer et Lumelsky [9] exploitent les variations de la frontière d'occultation. Lacroix et Chatila [10] ont développé des stratégies de perception et d'action utilisant soit un télémètre soit une paire d'images stéréo. Un algorithme de recherche fournit un chemin optimal au sein d'un graphe puis ce chemin est analysé pour définir la stratégie de perception associée. Marchand et Chaumette [12] utilisent la vision monoculaire active pour explorer des objets géométriques tels que cylindres ou polygones. La sélection du point de vue provient de la minimisation d'une fonction de coût.

La stratégie que nous développons consiste à réduire l'incertitude de la distribution associée à l'objet observé en intégrant des données visuelles. La modélisation gaussienne de l'incertitude ainsi qu'une linéarisation adéquate du processus d'acquisition visuelle nous permettent de construire une solution analytique au problème d'exploration optimale. En section 2, nous décrivons précisément le mélange de modèles utilisé. En section 3 nous développons les règles de propagation associées et les appliquons à la propagation de données visuelles. En section 4, nous décrivons un processus d'estimation qui tire parti du mouvement de la caméra pour reconstruire son environnement. Enfin, dans le contexte de l'exploration autonome, nous définissons le mouvement optimal de la caméra en section 5.

2 Modélisation

Pour tout objet \mathcal{O} d'une scène \mathcal{S} et pour tout point $x \in \mathcal{S}$, nous souhaitons calculer la probabilité que $x \in \mathcal{O}$ notée $\mathcal{P}(x \in \mathcal{O})$. Si nous considérons le vecteur aléatoire des coordonnées d'un point $c \in \mathcal{O}$ de distribution $\mathcal{P}_c(x)$ alors :

$$\mathcal{P}_c(x) = \mathcal{P}_c(x|x \in \mathcal{O}) \cdot \mathcal{P}(x \in \mathcal{O}) + \mathcal{P}_c(x|x \notin \mathcal{O}) \cdot \mathcal{P}(x \notin \mathcal{O}) \quad (1)$$

où $\mathcal{P}_c(x|x \notin \mathcal{O})$ est la probabilité que le point $c \in \mathcal{O}$ soit en x sachant que $x \notin \mathcal{O}$ et vaut donc 0; $\mathcal{P}_c(x|x \in \mathcal{O})$ est la probabilité que le point $c \in \mathcal{O}$ soit en x sachant que $x \in \mathcal{O}$,

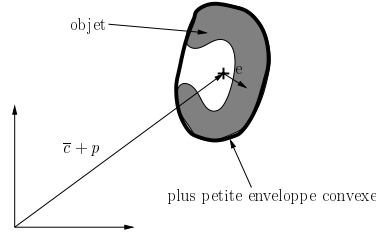


FIG. 1 – *Modèle additif.*

que nous modélisons naturellement par une loi uniforme dont la valeur peut être calculée après normalisation et pour tout \mathcal{O} . On obtient $\mathcal{P}_c(x|x \in \mathcal{O}) = 1/\text{Volume}(\mathcal{O})$. En résumé :

$$\mathcal{P}_c(x) = \frac{\mathcal{P}(x \in \mathcal{O})}{\text{Volume}(\mathcal{O})} \quad (2)$$

Modéliser \mathcal{S} revient donc à chercher pour tout \mathcal{O} une distribution adéquate de $\mathcal{P}_c(x)$. Dans ce but, nous décomposons c en la somme d’une constante \bar{c} et de deux vecteurs aléatoires indépendants (voir figure 1) :

$$c = \bar{c} + p + e \quad (3)$$

où p représente l’**incertitude** de localisation de l’objet et le support de l’**erreur** e définit son volume. Comme précisé en introduction, nous faisons les hypothèses suivantes : p suit une loi normale centrée de covariance P que nous noterons $\mathcal{N}(0, \Sigma)$ où $\Sigma = P^{-1}$ est la matrice d’information. Son utilisation est rendue nécessaire par le caractère partiel et donc dégénéré des observations visuelles. Notons que l’approximation gaussienne est satisfaisante pour de nombreuses sources d’incertitudes telles que la localisation de la caméra ou les erreurs de calibrage. D’autre part e est uniformément distribuée sur un ellipsoïde (assimilé à sa matrice E).

La distribution associée à un objet est donc entièrement déterminée par \bar{c} , Σ et E . Plus précisément, c’est la distribution de la somme de deux variables aléatoires indépendantes, c’est-à-dire le produit de convolution entre la loi uniforme \mathcal{U}_E et la loi normale $\mathcal{N}(\bar{c}, \Sigma)$. Nous appelons cette distribution “loi d’ensemble” et la notons :

$$\mathcal{E}(\bar{c}, \Sigma, E) = \mathcal{N}(\bar{c}, \Sigma) * \mathcal{U}_E \quad (4)$$

En résumé, d’après l’équation (2) et pour tout objet \mathcal{O} , $\mathcal{P}(x \in \mathcal{O})$ est proportionnel à une loi d’ensemble $\mathcal{E}(\bar{c}, \Sigma, E)$. Les travaux que nous présentons dans la suite concernent l’estimation des paramètres (\bar{c}, Σ, E) la constituant.

Pour appliquer ce modèle au cas particulier des données visuelles, nous considérons trois étapes dans la chaîne des observations visuelles (figure 2) :

1. Dans l’image, la mesure est une loi d’ensemble 2D : $\mathcal{E}^i(\bar{c}^i, \Sigma^i, E^i)$ où \bar{c}^i et E^i représentent le centre et la matrice de la plus petite ellipse englobant la projection de l’objet. Nous verrons en section 6 comment extraire cette segmentation en pratique. Quant à \bar{c}^i et E^i , ils s’obtiennent grâce à des algorithmes tels que celui proposé dans [16]. L’information Σ^i doit, pour sa part, prendre en compte toutes les sources d’incertitudes inhérentes au calcul de l’ellipse : erreur sur les paramètres intrinsèques et imprécision du traitement d’images.
2. L’opération transformant les mesures 2D en observations 3D est appelée “rétro-projection”. La loi d’ensemble 3D associée est notée $\mathcal{E}^c(\bar{c}^c, \Sigma^c, E^c)$.

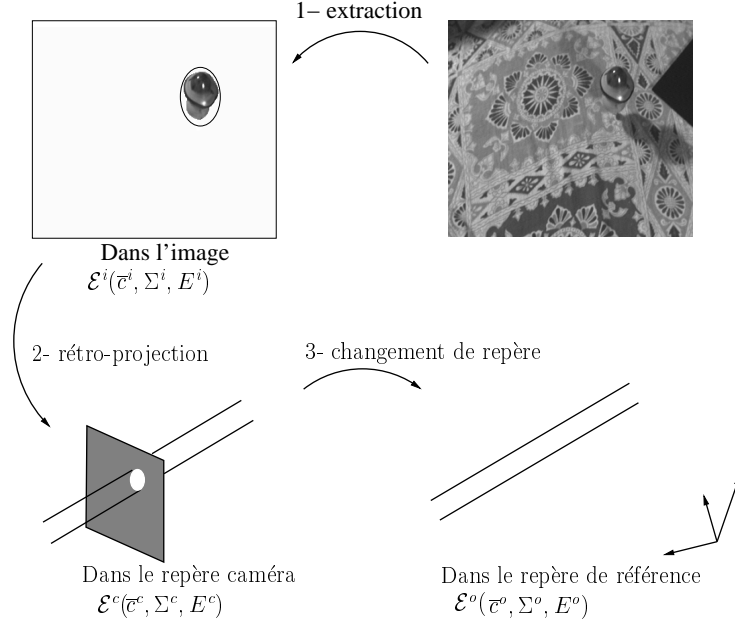


FIG. 2 – Evolution des données visuelles.

3. Enfin, nous devons exprimer chaque observation dans un repère commun de référence \mathcal{R} .
La loi d'ensemble associée est notée $\mathcal{E}^o(\bar{c}^o, \Sigma^o, E^o)$.

Ces étapes sont indispensables pour comparer et fusionner les informations visuelles mais elles requièrent la définition de règles de propagation des lois d'ensembles.

3 Règles de propagation

Nous définissons à présent les transformations induites par un changement d'espace des paramètres.

Règle 1 (transformation d'une loi d'ensemble)

Soit c un vecteur aléatoire de loi $\mathcal{E}(\bar{c}, \Sigma, E)$. Si l'on note $J = \left. \frac{\partial T^{-1}}{\partial c} \right|_{\bar{c}}$ le jacobien de la transformation inverse, en première approximation, le vecteur aléatoire $c' = T(c)$ suit une loi $\mathcal{E}'(\bar{c}', \Sigma', E')$ où :

$$\begin{cases} \bar{c}' &= T(\bar{c}) \\ \Sigma' &= J^T \Sigma J \\ E' &= J^T E J \end{cases} \quad (5)$$

La démonstration découle d'une approximation de T au premier ordre :

$$c' = T(c) \approx T(\bar{c}) + \left. \frac{\partial T}{\partial c} \right|_{\bar{c}} (c - \bar{c}) = T(\bar{c}) + \left. \frac{\partial T}{\partial c} \right|_{\bar{c}} p + \left. \frac{\partial T}{\partial c} \right|_{\bar{c}} e \quad (6)$$

et par application de résultats sur le calcul gaussien et le calcul ellipsoïdal [6]. Lorsque T dépend également de paramètres externes : $c' = T(c, p_e)$ (où p_e est $\mathcal{N}(\bar{p}_e, P_e)$), il est possible de linéariser T autour de \bar{c} et \bar{p}_e afin de prendre en compte d'autres sources d'incertitude. Mais, dans tous les

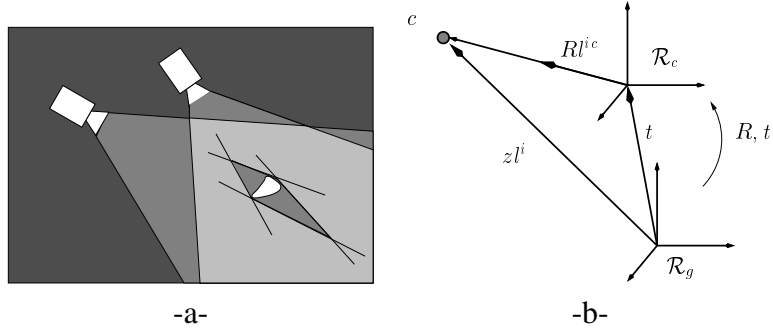


FIG. 3 – Initialisation de la distribution avec deux images.

cas, l'existence de T^{-1} requiert implicitement que T soit un difféomorphisme. En particulier, l'espace de définition et l'espace image de T doivent avoir la même dimension. Si ce n'est pas le cas, une règle de projection sur un sous-espace est nécessaire :

Règle 2 (projection sur un sous-espace)

Soit $c = (c_1, c_2)^T$ un vecteur aléatoire de loi :

$$\mathcal{E}\left(\begin{pmatrix} \bar{c}_1 \\ \bar{c}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}, \begin{pmatrix} E_{11} & E_{12} \\ E_{12}^T & E_{22} \end{pmatrix}\right) \quad (7)$$

Le sous-vecteur aléatoire c_1 suit une loi d'ensemble $\mathcal{E}'(\bar{c}', \Sigma', E')$ où :

$$\begin{cases} \bar{c}' &= \bar{c}_1 \\ \Sigma' &= \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T \\ E' &= E_{11} - E_{12} E_{22}^{-1} E_{12}^T \end{cases} \quad (8)$$

La preuve complète est donnée dans [6].

Grâce aux règles précédentes, nous pouvons déduire d'autres règles dédiées à la propagation d'informations visuelles. Au vu de la chaîne des données visuelles décrite en section 2, nous avons donc dérivé des règles spécialisées pour la rétro-projection et le changement de repère.

3.1 Rétro-projection

L'opération de rétro-projection dépend de la configuration utilisée.

- Si l'on note $c^i = (X^i, Y^i)^T$ et $c^c = (x^c, y^c, z^c)^T$, en configuration **monoculaire**, la transformation de rétro-projection est caractérisée par l'équation suivante : $c^c = (z^c X^i, z^c Y^i, z^c) = T(X^i, Y^i, z^c)$. La règle de rétro-projection se déduit donc de la règle 1 après avoir fixé la distribution de z^c à la distribution d'information nulle traduisant l'inobservabilité le long de cet axe : $\mathcal{E}(\bar{z}^c, 0, 0)$, où \bar{z}^c est issu des itérations précédentes lors du processus d'estimation (voir section 4).
- En configuration **binoculaire**, deux images du même objet sont disponibles, nous les appelons "image de référence" et "image courante". Un point c se projette en c^i dans l'image de référence et en c^{ic} dans l'image courante. Si l'on note $l^i = (c^i, 1)$, $l^{ic} = (c^{ic}, 1)$ et (R, t) le déplacement entre les deux caméras (voir figure 3) alors :

$$\begin{cases} (c - t) \wedge R l^{ic} &= 0 \\ c \wedge l^i &= 0 \end{cases} \quad (9)$$

C'est un système de six équations dont quatre sont indépendantes lorsque les caméras ne sont pas alignées. Nous pouvons trouver une solution aux moindres carrés $c = T(R, t, c^i, c^{ic})$ puis, grâce à la règle 1, déterminer les paramètres de \mathcal{E}^c (voir [6]).

3.2 Changement de repère

Les paramètres de déplacement entre \mathcal{R} et \mathcal{R}_c sont notés p_e . Nous les modélisons par une loi normale : $p_e = \bar{p}_e + \mathcal{N}(0, P_e)$. Si c^o exprime les coordonnées de c^c dans \mathcal{R} alors :

$$c^o = Rc^c + t = T(c^c, p_e) \quad (10)$$

où R et t dépendent de p_e . L'application de la règle 1 est alors immédiate. Pour un robot manipulateur, p_e représente la position des articulations et l'étalonnage caméra/effecteur.

4 Estimation

A l'initialisation, deux images du même objet sont nécessaires. En utilisant la règle de rétro-projection dans le cas binoculaire (section 3), nous pouvons calculer les paramètres de la distribution \mathcal{E}_0 qui initialise le modèle de connaissance.

Comme le montre la figure 3, deux images ne suffisent à donner une bonne estimation ni de la forme, ni de la position de l'objet. Lorsque la caméra est en mouvement, d'autres images peuvent affiner l'estimation. A l'instant k , la distribution a priori est $\mathcal{E}_k(\bar{c}_k, \Sigma_k, E_k)$. A l'instant $k + 1$, la vraisemblance de l'observation est $\mathcal{E}_{k+1}^o(\bar{c}_{k+1}^o, \Sigma_{k+1}^o, E_{k+1}^o)$. Nous estimons séparément les paramètres liés à l'incertitude et ceux liés à l'erreur.

La mise à jour de l'incertitude est un problème d'estimation gaussienne. On peut montrer (voir [6]) que la distribution a posteriori est $\mathcal{N}(\bar{c}_{k+1}, \Sigma_{k+1})$ où :

$$\begin{cases} \bar{c}_{k+1} = (\Sigma_k + \Sigma_{k+1}^o)^{-1}(\Sigma_k \bar{c}_k + \Sigma_{k+1}^o \bar{c}_{k+1}^o) \\ \Sigma_{k+1} = \Sigma_k + \Sigma_{k+1}^o \end{cases} \quad (11)$$

Quant au nouveau support de l'erreur, c'est l'intersection de deux ellipsoïdes concentriques E_k et E_{k+1}^o (le décentrage correspondant à p dans le modèle additif peut être négligé par rapport à la taille des ellipsoïdes). Nous avons approximé cette intersection par un ellipsoïde intermédiaire entre l'ellipsoïde interne et l'ellipsoïde externe (voir [6]).

Simulation

Les résultats que nous présentons sont issus d'une simulation en configuration monoculaire. A l'initialisation (figure 4-a), la caméra est placée dans deux positions voisines de sorte que l'ellipsoïde estimé E_0 est fortement allongé en direction de la caméra (axe de faible observabilité). Ensuite la caméra effectue une trajectoire circulaire à vitesse constante (3 cm par itération). Au centre de cette trajectoire, se trouve l'objet : une sphère dont la position ($X = -4.7$ cm, $Y = 4.7$ cm et $Z = 63$ cm) et le rayon (4 cm) sont connus. L'erreur sur la position de la caméra est un bruit additif gaussien centré, de variance 10 cm en translation et 5 degrés en rotation. L'estimation s'effectue sur 200 itérations.

La figure 4-b montre la convergence de la position estimée \bar{c}_k vers sa valeur réelle. La figure 4-c montre la convergence des axes de l'ellipsoïde englobant E_k . Les deux axes parallèles au

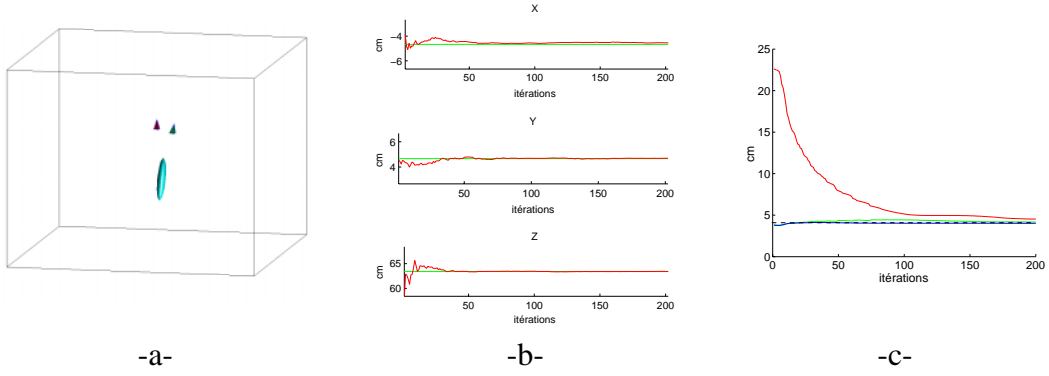


FIG. 4 – *Simulation de l'estimation.*

plan image à l'initialisation sont bien estimés dès le départ. En revanche, l'axe parallèle à l'axe optique au moment de l'initialisation utilise les observations acquises au cours du mouvement pour converger vers sa valeur finale.

5 Exploration

Pour générer automatiquement les mouvements exploratoires de la caméra, l'idée est de minimiser l'incertitude de la connaissance prédite à la prochaine itération. A l'instant k nous avons estimé $\mathcal{E}_k(\bar{c}_k, \Sigma_k, E_k)$ (exprimée dans le repère de la caméra). Si à $k + 1$ la caméra se déplace de (R, t) , grâce aux règles 1 et 2, nous pouvons calculer la connaissance prédite. Si l'on décompose :

$$R^T \Sigma_k R = \begin{pmatrix} A & B \\ B^T & c \end{pmatrix} \quad (12)$$

où A est $(2,2)$, B $(2,1)$ et c scalaire alors son expression est donnée par (voir [6]) :

$$\widehat{\Sigma}_{k+1} = \begin{pmatrix} 2A - \frac{BB^T}{c} & B \\ B^T & c \end{pmatrix} \quad (13)$$

lorsque l'objet est centré dans l'image. En pratique nous utilisons un asservissement visuel pour réaliser ce centrage ce qui a pour effet de maintenir l'objet dans le champ de vue de la caméra durant tout le processus (voir [7]).

Le déplacement (R, t) doit être calculé de telle sorte que $\widehat{\Sigma}_{k+1}$ soit maximal pour un certain critère. On peut montrer que la profondeur z_{k+1} de la caméra par rapport à l'objet n'influence pas $\widehat{\Sigma}_{k+1}$. Ainsi la translation optimale peut être calculée dans l'image. Le degré de liberté restant peut servir à réguler la surface projetée dans l'image. Une autre conséquence est que la direction de translation t est liée à l'axe de rotation u par $t = z \wedge u$ où z est le vecteur unitaire normal au plan image. La commande peut donc être définie soit en translation (t) soit en rotation (u). Nous avons développé deux lois de commande :

Commande locale : la commande optimise localement le gain d'information que nous quantifions par $trace(\widehat{\Sigma}_{k+1})$. Son expression s'obtient, après développement, par résolution d'un système linéaire dont l'inconnue est le vecteur u de rotation de la caméra.

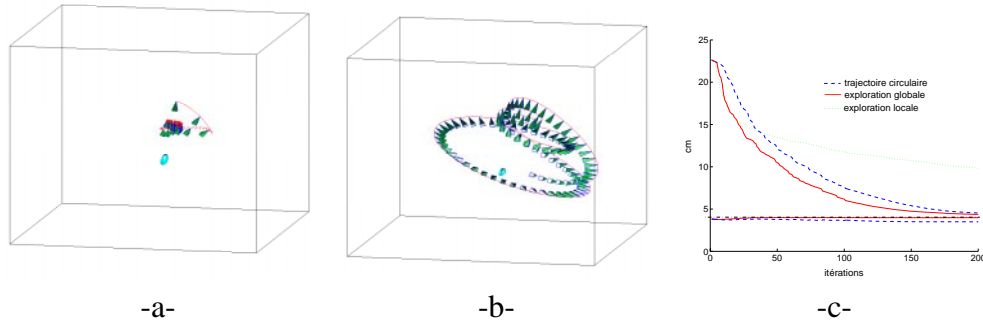


FIG. 5 – Simulation de l’exploration.

Commande globale : la commande tend à rejoindre le meilleur point de vue : celui qui conduit à la plus grande “information sphérique” ($\widehat{\Sigma}_{k+1}$ proportionnelle à l’identité). La notion de sphéricité est importante vis-à-vis de l’isotropie de la reconstruction. On peut montrer que ce meilleur point de vue est situé sur l’axe propre de Σ_k associé à la plus grande valeur propre.

Simulation

Les deux lois de commande ont été simulées dans les mêmes conditions que la section 4. L’exploration locale (figure 5-a) conduit à un minimum local. Ceci est dû au caractère local de l’optimisation (voir [6]) et correspond à des configurations pour lesquelles la commande calculée s’annule (par exemple lorsque la caméra est sur une direction propre de Σ_k). En revanche la commande globale (figure 5-b) conduit à une trajectoire très intuitive : la caméra effectue une spirale du pôle à l’équateur. Au terme de l’exploration, la commande locale aboutit à une réduction de 67% de la différence (*volume initial - volume réel*) contre 99.5% pour la commande globale. La figure 5-c témoigne du gain introduit par cette dernière commande par rapport à une trajectoire circulaire. La convergence des axes est à la fois plus rapide et plus précise. Le critère d’arrêt choisi pour la simulation de la commande globale est la trace de la matrice de covariance Σ^{-1} qui traduit l’incertitude du modèle reconstruit selon chaque axe. Le seuil correspondant à l’arrêt de la simulation ayant été arbitrairement fixé à 0.02. Notons qu’en raison du minimum local, ce seuil n’est jamais atteint lors de l’exploration locale.

6 Expérimentation

Pour valider notre étude en situation réelle, nous devons extraire le masque des objets d’intérêt. Afin de n’imposer aucune contrainte concernant l’aspect des objets (texture, couleur, niveau de gris,...), nous faisons l’hypothèse (peu restrictive) que la frontière des objets est marquée par une discontinuité de profondeur. Alors, pour tout mouvement de translation de la caméra, le mouvement projeté de chaque objet est discernable. Ainsi un algorithme de segmentation au sens du mouvement nous donnera le masque des objets. Nous avons choisi l’algorithme d’estimation paramétrique du mouvement de Odobez et Bouthemy [14] qui fournit une carte des points non conformes au mouvement dominant.

Le processus d’exploration a été implanté sur un robot cartésien à six axes. La rapidité de l’algorithme d’estimation nous permet d’estimer plusieurs objets en temps réel. La figure 6 décrit la

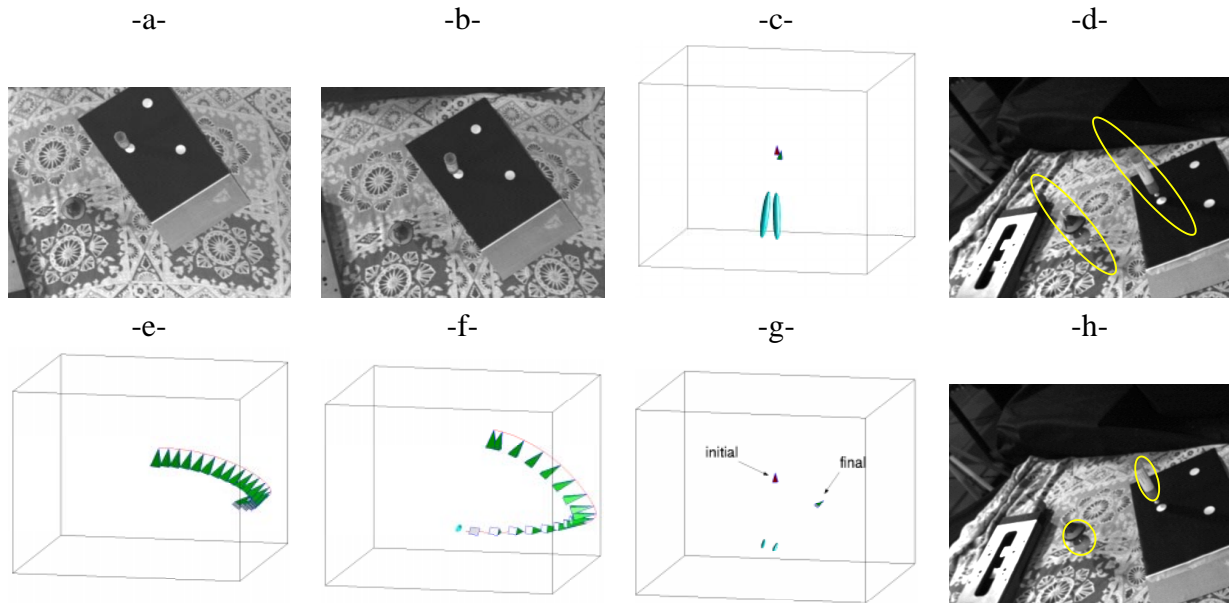


FIG. 6 – Résultats expérimentaux.

reconstruction approchée de deux objets : un champignon et un marqueur. A l'initialisation, deux images de la scène sont disponibles (figure 6-a et 6-b) et conduisent à l'estimation E_0 donnée figure 6-c (l'estimation de \bar{c} n'est pas analysé dans cette partie puisque nous ne disposons pas de la position réelle des objets). Puisqu'à l'initialisation, les caméras sont fixes, la détection de mouvement ne permet pas de segmenter les objets ; leur masque est grossièrement défini à la main. La figure 6-d représente la projection de E_0 dans l'image finale et témoigne de la nécessité d'affiner cette estimation trop largement englobante. Dans un deuxième temps, nous basons la stratégie de déplacement de la caméra sur l'exploration de l'un des deux objets (le champignon). Grâce au bruit inhérent à l'expérimentation, la commande locale (figure 6-e) ne rencontre plus de minimum local. La commande globale (figure 6-f) conduit à une trajectoire similaire à celle simulée même si le temps expérimental est plus court en raison des butées articulaires. L'estimation finale E_k des objets (figure 6-g) a été projetée dans l'image finale (figure 6-h) pour montrer l'efficacité de l'algorithme.

7 Conclusion

Nous avons défini un modèle représentant chaque objet par une distribution probabiliste approchée nous permettant de calculer pour chaque point de la scène sa probabilité d'appartenance à un objet. Ce modèle est peu coûteux puisqu'il requiert simplement l'estimation d'un vecteur 3D et de deux matrices 3D symétriques. Une règle de propagation associée à ce modèle nous a permis de construire un processus d'estimation rapide et robuste. Nous avons ensuite défini un processus d'exploration autonome et optimal.

Un point abordé en introduction et non encore implémenté est la reconstruction d'objets fortement concaves. Au vu de nos développements, une telle reconstruction apparaît clairement réalisable si toutefois nous sommes capables de découper l'objet en parties cohérentes (non fortement concaves) et de suivre chacune de ces parties au cours du processus. D'autre part, le

modèle que nous avons construit et les outils développés constituent une base intéressante pour imaginer des tâches de plus haut niveau. Nous nous sommes intéressés à l'exploration d'objets apparaissant entièrement dans le champ de la caméra. Nos prochains travaux concerneront la recherche de tous les objets de la scène.

Remerciements. Cette étude a reçu le soutien du projet LARA de l'INRIA et de la région Bretagne.

Références

- [1] N. Ayache. *Artificial Vision for Mobile Robots*. The MIT Press, Cambridge, Etats-Unis, 1991.
- [2] C. Connolly. The determination of next best views. In *Proc. IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 432–435, St Louis, Etats-Unis, mars 1995.
- [3] H. F. Durrant-Whyte. *Integration, Coordination, and Control of Multi-Sensor Robot Systems*. Kluwer Academic Publishers, Boston, Etats-Unis, 1987.
- [4] H. F. Durrant-Whyte. Uncertain geometry in robotics. *IEEE Journal of Robotics and Automation*, 4(1):23–31, 1988.
- [5] F. P. Ferrie et M. D. Levine. Deriving coarse 3D models of objects. In *CVPR'88 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Ann Arbor, MI, June 5–9, 1988)*, pages 345–353, Washington, DC., Etats-Unis, juin 1988. Computer Society Press.
- [6] G. Flandin et F. Chaumette. Complex objects exploration. Rapport de recherche, IRISA-INRIA, mars 2001.
- [7] G. Flandin, F. Chaumette, et E. Marchand. Eye-in-hand / eye-to-hand cooperation for visual servoing. In *IEEE Int. Conf. on Robotics and Automation*, San Francisco, Etats-Unis, avril 2000.
- [8] U. D. Hanebeck et G. Schmidt J. Horn. On combining set theoretic and bayesian estimation. In *IEEE Int. Conf. on Robotics and Automation*, 1996.
- [9] K. N. Kutulakos, C. R. Dyer, et V. J. Lumelsky. Provable strategies for vision-guided exploration in three dimensions. In *Proc. 1994 IEEE Int. Conf. Robotics and Automation*, pages 1365–1372, Los Alamitos, Etats-Unis, 1994. IEEE Computer Society Press.
- [10] S. Lacroix et R. Chatila. *Motion and perception strategies for outdoor mobile robot navigation in unknown environments*. Lecture Notes in Control and Information Sciences, 223. Springer-Verlag, New York, Etats-Unis, 1997.
- [11] D. G. Maksarov et J. P. Norton. State bounding with ellipsoidal set description of the uncertainty. *International Journal on Control*, 65(5):847–866, 1996.
- [12] E. Marchand et F. Chaumette. Active vision for complete scene reconstruction and exploration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(1), janvier 1999.
- [13] D. Marr et K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Royal Soc. London Bulletin*, B200:269–294, 1977.
- [14] J.M. Odobez et P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, 1995.
- [15] F. C. Schweppe. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control*, pages 22–28, 1968.
- [16] E. Welzl. Smallest enclosing disks (balls and ellipsoids). *Lecture Notes in Computer Science*, 555:359–370, 1991.
- [17] P. Whaitte et F. P. Ferrie. Autonomous exploration: Driven by uncertainty. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 339–346, Los Alamitos, Etats-Unis, Juin 1994. IEEE Computer Society Press.
- [18] H. S. Witsenhausen. Sets of possible states of linear systems given perturbed observations. *IEEE Transactions on Automatic Control*, pages 556–558, 1968.