

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

Visual data fusion for complex objects localization

par Grégory FLANDIN, François CHAUMETTE

IRISA-INRIA Rennes Campus de Beaulieu, 35042 Rennes Cedex, France e-mail : {prénom. Nom@irisa.fr}

résumé et mots clés

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.

Vision active, exploration, fusion, reconstruction, modélisation probabiliste.

abstract and key words

Visual sensors provide exclusively uncertain and partial knowledge of a scene. In this article, we present a suitable scene knowledge representation that makes integration and comparison of uncertain and partial data possible. It is based on a mixture of stochastic and set membership models. Considering a gaussian approximation of uncertainties and an ellipsoidal approximation for error bounds, we build an efficient estimation process integrating visual data online. Based on this estimation scheme, we perform online and optimal exploratory commands.

Active vision, exploration, fusion, reconstruction, stochastic modelling.

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. La connais-

sance de chaque objet de la scène peut alors se résumer à celle de son volume englobant (centre et enveloppe). 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 [15, 12]. 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 [9]. Dans cet article, nous tirons parti des deux modèles précédents au sein d'une représentation mixte des objets.

Un autre problème concernant la modélisation de l'environnement est la génération autonome de 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. Whaite et Ferries [14] 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 [7] exploitent les variations de frontière d'occultation. Lacroix et Chatila [8] 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 [10] utilisent la vision monoculaire active pour explorer des objets géométriques tels que cylindres ou polyèdres. 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 visuel nous permettent de construire une solution analytique au problème d'exploration optimale. En section 2, nous décrivons précisément le mélange de modèles utilisé ainsi qu'une règle de propagation associée qui sera appliquée à la propagation de données visuelles. En section 3, nous décrivons un processus d'estimation qui tire parti du mouvement de la caméra pour reconstruire son environnement. Dans le contexte de l'exploration autonome, nous définissons le mouvement optimal de la caméra en section 4.

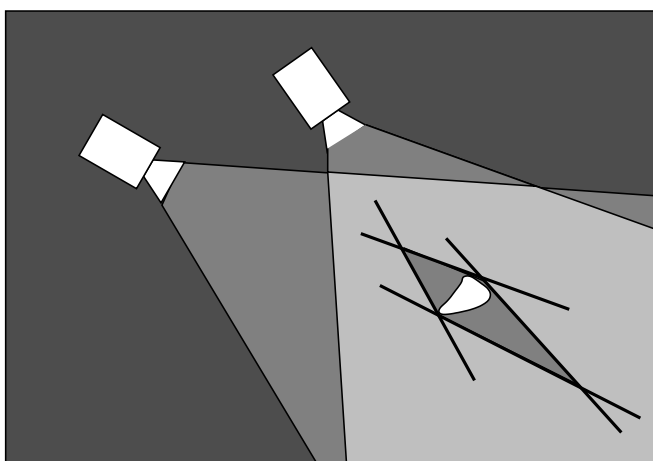


Figure 1. – Principe de l'exploration

2. modélisation et règle de propagation

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 :

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

Dans cette relation :

- $\mathcal{P}_c(x|x \notin \mathcal{O})$ est la probabilité que le point c de \mathcal{O} soit en x sachant que $x \notin \mathcal{O}$. $\mathcal{P}_c(x|x \notin \mathcal{O})$ vaut donc 0, par définition de c .
- $\mathcal{P}_c(x|x \in \mathcal{O})$ est la probabilité que le point c de \mathcal{O} soit en x sachant que $x \in \mathcal{O}$. Nous modélisons naturellement $\mathcal{P}_c(x|x \in \mathcal{O})$ par une loi uniforme dont la valeur se calcule, après normalisation.

En résumé, la probabilité $\mathcal{P}(x \in \mathcal{O})$ est proportionnelle à $\mathcal{P}_c(x)$. 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 2) :

$$c = \bar{c} + p + e$$

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 $\mathcal{N}(0, P)$ (approximation satisfaisante pour de nombreuses sources d'incertitudes) et e est uniformément distribuée sur un ellipsoïde (assimilé à sa matrice E). Afin de pouvoir tenir compte de l'inobservabilité partielle des capteurs visuels, nous préférons, par la suite, utiliser la **matrice d'information** $\Sigma = P^{-1}$ (analogie avec la matrice d'information de Fisher) plutôt que la matrice de covariance. Une variance infinie le long de l'axe d'inobservabilité est ainsi remplacée par une informa-

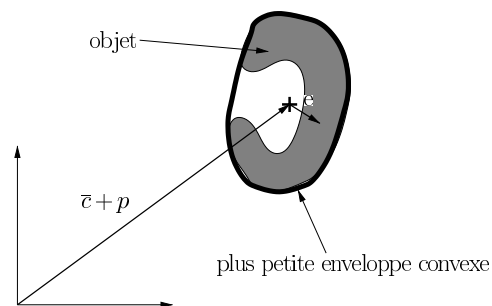


Figure 2. – Le modèle additif unifie l'aspect géométrique représenté par \bar{c} (position de l'objet) et e (dont le support représente la forme de l'objet) et l'aspect probabiliste représenté par p (incertitude de localisation).

tion nulle. Par abus de langage, nous dirons que p suit une loi $\mathcal{N}(0, \Sigma)$. La distribution associée à un objet est donc entièrement déterminée par \bar{c} , Σ et E . Il s'agit de la convolution d'une loi uniforme \mathcal{U}_E sur E par une loi normale $\mathcal{N}(\bar{c}, \Sigma)$. Nous appelons cette distribution loi d'ensemble et la notons :

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

Nous définissons, à présent, les transformations induites par un changement d'espace des paramètres. La première règle est consacrée aux transformations qui conservent la dimension de l'espace des paramètres et la complétude de leur représentation : les difféomorphismes.

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

Soit T un difféomorphisme et c un vecteur aléatoire de loi $\mathcal{E}(\bar{c}, \Sigma, E)$ sur l'espace de définition de T . 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 $\mathcal{E}'(\bar{c}', \Sigma', E')$ sur l'espace image de T où :

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

La démonstration ([5]) découle d'une approximation de T au premier ordre. 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 ces règles requièrent que T soit un difféomorphisme par rapport à c . Lorsque nous avons à faire à une transformation qui ne conserve pas la complétude de la représentation, ce qui est typiquement le cas des observations visuelles du fait de leur nature projective, une règle de projection sur un sous-espace est nécessaire :

Règle 2 (Projection sur un sous-espace)

Soit c un vecteur aléatoire se décomposant en $c = (c_1, c_2)^T$. Si c suit une loi

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

le vecteur aléatoire c_1 , projection de c sur un sous-espace, suit une loi d'ensemble $\mathcal{E}'(\bar{c}', \Sigma', E')$ où :

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

Grâce à ces règles, nous pouvons déduire d'autres règles dédiées à la propagation d'informations visuelles. Nous considérons trois étapes dans la chaîne des observations visuelles (figure 3) :

1. Dans l'image, la mesure est une loi d'ensemble 2D $\mathcal{E}^i(\bar{c}^i, \Sigma^i, E^i)$. Ainsi \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 5 comment extraire cette projection en pratique. Quant à \bar{c}^i et E^i , ils s'obtiennent grâce à des algorithmes tels que celui proposé dans [13]. Σ^i doit, pour sa part, prendre en compte toutes les sources d'incertitudes inhérentes au calcul de l'ellipse : paramètres intrinsèques, imprécision du traitement d'image.
2. L'opération transformant les mesures 2D en observations 3D est appelée rétro-projection. La loi d'ensemble 3D associée est notée $\mathcal{E}^c(\bar{c}^c, \Sigma^c, E^c)$. L'opération de rétro-projection dépend de la configuration utilisée. Dans [5], nous présentons les cas monoculaires et binoculaires. Dans la suite, nous nous contentons de présenter, sans démonstration, le cas monoculaire.

Règle 3 (Rétro-projection d'une loi d'ensemble)

La rétro-projection c^c de c^i suit, en première approximation, une loi d'ensemble $\mathcal{E}^c(\bar{c}^c, \Sigma^c, E^c)$:

$$\begin{cases} \bar{c}^c &= (\bar{z}^c \bar{X}^i, \bar{z}^c \bar{Y}^i, \bar{z}^c)^T \\ \Sigma^c &= J^T \begin{pmatrix} \Sigma^i & 0 \\ 0 & 0 \end{pmatrix} J \\ E^c &= J^T \begin{pmatrix} E^i & 0 \\ 0 & 0 \end{pmatrix} J \end{cases}$$

$$\text{où } J = \left. \frac{\partial T^{-1}}{\partial c} \right|_{\bar{c}} = \begin{pmatrix} 1/\bar{z}^c & 0 & -\bar{X}^i/\bar{z}^c \\ 0 & 1/\bar{z}^c & -\bar{Y}^i/\bar{z}^c \\ 0 & 0 & 1 \end{pmatrix}$$

Dans cette règle, \bar{z}^c est *a priori* inconnu mais nous verrons en section 3 que sa valeur peut être fixée par l'estimation à l'instant précédent. Réciproquement à la rétro-projection, se pose le problème de la projection perspective qui n'est pas un difféomorphisme mais que l'on peut approximer par une projection paraperspective (composition d'une projection et d'un difféomorphisme) [5].

Règle 4 (Projection perspective d'une loi d'ensemble)

Si l'on note

$$E^c = \begin{pmatrix} E_{11}^c & E_{12}^c \\ E_{12}^{cT} & E_{22}^c \end{pmatrix} \text{ et } \Sigma^c = \begin{pmatrix} \Sigma_{11}^c & \Sigma_{12}^c \\ \Sigma_{12}^{cT} & \Sigma_{22}^c \end{pmatrix}$$

où E_{11}^c et Σ_{11}^c sont (2,2), E_{12}^c et Σ_{12}^c sont (2,1) et E_{22}^c et Σ_{22}^c scalaires alors la projection perspective \bar{c}^i de c^c suit, en première approximation, une loi d'ensemble $\mathcal{E}(\bar{c}^i, \Sigma^i, E^i)$ où :

$$\begin{cases} \bar{c}^i &= (\bar{x}^c / \bar{z}^c, \bar{y}^c / \bar{z}^c)^T \\ \Sigma^i &= J^T [\Sigma_{11}^c - \frac{\Sigma_{12}^c \Sigma_{12}^{cT}}{\Sigma_{22}^c}] J \quad \text{où } J = \begin{pmatrix} \bar{z}^c & 0 \\ 0 & \bar{z}^c \end{pmatrix} \\ E^i &= J^T [E_{11}^c - \frac{E_{12}^c E_{12}^{cT}}{E_{22}^c}] J \end{cases}$$

3. Enfin, nous devons exprimer chaque observation dans un repère commun de référence \mathcal{R} . La loi d'ensemble associée est notée $\mathcal{E}^o(\bar{c}^o, \Sigma^o, E^o)$. La règle correspondante est appelée règle de changement de repère [5].

3. estimation

À l'initialisation, deux images du même objet sont nécessaires. En utilisant la règle de rétro-projection dans le cas binoculaire (section 2), nous pouvons calculer les paramètres de la distribution \mathcal{E}_0 qui initialise le modèle de connaissance. Comme le montre la figure 1, deux images ne suffisent pas 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. À l'instant k , la distribution *a priori* est $\mathcal{E}_k(\bar{c}_k, \Sigma_k, E_k)$. À 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 que la distribution *a posteriori* est $\mathcal{N}(\bar{c}_{k+1}, \Sigma_{k+1})$ où :

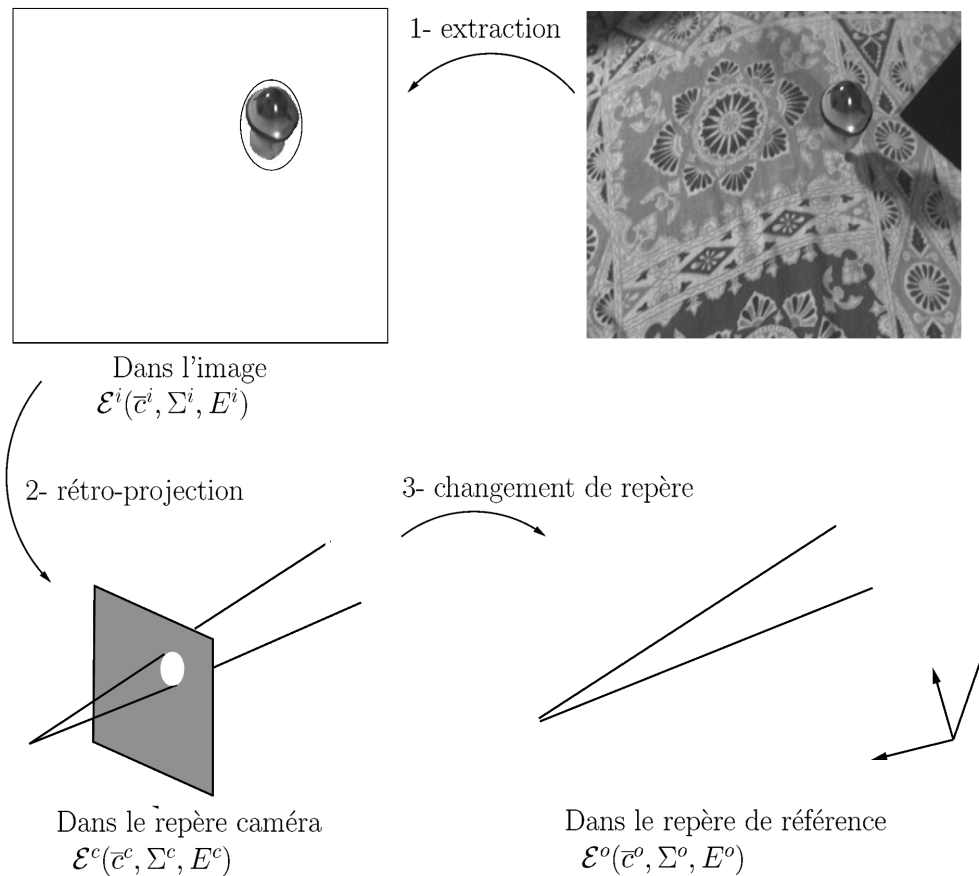


Figure 3. – Transformations successives des données visuelles 1-Extraction de la projection de l'objet dans l'image et calcul de la loi 2D associée. 2-Rétro-projection de la loi 2D dans l'espace 3D. 3-Expression de l'observation dans le repère de référence.

$$\overline{c_{k+1}} = (\Sigma_k + \Sigma_{k+1}^o)^{-1}(\Sigma_k \overline{c_k} + \Sigma_{k+1}^o \overline{c_{k+1}^o})$$

$$\Sigma_{k+1} = \Sigma_k + \Sigma_{k+1}^o$$

Quant au nouveau support de l'erreur, c'est l'intersection de deux ellipsoïdes (E_k and E_{k+1}^o) supposés centrés. Cette intersection doit être approchée par un ellipsoïde. Il existe deux types d'approximation classiques : l'approximation externe notée E_{k+1}^+ et l'approximation interne notée E_{k+1}^- dont les calculs sont disponibles dans [5]. Puisque très pessimiste, l'utilisation de E^+ est plus robuste que celle de E^- mais révèle un taux de convergence très lent. Il semble judicieux d'utiliser un compromis intermédiaire $E^- \subset E \subset E^+$. Par exemple en calculant une somme pondérée de E^+ et E^- . En effet on peut montrer que l'ellipsoïde $\alpha E^+ + (1 - \alpha)E^-$ est inclus dans E^+ et contient E^- . Nous verrons qu'un choix empirique de α peut être fixé convenablement pour toutes les simulations et expérimentations effectuées.

simulation

Les résultats que nous présentons sont issus d'une simulation en configuration monoculaire. Après l'étape d'initialisation (figures 4-a,b)), la caméra effectue une trajectoire circulaire à vitesse constante (3cm par itération). Au centre de cette trajectoire, se trouve l'objet : une sphère dont la position et le rayon sont connus. L'incertitude 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 sans critère d'arrêt.

La figure 4-b montre la convergence de la position estimée vers sa valeur réelle. La convergence des axes de l'ellipsoïde a été simulée en utilisant E^+ , E^- et $E = 0.98E^+ + 0.02E^-$. La figure 4-c montre la variation de la convergence en fonction de l'approximation utilisée.

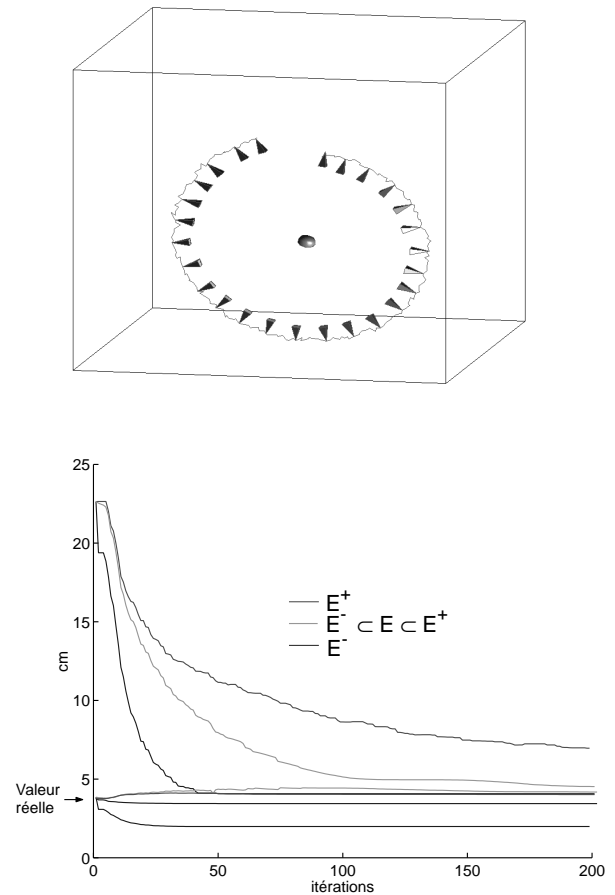
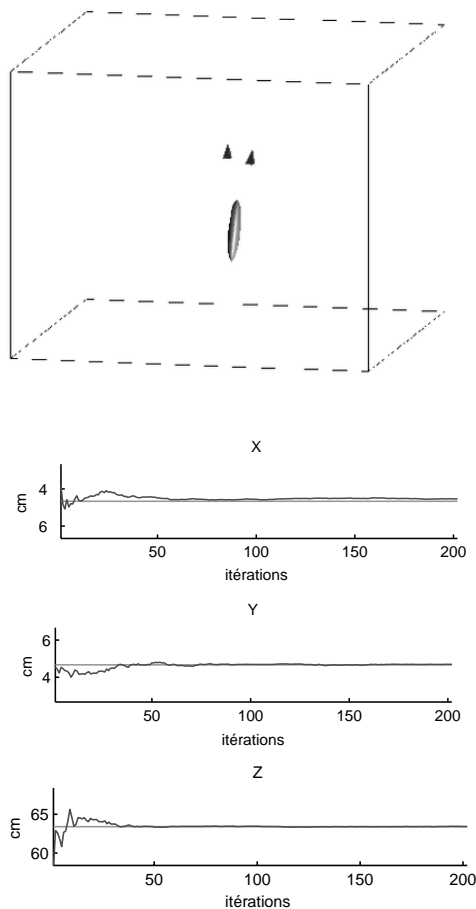


Figure 4. – Simulation de l'estimation. a) À l'initialisation, la caméra est placée dans deux positions voisines de sorte que l'ellipsoïde estimé est fortement allongé en direction de la caméra. b) Ensuite, la caméra effectue une trajectoire circulaire à vitesse constante. c) Estimation de la localisation de l'objet. d) Estimation des axes de l'enveloppe 3D.

4. 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. À 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 à la règle 1, nous pouvons calculer la connaissance prédite. Son expression est donnée par (voir [5]) :

$$\widehat{\Sigma}_{k+1} = R^T \Sigma_k R + J_1^T \begin{pmatrix} J_0^T R^T \Sigma_k R J_0 & 0 \\ 0 & 0 \end{pmatrix} J_1$$

où

$$\begin{cases} J_0 = \begin{pmatrix} z_{k+1} & 0 \\ 0 & z_{k+1} \\ 0 & 0 \end{pmatrix} \\ J_1 = \begin{pmatrix} 1/z_{k+1} & 0 & -X_{k+1}^i/z_{k+1} \\ 0 & 1/z_{k+1} & -Y_{k+1}^i/z_{k+1} \\ 0 & 0 & 1 \end{pmatrix} \end{cases}$$

X_{k+1}^i et Y_{k+1}^i sont les coordonnées prédites de \bar{c}_i à $k + 1$. En pratique nous utilisons un asservissement visuel pour centrer l'objet dans l'image ($\forall k, X_k^i = Y_k^i = 0$) afin d'assurer la visibilité de l'objet durant tout le processus (voir [6]).

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 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 t soit en u . Nous avons développé deux lois de commande :

commande locale (LO) : Dans ce cadre, le mouvement de la caméra optimise localement l'accroissement d'information soit $\widehat{\Sigma}_{k+1}$. Le critère choisi pour quantifier cet accroissement est la trace de $\widehat{\Sigma}_{k+1}$. En maximisant la somme des valeurs propres de la matrice d'information nous espérons capter la notion de sphéricité nécessaire à l'isotropie de l'exploration dans les sens où une valeur propre faible pénaliserait le critère. À l'instant $k + 1$, la caméra aura pivoté d'un angle $\alpha \geq 0$ autour du vecteur unitaire $u = (u_x, u_y, 0)$. L'angle α détermine la vitesse de la caméra. Pour α fixé, le mouvement optimal de la caméra est défini par son axe de rotation. Soit :

$$(u_x, u_y) = \operatorname{argmax} \operatorname{tr}[\widehat{\Sigma}_{k+1}]$$

Son expression s'obtient, après développement, par résolution d'un système linéaire d'inconnues u_x et u_y .

stratégie de meilleur point de vue (BVP) : À présent, plutôt que d'optimiser localement $\widehat{\Sigma}_{k+1}$, le déplacement est calculé dans la direction du meilleur point de vue : celui qui conduit à la plus grande information sphérique $\widehat{\Sigma}_{k+1}$, c'est-à-dire proportionnelle à la matrice identité. Nous montrons le résultat suivant : le meilleur point de vue est situé sur la direction propre de Σ_k de plus grande information. Le vecteur vitesse associé est donc dirigé vers l'intersection entre le plan image et l'axe de plus grande information. Cette commande est recalculée à chaque itération sans attendre d'avoir atteint le meilleur point de vue. La norme de la vitesse est indépendante du processus d'exploration et peut être fixée une fois pour toutes.

simulation

Les deux lois de commande ont été simulées sous les mêmes conditions que la section 3. L'exploration locale (figure 5-a) conduit à un minimum local. Ceci est dû au caractère local de l'optimisation. En revanche la commande BVP (figure 5-b) conduit à une trajectoire très intuitive : la caméra effectue une spirale du pôle à l'équateur. Le critère d'arrêt choisi pour l'exploration est basé sur la trace de la matrice de covariance Σ^{-1} qui traduit la somme des incertitudes du modèle reconstruit selon chaque axe. Le seuil correspondant à l'arrêt de la simulation a été arbitrairement fixé à 0.002 pour la commande BVP. L'exploration LO n'atteignant jamais cette valeur du critère d'arrêt (en raison du minimum local), le nombre d'itérations a été arbitrairement fixé à 200. Au terme de la simulation, la commande locale aboutit à une réduction de 67 % de la différence (*volume initial* - *volume réel*) contre 99.5 % pour la commande BVP. 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.

5. 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

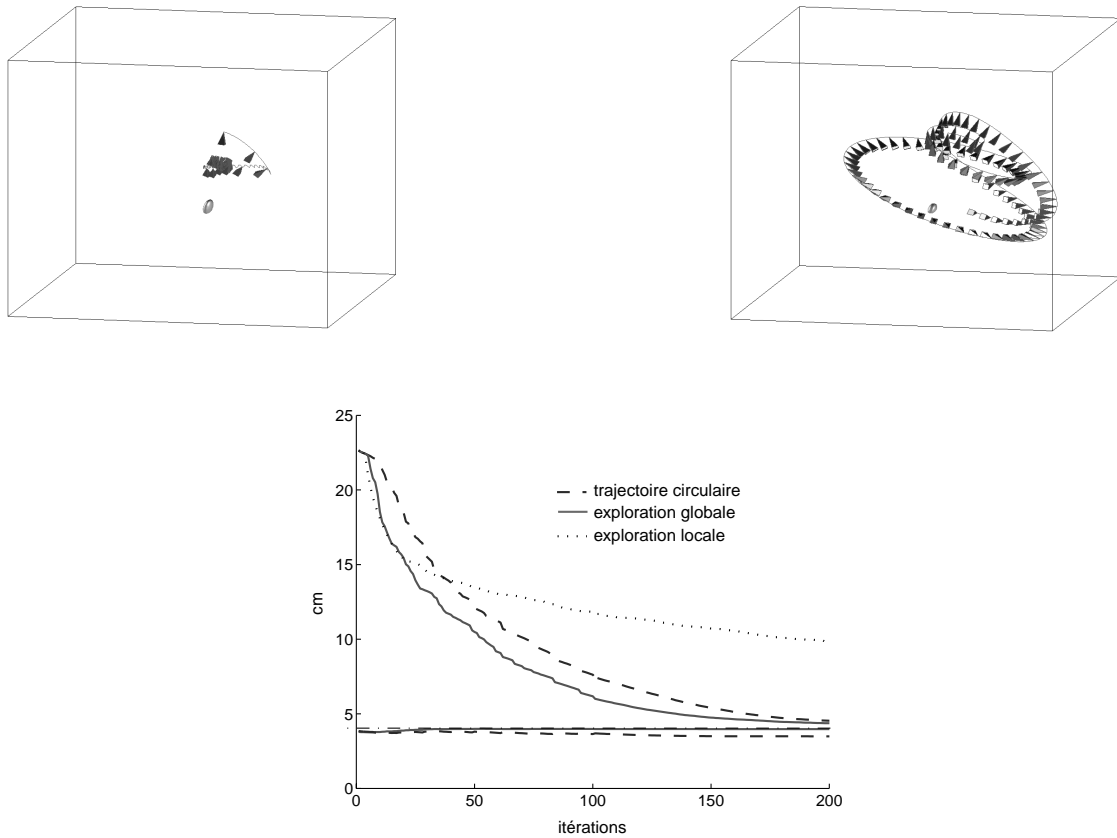


Figure 5. – Trajectoire simulée pour a) la loi commande LO, b) la loi de commande BVP. La loi de commande LO conduit à un minimum local alors que la loi de commande BVP génère une trajectoire très intuitive. c) Estimation des axes de l'enveloppe 3D.

Bouthemy [11] qui fournit une carte des points non conformes au mouvement dominant.

Le processus d'exploration a été implanté sur un robot cartésien à 6 axes. La rapidité de l'algorithme d'estimation nous permet d'estimer plusieurs objets en temps réel. La figure 6 est un exemple constitué de deux objets. À l'initialisation, deux images de la scène sont disponibles (figure 6-a et 6-b) et conduisent à l'estimation donnée figure 6-c. La figure 6-d représente la projection de cette estimation dans l'image finale et nous convainc de la nécessité d'affiner cette estimation. Dans un deuxième temps la période d'exploration est fixée à 20 s et la stratégie est basée sur l'exploration de l'un des deux objets. Grâce au bruit inhérent à l'expérimentation, la commande locale (figure 6-e) ne rencontre plus de minimum local. La commande BVP (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 de la scène (figure 6-g) a été projetée dans l'image finale (figure 6-h) pour montrer l'efficacité de l'algorithme.

6. conclusion

Notre approche se situe dans la classe des techniques d'estimation approchée de modèles. Elle diffère des travaux précédents sur plusieurs points. Tout d'abord, la décomposition additive que nous utilisons permet de représenter les objets sous forme d'un mélange de modèles stochastiques et à erreur bornée. Cette modélisation permet d'accéder simultanément à des informations probabilistes sur l'occupation des points de la scène et à des informations géométriques immédiates. D'autre part, la description du modèle dans un schéma probabiliste unifié le rend robuste à diverses sources d'incertitudes et fournit un cadre général à la fusion et à l'estimation approchée. Il nous a permis de développer des règles générales de propagation de la connaissance. En conséquence la méthode peut s'appliquer à n'importe quel capteur. L'idée sous-jacente à l'estimation du modèle tel que nous l'avons défini résulte d'une hypothèse d'indépendance entre l'incertitude de position et l'enveloppe de l'objet. Celle-ci

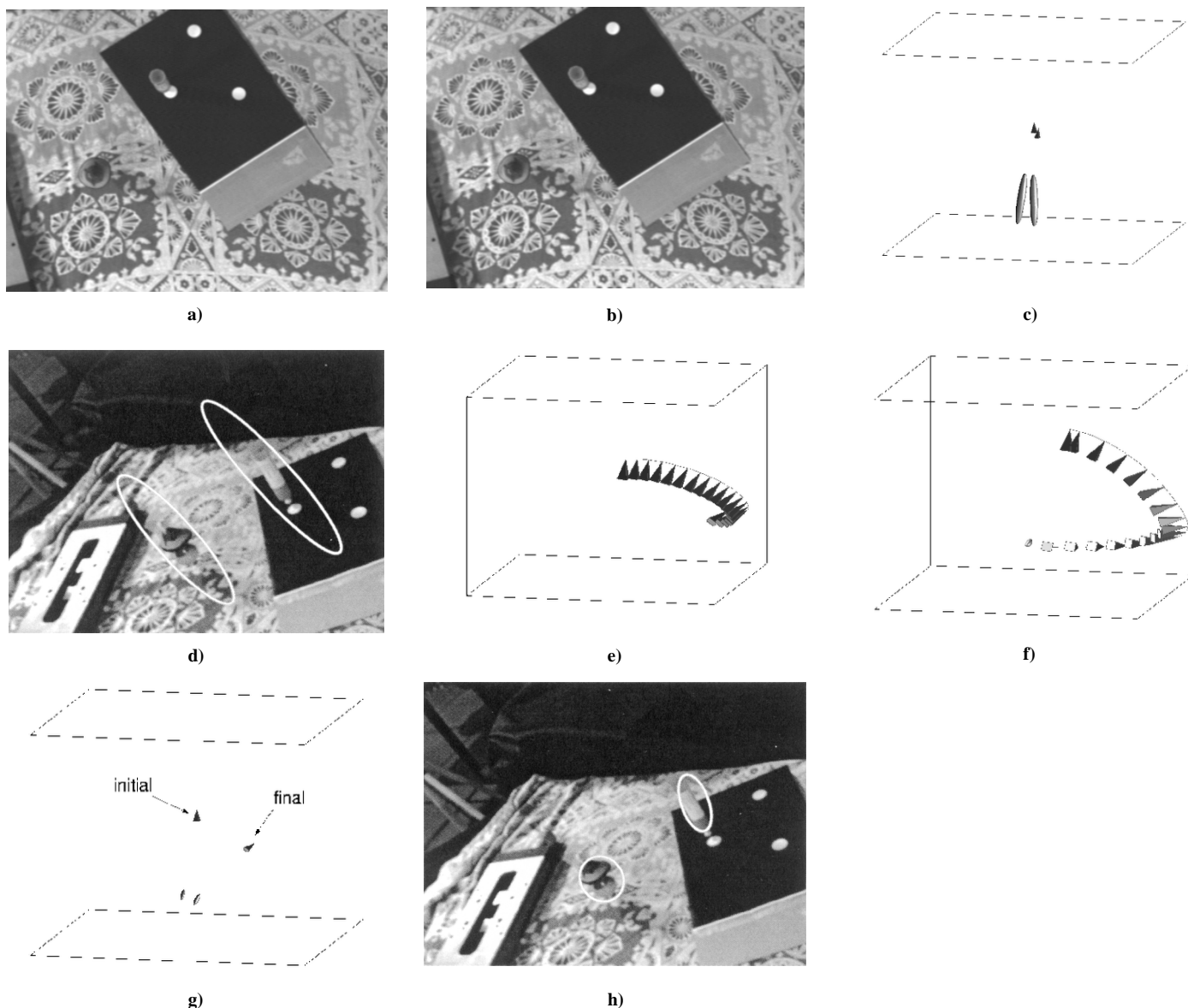


Figure 6. – Résultats expérimentaux

nous permet de retomber sur un schéma classique d'estimation *a posteriori* de la position de l'objet et d'approximation ellipsoïdale du support de l'erreur. Ayant pu dériver une expression analytique de cette approximation, la formulation de l'estimation demeure relativement compacte et rapide d'utilisation. Par ailleurs, nous avons dérivé une solution analytique au problème de l'exploration de telle sorte que le calcul d'une trajectoire d'exploration réduisant l'incertitude du modèle devient un problème soluble en ligne. Même si la reconstruction n'est qu'approchée, répondre à la question *quel mouvement de la caméra améliorera notre connaissance de l'objet ?* devient un problème bien posé. Le modèle ainsi 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.

BIBLIOGRAPHIE

- [1] N. Ayache. *Artificial Vision for Mobile Robots*. The MIT Press, Cambridge, MA, 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, Missouri, mars 1995.
- [3] H. F. Durrant-Whyte. *Integration, Coordination, and Control of Multi-Sensor Robot Systems*. Kluwer Academic Publishers, Boston, 1987.
- [4] H. F. Durrant-Whyte. Uncertain geometry in robotics. *IEEE Journal of Robotics and Automation*, 4(1) : 23-31, 1988.
- [5] G. Flandin et F. Chaumette. Complex objects exploration. Technical report, IRISA-INRIA, Janvier 2001.
- [6] 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, CA, Avril 2000.
- [7] 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, CA, 1994.
- [8] S. Lacroix et R. Chatila. *Motion and perception strategies for outdoor mobile robot navigation in unknown environments*. LNCIS, 223. Springer-Verlag, New York, 1997.
- [9] 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.
- [10] E. Marchand et F. Chaumette. Active vision for complete scene reconstruction and exploration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(1), January 1999.
- [11] 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.
- [12] F. C. Schweppe. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Trans. on Automatic Control*, pages 22-28, 1968.
- [13] E. Welzl. Smallest enclosing disks (balls and ellipsoids). *LNCS*, 555 : 359-370, 1991.
- [14] P. Whaite et F. P. Ferrie. Autonomous exploration: Driven by uncertainty. In *Proc. of the Conf. on Computer Vision and Pattern Recognition*, pages 339-346, Los Alamitos, CA, USA, June 1994.
- [15] H. S. Witsenhausen. Sets of possible states of linear systems given perturbed observations. *IEEE Trans. on Automatic Control*, pages 556-558, 1968.

Manuscrit reçu le 10 mai 2001

LES AUTEURS

François CHAUMETTE



François CHAUMETTE est né à Nantes en 1963. Après une Maîtrise de Mathématiques Appliquées et un diplôme d'ingénieur ENSM en automatique, obtenus à Nantes en 1985 et 1987, il a soutenu à l'Université de Rennes 1 une thèse en informatique en 1990, puis une Habilitation à diriger des recherches en 1998. Chercheur à l'Irisa depuis sa thèse, il est à présent Directeur de recherches INRIA au sein du projet VISTA. Il est également Editeur associé de la revue IEEE

Trans. on Robotics and Automation. Son domaine de recherches porte sur l'asservissement visuel et la perception active en vision robotique.

Grégory FLANDIN



Grégory FLANDIN est né à Roanne en 1975. Titulaire d'un diplôme d'ingénieur en automatique et d'un DEA en automatique et informatique industrielle obtenus à SUPAERO en 1997, il a soutenu, en 2001, une thèse à l'INRIA de Rennes portant sur l'exploration par vision active. Il est à présent ingénieur de recherche dans la société Astrium au centre de Toulouse où ses travaux portent sur le traitement d'images et ses

applications spatiales.