

Estimation de la structure et du mouvement 3D dans un contexte multi-capteurs

Christophe BOUCHER, Jean-Charles NOYER, Mohammed BENJELLOUN

Laboratoire d'Analyse des Systèmes du Littoral (EA 2600)
Université du Littoral Côte d'Opale

50 Rue Ferdinand Buisson, B.P. 699, 62228 Calais Cedex, France

boucher@lasl.univ-littoral.fr, noyer@lasl.univ-littoral.fr, benjello@lasl.univ-littoral.fr

Résumé – Cet article présente une méthode d'estimation multi-capteurs de la structure et du mouvement rigide d'un objet 3D, basée sur le filtrage particulaire. L'intérêt de cette méthode réside dans sa capacité à prendre en compte des modèles non-linéaires et/ou des statistiques non-gaussiennes sans étape de linéarisation, comme c'est le cas pour le filtre de Kalman étendu, par exemple. Des résultats expérimentaux sont présentés pour un système de vision composé d'un capteur délivrant des séquences d'images de profondeur et de réflexion.

Abstract – This article deals with a multisensor particle estimation method of the structure and rigid motion of a 3D object. The benefits of the Particle Filter lies on its ability to deal with non-linear models and/or non-Gaussian statistics without any linearization stage as the Extended Kalman Filter does, for example. Experimental results are shown for a vision system composed by a sensor delivering range and reflectance image sequences.

1 Introduction

Dans le domaine de la vision par ordinateur, un problème clé concerne l'estimation conjointe du mouvement et de la structure 3D à partir de données issues de capteurs différents. Il existe de nombreux travaux sur le sujet qui peuvent être principalement classés en trois catégories : les méthodes basées sur l'étude de primitives, celles basées sur le calcul du gradient et les méthodes de corrélation.

Nos précédents travaux sur le sujet [2] utilisant un filtre de Kalman étendu ont montré une bonne précision en estimation ainsi qu'une robustesse à la perte de mesures. Ce type d'approche nécessite cependant une linéarisation des équations d'état, ce qui peut conduire à des problèmes de stabilité et de convergence. Afin de palier cette difficulté, nous proposons une approche "particulaire" de ce problème d'estimation.

Le filtrage particulaire [6] est une technique d'estimation qui permet de traiter des modèles non-linéaires et des statistiques non-gaussiennes. L'idée principale de la méthode est de développer une approximation par particules aléatoires de la densité de probabilité de l'état conditionnelle (aux mesures), solution au problème d'estimation. On retrouve des approches similaires sous le nom de filtre "bootstrap", d'algorithme de condensation, etc. qui proposent également de l'approximer. A noter qu'il existe pour la méthode particulaire des résultats de convergence qui peuvent être trouvés dans [7], par exemple.

2 Modélisation 3D

Nous détaillons ici une modélisation d'état de ce problème d'estimation multi-capteurs de la structure et du

mouvement 3D. L'approche développée est basée sur les primitives de type point et segment qui vont permettre d'avoir une description filaire des objets 3D de la scène et sur un modèle rigide de leur mouvement. Le vecteur d'état est composé des coordonnées des extrémités du segment et des paramètres de mouvement rigide.

2.1 Equations de dynamique

L'évolution d'un objet 3D dans le référentiel des capteurs peut s'écrire :

$$\vec{\Theta}_{t+1} = R_t \vec{\Theta}_t + \vec{T}_t \quad (1)$$

où $\vec{\Theta}_t (X_t, Y_t, Z_t)^T$ est le vecteur des coordonnées cartésiennes 3D d'un point du segment à l'instant t et (R_t, \vec{T}_t) sont les paramètres du mouvement rigide (R_t est une matrice de rotation 3×3 suivant les angles $(\alpha_t^X, \alpha_t^Y, \alpha_t^Z)$ et $\vec{T}_t (T_t^X, T_t^Y, T_t^Z)^T$ est un vecteur 3D de translation).

Les équations de dynamique peuvent s'écrire :

$$\begin{cases} \Theta_{t+1}^i &= A_t \Theta_t^i + D_t + W_t^{\Theta^i} & i = \{1, 2\} \\ \alpha_{t+1}^i &= \alpha_t^i + w_t^{\alpha^i} & i = \{X, Y, Z\} \\ T_{t+1}^i &= T_t^i + w_t^{D^i} & i = \{X, Y, Z\} \end{cases} \quad (2)$$

soit, sous forme matricielle :

$$\Omega_{t+1} = F_t(\Omega_t) + W_t \quad (3)$$

où :

- $\Omega_t (\vec{\Theta}_t^1, \vec{\Theta}_t^2, \alpha_t^X, \alpha_t^Y, \alpha_t^Z, T_t^X, T_t^Y, T_t^Z)^T$ est le vecteur d'état de dimension 12 ;
- $W_t (W_t^{\Theta^1}, W_t^{\Theta^2}, w_t^{\alpha^X}, w_t^{\alpha^Y}, w_t^{\alpha^Z}, w_t^{D^X}, w_t^{D^Y}, w_t^{D^Z})^T$ est un vecteur de bruit blanc gaussien additif de moyenne nulle et de matrice de covariance Q_t .

2.2 Equations de mesure

Afin d'évaluer l'apport du filtrage particulière à l'estimation du mouvement et de la structure 3D, on applique cette méthode au cas d'un système de vision composé d'une camera LRC (*Laser Range Camera*) délivrant des images de profondeur et de réflexion. Ces images sont segmentées en utilisant les méthodes usuelles [3, 5, 8] pour ce type d'information. Les équations de mesures peuvent donc s'écrire comme suit :

$$\begin{cases} m_{t+1}^R &= h_R(\Omega_{t+1}) + V_{t+1}^R \\ m_{t+1}^I &= h_I(\Omega_{t+1}) + V_{t+1}^I \end{cases} \quad (4)$$

où :

- $m_{t+1}^R (u_{t+1}^{1,R}, v_{t+1}^{1,R}, Z_{t+1}^{1,R}, u_{t+1}^{2,R}, v_{t+1}^{2,R}, Z_{t+1}^{2,R})^T$ est la mesure de profondeur sur la position des extrémités du segment (u, v sont les coordonnées pixel et Z est la distance "caméra-objet" renvoyée);
- $m_{t+1}^I (u_{t+1}^{1,I}, v_{t+1}^{1,I}, u_{t+1}^{2,I}, v_{t+1}^{2,I})^T$ est la mesure de réflexion;
- V^R et V^I sont deux bruits blanc gaussiens additifs de moyenne nulle et de matrices de covariance respectives R_t^R et R_t^I ;
- les fonctions non-linéaires h_R et h_I sont les fonctions de mesures associées aux images de profondeur et de réflexion délivrées par la caméra laser de profondeur, de marque *Odetics*, dont le modèle projectif est :

$$\begin{cases} u_t^{R,I} &= \alpha_u \tan^{-1}\left(\frac{Y_t}{Z_t}\right) + u_0 \\ v_t^{R,I} &= \alpha_v \tan^{-1}\left(\frac{X_t}{Z_t}\right) + v_0 + \beta_v u_t^{R,I} \end{cases} \quad (5)$$

où $(\alpha_u, \alpha_v, \beta_v, u_0, v_0)$ sont les paramètres intrinsèques identifiés du capteur [9].

3 Solution particulière

Les équations 3 et 4 font apparaître l'aspect non-linéaire de ce problème d'estimation. Il existe diverses solutions pour résoudre ces équations de filtrage non-linéaires. La plus directe reste celle du filtre de Kalman étendu qui repose sur une étape de linéarisation des équations d'état, ce qui peut poser des problèmes de stabilité.

Cet article propose donc une solution particulière à ce problème d'estimation. L'idée directrice de cette approche est de développer une approximation de type Monte-Carlo basée sur une discrétisation particulière de l'espace de probabilité. Cette approche fournit une solution qui est uniquement limitée par le nombre fini de particules N et qui converge dès lors que ce nombre augmente [7].

On traite ici les processus de Markov à temps discret décrits par les équations suivantes :

$$\begin{cases} X_t = F(X_{t-1}, \pi_t) \\ Y_t = H(X_t) + \nu_t \end{cases} \quad (6)$$

où F et H sont des fonctions non-linéaires et π_t et ν_t qui sont des vecteurs de bruits indépendants à valeurs réelles.

L'estimateur optimal \hat{X}_t (au sens du minimum de l'erreur quadratique) de X_t , à partir de la connaissance des mesures $Y_0^t = \{Y_0, \dots, Y_t\}$, est donnée par :

$$\hat{X}_t = E(X_t | Y_0^t) = \int_{X_t} X_t dP(X_t | Y_0^t)$$

La construction de l'estimateur repose donc sur la connaissance de la densité de probabilité de l'état conditionnelle aux observations $P(X_t | Y_0^t)$.

En utilisant la loi de Bayes et l'équation de Chapman-Kolmogorov, on obtient une décomposition de la densité de probabilité [6] :

$$P(X_t | Y_0^t) = \frac{\int_{X_0^{t-1}} \prod_{\tau=1}^t P(X_\tau | X_{\tau-1}) P(X_0) \prod_{\tau=1}^t P(Y_\tau | X_\tau) dX_0^{t-1}}{\int_{X_0^t} \prod_{\tau=1}^t P(X_\tau | X_{\tau-1}) P(X_0) \prod_{\tau=1}^t P(Y_\tau | X_\tau) dX_0^t} \quad (7)$$

En résumé, les processus de Markov mesurés sont entièrement définis par :

$$\begin{cases} P(X_\tau | X_{\tau-1}) & \text{loi de transition} \\ P(Y_\tau | X_\tau) & \text{loi d'observation} \end{cases}$$

Cette remarque constitue la base de l'approche particulière des équations de filtrage. Elle repose sur une représentation de l'espace de probabilité par N mesures de Dirac $p_t^i \delta_{X^i}(dX)$ dont les supports X_t^i et les poids p_t^i sont conditionnés par les mesures.

On doit maintenant détailler l'évolution des particules dans l'espace de probabilité et le calcul des poids associés. Ceci est réalisé selon les transformations données précédemment : l'évolution est donnée par la loi de transition et la pondération par la loi d'observation. La solution particulière peut être principalement scindée en deux parties :

1. une étape d'évolution où chaque particule X_t^i de l'espace d'état évolue selon le flot du système (éq. 6a), via N tirages aléatoires suivant la loi $P(\pi_t)$:

$$X_t^i = F(X_{t-1}^i, \pi_t^i) \quad \text{pour } i = \{1, \dots, N\} \quad (8)$$

2. une étape de pondération où les poids p_t^i sont calculés à l'aide des mesures disponibles jusqu'à l'instant t (éq. 6b) :

$$p_t^i = p_{t-1}^i \frac{P(Y_t | X_t^i)}{\sum_{j=1}^N P(Y_t | X_t^j)} \quad (9)$$

L'estimée est alors calculée par : $\hat{X}_t = \sum_{i=1}^N p_t^i X_t^i$

4 Estimation du mouvement et de la structure 3D

Ce problème d'estimation multi-capteurs est résolu par un filtre qui traite directement les mesures issues de chaque capteur. Celui-ci repose sur une description point/segment de l'objet et délivre une estimation de la position 3D des extrémités de chacun de ses segments ainsi que des paramètres du mouvement rigide. La structure du filtre peut être résumé comme suit.

4.1 Initialisation

Les N particules sont initialisées selon l'information dont on dispose pour $i = \{1, \dots, N\}$:

$$\Omega_0^i [(\tilde{\Theta}_0^1)^i, (\tilde{\Theta}_0^2)^i, (\alpha_0^X)^i, (\alpha_0^Y)^i, (\alpha_0^Z)^i, (T_0^X)^i, (T_0^Y)^i, (T_0^Z)^i]^T$$

4.2 Evolution et mise en correspondance

Les particules évoluent selon l'équation de la dynamique (éq. 3) via N réalisations indépendantes W_t^i de loi $P(W_t^i)$:

$$\Omega_{t+1}^i = F_t \Omega_t^i + W_t^i \quad \text{pour } i = \{1, \dots, N\} \quad (10)$$

Dans ce problème d'estimation, les n mesures sont prétraitées pour délivrer les primitives de type segment/points. Ceci peut générer un nombre important de candidats potentiels à poursuivre dans les séquences de données.

Dans ce contexte de suivi "multi-cibles", on doit identifier la mesure la plus vraisemblable sur chaque capteur. Il existe pour cela de nombreuses méthodes comme, par exemple, l'association probabiliste de données (PDAF), les techniques de suivi multi-hypothèses (MHT) [1], la minimisation d'une fonction de similitude [4], etc. Nous utilisons ici la distance de Mahalanobis comme critère de mise en correspondance, pour associer les données dans les différentes séquences. Les mesures prédites sont calculées pour chaque particule sur les deux sources, en utilisant les équations de mesure (éq. 4), pour $i = \{1, \dots, N\}$:

$$\begin{cases} (\widehat{m}_{t+1}^R)^i = h_R(\Omega_{t+1}^i) \\ (\widehat{m}_{t+1}^I)^i = h_I(\Omega_{t+1}^i) \end{cases} \quad (11)$$

Les distances de Mahalanobis entre les mesures prédites et les mesures disponibles à l'instant $t+1$ sont calculées pour chaque segment de l'objet et le correspondant le plus probable est ainsi détecté sur chaque source j , pour $j = \{R, I\}$. $m_{t+1}^{j,k}$ correspond à la $k^{\text{ème}}$ mesure potentielle ($k = \{1, \dots, n(j)\}$) à l'instant $t+1$ sur la source j .

4.3 Pondération

Les poids normalisés p_t^i sont corrigés par les n mesures mises en correspondance m_{t+1}^{j,k^*} , pour $i = \{1, \dots, N\}$:

$$p_{t+1}^i = \frac{\prod_{j=1}^n P(m_{t+1}^{j,k_j^*} | \Omega_{t+1}^i)}{\sum_{i=1}^N \prod_{j=1}^n P(m_{t+1}^{j,k_j^*} | \Omega_{t+1}^i)} p_t^i \quad (12)$$

où $P(m_{t+1}^{j,k_j^*} | \Omega_{t+1}^i)$ est la loi d'observation de la source j .

4.4 Estimation

L'estimation finale est alors calculée comme suit :

$$\widehat{\Omega}_{t+1} = \sum_{i=1}^N p_{t+1}^i \Omega_{t+1}^i \quad (13)$$

Pour un traitement dynamique, on itère les étapes 2 à 4.

On peut donc remarquer (éqs. 10-11) que le filtre particulaire ne procède à aucune linéarisation des équations d'état, contrairement au filtre de Kalman étendu. De plus, il fournit une structure parallèle de traitement et permet une intégration naturelle de tout type de capteur.

5 Résultats expérimentaux

L'algorithme proposé est appliqué à une séquence d'images réelles issue de la banque de données de profondeur de

l'Université de Floride¹. Les images de profondeur et de réflexion sont délivrées par une caméra laser de profondeur *Odetics* fixe dont la résolution est 128×128 pixels. La scène 3D contient un objet mobile qui peut être décrit par un modèle polygonal. Les figures 2 présentent des images de profondeur et de réflexion à différents instants de la séquence : l'objet polyédrique évolue suivant un mouvement de translation pure durant 12 images. L'amplitude de la translation entre chaque instant est d'environ 9%.

5.1 Estimation du mouvement rigide 3D

Les figures 1 montrent l'estimation des paramètres du mouvement rigide (R_t, \vec{T}_t) pour l'un des segments de l'objet. Le filtre particulaire délivre ici une bonne estimation (erreur $< 10^{-3}$) de ce mouvement uniforme dont il identifie les caractéristiques à partir de quelques images.

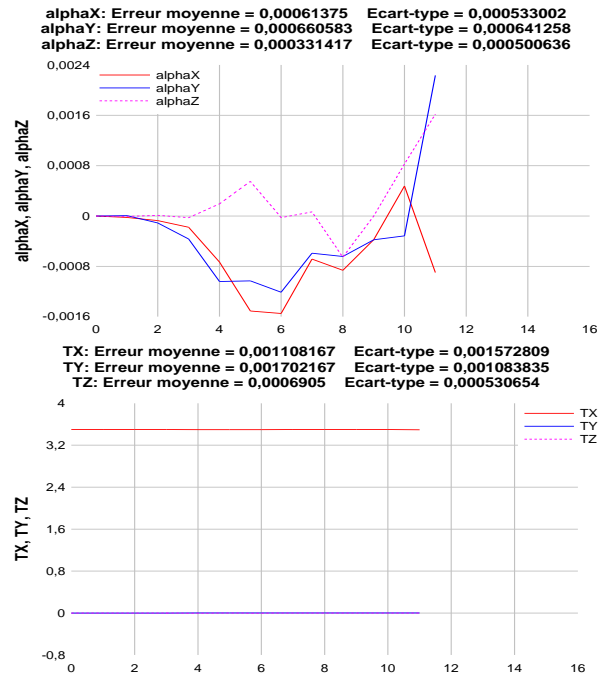


FIG. 1 – Estimation du mouvement rigide 3D

5.2 Estimation des positions 3D

Les figures 2 montrent l'estimation de la position d'une des extrémités du même segment comparée à sa trajectoire mesurée. Les faibles erreurs moyennes et écart-types sur chaque coordonnées 3D permettent d'attester de la précision de la méthode dans un cas réel d'étude.

Finalement, les figures 3 montrent les scènes 3D reconstruites en utilisant l'ensemble des segments suivis, respectivement projetés dans les images de profondeur et de réflexion via les relations 5. Les cercles définissent les mesures disponibles, les segments en trait plein sont les projections des données 3D estimées, les carrés sont les mesures utilisées pour mettre à jour l'estimation des positions 3D. Les segments en traits pointillés indiquent que le filtre a fonctionné uniquement en prédiction et qu'il n'y a pas eu d'association de mesure.

¹Disponible à marathon.csee.usf.edu/range/DataBase.html



FIG. 2 – Estimation de la position 3D d’une extrémité

6 Conclusion

Cet article montre l’apport de la méthode particulière en estimation conjointe de la structure et du mouvement d’objets 3D rigides. L’intérêt réside dans son aptitude à traiter des modèles non-linéaires et des statistiques non-gaussiennes. La solution repose sur un filtre global qui fusionne les informations issues de chaque source et permet une amélioration des précisions en estimation par rapport à une méthode de type Kalman étendu [2].

Des résultats sont présentés dans le cas d’un système de vision composé d’une caméra laser de profondeur mais la méthode peut être aisément étendue à d’autres types de capteur, sans modification profonde de la solution.

Références

[1] Y. Bar-Shalom and X. R. Li. *Multitarget Multisensor Tracking*. YBS Publication, 1995.

[2] C. Boucher, J.-C. Noyer, and M. Benjelloun. 3D structure and motion recovery by fusing range and intensity image sequences. In *Proceedings of the 3rd International Conference on Information Fusion*, Paris, France, July 2000.

[3] R. Deriche. Using Canny’s criteria to derive an optimal edge detector recursively implemented. *The Interna-*

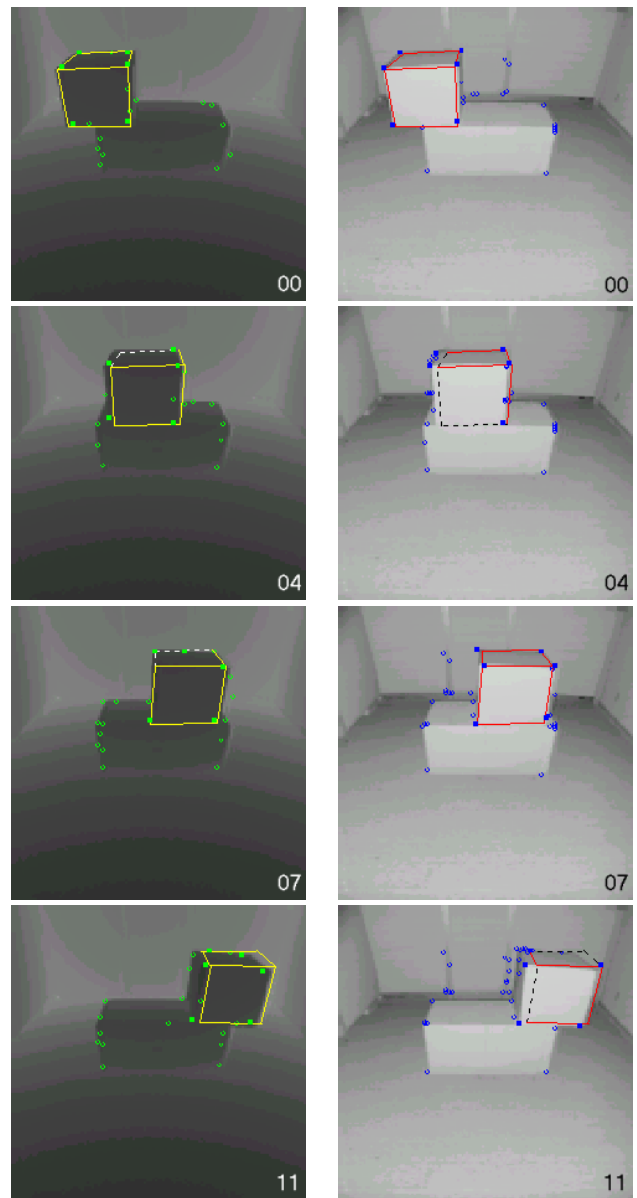


FIG. 3 – Scènes reconstruites et projetées

tional Journal of Computer Vision, April 1987.

[4] R. Deriche and O. Faugeras. Tracking line segments. *Image and Vision Computing*, 8(4) :261–270, 1990.

[5] A. Hoover et al. An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7) :673–689, July 1996.

[6] P. Del Moral, J.-C. Noyer, G. Rigal, and G. Salut. Résolution particulière et traitement non-linéaire du signal : Applications Radar/Sonar. *Traitement du Signal*, 12(4) :287–301, 1995.

[7] P. Del Moral and G. Salut. Particle interpretation of non-linear filtering and optimization. *Russian Journal of Mathematical Physics*, 5(3) :355–372, 1997.

[8] S. M. Smith and J. M. Brady. SUSAN : A new approach to low level image processing. *International Journal of Computer Vision*, 23(1) :45–78, May 1997.

[9] K. Storjohann. Laser range camera modeling. Technical Report ORNL/TM-11530, ORNL, 1990 1990.