Filtrage particulaire pour la capture de mouvement dédiée à l'interaction homme-robot

M. Fontmarty[†], F. Lerasle[†‡], P. Danès[†‡], P. Menezes[¶]

[†] LAAS-CNRS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4
[‡] Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cédex
[¶] ISR/DEEC, Université de Coimbra - Polo II, 3030-290 Coimbra
e-mail :{mfontmar, lerasle, danes}@laas.fr, paulo@deec.uc.pt

Résumé

Cet article traite du suivi visuel de structure 3D articulée à partir d'un système binoculaire embarqué sur un robot mobile en environnement humain, a priori *encombré. Une telle structure est définie par un ensemble de cônes tronqués reliés par des articulations admettant un ou plusieurs degrés de liberté. Le filtrage particulaire et notamment l'échantillonnage partitionné sont adaptés pour estimer la configuration de la structure articulaire compatible avec les deux vues. Cette stratégie ainsi que la* CONDENSATION *classique sont implémentées et comparées dans notre contexte applicatif. Ces deux stratégies fusionnent, dans leur modèle de mesure, des attributs hétérogènes tels que des mesures d'apparence (couleur, contours) et des mesures 3D issues d'une reconstruction éparse. L'apport de l'échantillonnage partitionné est mis en exergue aux travers de quelques évaluations quantitatives ou qualitatives. Les extensions de ces travaux sont discutées en final.*

Mots Clef

Robotique mobile, vision binoculaire, structures articulées, suivi 3D, fusion de données, filtrage particulaire.

Abstract

This paper deals with visual tracking of 3D articulated structure from a binocular system mounted on a mobile robot in a human – a priori *cluttered* – *environment. Such a structure is composed of a set of truncated cones connected by joints admitting one or more degrees of freedom. Particle filtering and especially the partitioned sampling strategy is well suited to estimate the 3D model configuration which is consistent with the two perceived images. This scheme as well as the well-known* CONDENSATION *algorithm are implemented and compared in our context. These two strategies noticeably fuse, in their measurement model, heterogeneous cues such as appearance measures (color, edges) and 3D measures issued from a sparse reconstruction. The benefit of the partitioned sampling is highlighted through some quantitative and qualitative evaluations. The extensions of this work are discussed at the end.*

Keywords

Mobile robotics, binocular vision, articulated structures, 3D tracking, data fusion, particle filtering.

1 Introduction

Durant ces dernières années, de nombreux travaux dans la communauté Vision pour la Robotique ont porté sur l'interaction entre l'homme et la machine. Un défi majeur de la Robotique est sans doute celui du robot personnel avec la perspective de voir un robot interagir à distance avec ses interlocuteurs par la reconnaissance de gestes, notamment pour le pointage et la manipulation d'objets (figure 1).

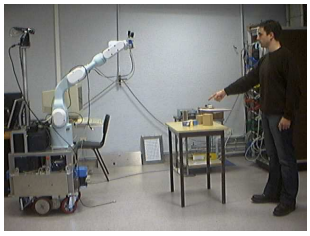


FIG. 1 – Gestes déictiques pour le robot personnel Jido.

Le synoptique d'un système de reconnaissance de gestes distingue classiquement une phase d'analyse — suivi ou capture de mouvement — et une phase d'interprétation qui permet l'identification du geste effectué parmi une base de gestes connus. Cette dernière phase sortant du cadre de cet article, le lecteur pourra se référer à [12] pour une synthèse. L'objectif est ici d'estimer dans le flot vidéo la variation des paramètres de position/orientation de tout ou partie des membres corporels observés.

De nombreux travaux insistent sur l'intérêt de gérer à chaque instant plusieurs hypothèses sur ces paramètres à estimer [5, 16, 20]. Le filtrage particulaire, abondamment référencé dans ce contexte [5, 10, 11, 15, 20], s'affranchit de toute hypothèse restrictive quant aux distributions de probabilité entrant en jeu dans

la caractérisation du problème. Néanmoins, le nombre parfois prohibitif de particules nécessaires à l'exploration d'un espace d'état de grande dimension peut constituer un obstacle à sa mise en œuvre en temps réel. L'échantillonnage partitionné a été introduit dans le cadre du suivi 2D d'objets articulés en vue de pallier ce problème [10]. Son principe, présenté section 2, nous semble transposable au suivi 3D des membres corporels. Par ailleurs, le filtrage particulaire permet, dans un cadre théorique étayé, de fusionner aisément différents types de mesures. Or notre robot est censé évoluer dans des environnements *a priori* encombrés et sujets à des changements d'illumination. Il est alors opportun d'exploiter plusieurs sources de mesures.

A l'instar de nombreux travaux [4, 5, 16, 20], nous modélisons le corps humain comme un ensemble de parties rigides. La littérature distingue alors les approches multi-oculaires basées sur une reconstruction 3D [4, 19] des approches monoculaires majoritairement basées sur l'apparence du modèle après projection image [5, 11, 15, 16, 20]. Par le passé, nous avons privilégié une telle stratégie, légère et simple à mettre en œuvre (une seule caméra étalonnée) [11]. La plus grande difficulté résidait alors dans l'estimation des mouvements non fronto-parallèles au plan image. Les développements actuels reprennent et étendent ces travaux à un système stéréoscopique. L'approche retenue se démarque de la classification énoncée au sens où elle repose, comme dans [3], sur des mesures d'apparence mais également sur des mesures/contraintes géométriques tirant partie d'une reconstruction 3D éparse.

L'article est structuré comme suit. La section 2 rappelle sommairement le formalisme bien connu du filtrage particulaire puis caractérise la stratégie d'échantillonnage partitionné. La section 3 décrit notre modèle 3D et sa projection perspective. La section 4 spécifie les diverses mesures impliquées. La section 5 décrit l'implémentation de notre filtre partitionné. Dans la section 6, les stratégies d'échantillonnage partitionné et SIR sont évaluées et comparées dans notre contexte applicatif. Enfin, la section 7 résume notre contribution et propose quelques extensions envisagées.

2 Filtrage particulaire

2.1 Un algorithme générique

Les techniques de filtrage particulaire sont des méthodes de Monte Carlo pour l'estimation récursive du vecteur d'état d'un système stochastique Markovien [2, 6, 7]. Leur but est d'approximer la densité de probabilité *a posteriori* $p(x_k|z_{1:k})$ du vecteur d'état x_k à l'instant k conditionnellement aux mesures $z_{1:k} = z_1, \ldots, z_k$, par une distribution ponctuelle

$$p(x_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)}),\ \sum_{i=1}^{N} w_k^{(i)} = 1, \quad (1)$$

exprimant la sélection d'une valeur – ou « particule » – $x_k^{(i)}$ avec la probabilité – ou « poids » – $w_k^{(i)}$, $i = 1, \ldots, N$. Une approximation de l'espérance *a posteriori* d'une fonction quelconque de x_k, *e.g.* l'estimé du minimum d'erreur quadratique moyenne (MMSE) $\mathrm{E}[x_k|z_{1:k}]$, se déduit immédiatement.

Soit le système de loi de dynamique $p(x_k|x_{k-1})$ et de densité d'observation $p(z_k|x_k)$. L'algorithme de filtrage particulaire générique – ou "Sampling Importance Resampling" (SIR) – est présenté Table 1. Son initialisation consiste en une séquence indépendante identiquement distribuée (i.i.d.) selon la distribution *a priori* $p(x_0)$. Aux instants ultérieurs $k \geq 1$, chaque particule $x_k^{(i)}$ est échantillonnée selon une *fonction d'importance* $q(x_k|x_{k-1}^{(i)}, z_k)$ visant à explorer adaptativement les zones « pertinentes » de l'espace d'état. Afin de garantir la cohérence de l'approximation (1), $x_k^{(i)}$ est affectée d'un poids $w_k^{(i)}$ dépendant de sa *vraisemblance* $p(z_k|x_k^{(i)})$ par rapport à la mesure z_k ainsi que des évaluations de la loi de dynamique et de la fonction d'importance (étape 5).

Toute méthode séquentielle de type Monte Carlo souffre du phénomène de dégénérescence, selon lequel une seule particule concentre rapidement l'essentiel des poids. C'est pourquoi l'algorithme SIR inclut dans son étape 8 un rééchantillonnage du nuage, *e.g.* le "rééchantillonnage systématique" défini en [9]. Ainsi, les particules affectées de poids élevés sont dupliquées, au détriment de celles, faiblement pondérées, qui disparaissent. Notons que cette étape de redistribution ne doit être déclenchée que lorsque l'efficacité du filtre – liée au nombre de particules "utiles" – se situe en deçà d'un certain seuil [7, 2]. Le calcul des moments de (1) doit de préférence faire intervenir l'ensemble de particules pondérées avant rééchantillonnage (étape 7).

2.2 Éléments de définition d'une stratégie de filtrage

Complexité de l'algorithme Une première considération à prendre en compte dans le contexte de la robotique est bien sûr la complexité calculatoire de l'algorithme. La dynamique du système $p(x_k|x_{k-1})$ est généralement une loi relativement simple, *e.g.* une Gaussienne permettant de modéliser une marche aléatoire, un mouvement à vitesse constante, etc. (cf. § 5). Il s'en suit que la stratégie la moins coûteuse est l'instantiation de l'algorithme SIR au cas où $q(x_k|x_{k-1}^{(i)}, z_k) = p(x_k|x_{k-1}^{(i)})$, introduite pour la première fois dans [8] sous le vocable CONDENSATION – pour "Conditional Density Propagation". En effet, non seulement l'échantillonnage

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N = \mathrm{SIR}([\{x_{k-1}^{(i)}, w_{k-1}^{(i)},\}]_{i=1}^N, z_k)$

1: **SI** $k=0$**, ALORS** Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$ **FIN SI**

2: **SI** $k \geq 1$ **ALORS** $\{$—$[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^N$ désignant une approximation particulaire de $p(x_{k-1}|z_{1:k-1})$—$\}$

3: **POUR** $i = 1, \ldots, N$, **FAIRE**

4: « Propager » la particule $x_{k-1}^{(i)}$ en simulant de manière indépendante $x_k^{(i)} \sim q(x_k|x_{k-1}^{(i)}, z_k)$

5: Mettre à jour le poids $w_k^{(i)}$ associé à $x_k^{(i)}$ selon $w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_k)}$, (le symbole "$\propto$" signifiant "préalablement à une étape de normalisation assurant que $\sum_i w_k^{(i)} = 1$")

6: **FIN POUR**

7: Calculer l'espérance *a posteriori* d'une fonction quelconque de x_k, *e.g.* $\mathrm{E}_{p(x_k|z_{1:k})}[x_k]$, à partir de l'approximation $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ de $p(x_k|z_{1:k})$

8: À chaque instant ou selon un critère d'« efficacité », redistribuer la description $[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$ de $p(x_k|z_{1:k})$ en l'ensemble équivalent $[\{x_k^{(s^{(i)})}, \frac{1}{N}\}]_{i=1}^N$, en échantillonnant avec remise les indices $s^{(1)}, \ldots, s^{(N)}$ dans $\{1, \ldots, N\}$ selon $P(s^{(i)} = j) = w_k^{(j)}$; affecter $x_k^{(i)}$ et $w_k^{(i)}$ avec $x_k^{(s^{(i)})}$ et $\frac{1}{N}$

9: **FIN SI**

TAB. 1 – Algorithme de filtrage particulaire SIR.

des particules selon la dynamique ne pose pas de problème majeur, mais l'étape de mise à jour des poids se simplifie trivialement en $w_k^{(i)} \propto w_{k-1}^{(i)} p(z_k|x_k^{(i)})$.
Toutefois, choisir la dynamique du système comme fonction d'importance peut s'avérer inadapté. Pour peu que les modes de la fonction de vraisemblance $p(z_k|.)$ soient très prononcés, de nombreuses particules $x_k^{(i)}$, échantillonnées sans considérer la mesure z_k, risquent d'être affectées d'un poids faible à l'issue de l'étape 5 de l'algorithme SIR. Le rééchantillonnage entraîne alors une perte de diversité dans l'exploration de l'espace d'état, ce qui revient au phénomène de dégénérescence précédemment mentionné.

Efficacité d'un filtre particulaire Comme cela a été mentionné dans les parties précédentes, un filtre est intuitivement convenable si toutes les particules sont associées à des poids suffisamment élevés. Ce problème se formalise rigoureusement sur le plan théorique en terme de comportement statistique, *cf.* par exemple l'analyse de [7].
Considérons l'algorithme SIR auquel on supprime l'étape 8 de rééchantillonnage. La variance d'un estimé calculé dans l'étape 7 – qui est asymptotiquement non biaisé lorsque le nombre de particules croît indéfiniment – est fonction du choix de la fonction d'importance, au sens où elle est d'autant plus faible que les poids sont peu dispersés. On définit ainsi un scalaire inférieur à N, appelé *nombre de particules efficaces*, inversement proportionnel à la variance de l'estimé. L'expression analytique de ce scalaire ne peut pas être évaluée numériquement, cependant il est possible de l'estimer à tout instant k au moyen de la quantité

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} (w_k^{(i)})^2},$$

qui vaut N lorsque les particules sont réparties uniformément, et $\frac{1}{N}$ lorsque toutes sauf une admettent un poids nul. Le rééchantillonnage peut être déclenché lorsque N_{eff} se situe en deçà d'un certain seuil, *e.g.* $\frac{N}{3}$.

2.3 Échantillonnage partitionné

L'échantillonnage partitionné peut améliorer significativement l'efficacité d'un filtre particulaire si la dynamique du système se présente comme l'application successive de dynamiques élémentaires dès lors que des vraisemblances intermédiaires du vecteur d'état peuvent être définies après application de chaque modèle d'évolution partiel. Le nuage de particules est alors géré au moyen d'une stratégie stratifiée : grâce à une succession d'échantillonnages suivis de rééchantillonnages basés sur les vraisemblances intermédiaires, l'exploration de l'espace d'état peut être guidée de façon que chaque itération raffine le nuage résultant de l'itération précédente.
Pour présenter la seconde stratégie de filtre particulaire partitionné proposée dans [10], supposons que le vecteur d'état x_k à tout instant k puisse être partitionné de manière analogue en M sous-vecteurs $x_k^1, \ldots, x_k^M$, et que la loi de dynamique s'écrive $p(x_k|x_{k-1}) = \prod_{m=1}^M p(x_k^m|x_{k-1}^m)$. Le lien état-mesure $p(z_k|x_k)$ doit quant à lui pouvoir se factoriser en $p(z_k|x_k) = \prod_{m=1}^M p_m(z_k|x_k)$, où les vraisemblances intermédiaires $p_m(z_k|x_k)$ sont de la forme $p_m(z_k|x_k) = l_m(z_k|x_k^1, \ldots, x_k^m)$, *i.e.* concernent un sous-ensemble du vecteur d'état d'autant plus important que $m \longrightarrow M$. Le filtre suit alors l'algorithme décrit succintement Table 2.
L'échantillonnage partitionné a été appliqué avec succès au suivi visuel d'une chaîne cinématique ouverte dans [10], en constituant le vecteur d'état de telle sorte que ses premières composantes paramètrent les éléments de début de chaîne – à positionner d'abord pour plus d'efficacité –, les dernières composantes étant reliées aux extrémités. Un algorithme ramifié permet le suivi de plusieurs personnes dans [10].
Signalons également l'existence du filtre particulaire hiérarchisé [13], qui généralise le schéma partitionné

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$= PARTITIONNE $\left([\{x_{k-1}^{(i)}, w_{k-1}^{(i)},\}]_{i=1}^N, z_k\right)$

1: **SI** $k=0$**, ALORS** Échantillonner $x_0^{(1)}, \ldots, x_0^{(i)}, \ldots, x_0^{(N)}$ i.i.d. selon $p(x_0)$, et poser $w_0^{(i)} = \frac{1}{N}$ **FIN SI**
2: **SI** $k \geq 1$ **ALORS** —$[\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^N$ désignant une approximation particulaire de $p(x_{k-1}|z_{1:k-1})$—
3: Poser $\{\xi_0^{(i)}, \tau_0^{(i)}\} = \{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}$
4: **POUR** $m = 1, \ldots, M$, **FAIRE**
5: **POUR** $i = 1, \ldots, N$, **FAIRE** Échantillonner de manière indépendante $\xi_m^{(i)} \sim \tilde{q}_m(\xi_m|\xi_{m-1}^{(i)})$, avec $\tilde{q}_m(\xi_m|\xi_{m-1}) = p(\xi_m^m|\xi_{m-1}^m) \prod_{r \neq m} \delta(\xi_m^r - \xi_{m-1}^r)$ –*i.e.* « propager » la $m^{\text{ième}}$ partition $x_{k-1}^{(i)\,m}$ de la particule $x_{k-1}^{(i)}$ selon la dynamique élémentaire $p(x_k^m|x_{k-1}^{(i)\,m})$– et associer à $\xi_m^{(i)}$ le poids $\tau_m^{(i)} \propto \tau_{m-1}^{(i)} p_m(z_k|\xi_m^{(i)})$, où $p_m(z_k|\xi_m) = l_m(z_k|\xi_m^1, \ldots, \xi_m^m)$ **FIN POUR**
6: Redistribuer $[\{\xi_m^{(i)}, \tau_m^{(i)}\}]_{i=1}^N$ en l'ensemble équivalent $[\{\xi_m^{(s^{(i)})}, \frac{1}{N}\}]_{i=1}^N$; renommer $[\{\xi_m^{(s^{(i)})}, \frac{1}{N}\}]_{i=1}^N$ en $[\{\xi_m^{(i)}, \tau_m^{(i)}\}]_{i=1}^N$
7: **FIN POUR**
8: Poser $\{x_k^{(i)}, w_k^{(i)}\} = \{\xi_M^{(i)}, \tau_M^{(i)}\}$, qui est une description cohérente de $p(x_k|z_{1:k})$
⋮
(...) : **FIN SI**

TAB. 2 – Algorithme de filtrage particulaire PARTITIONNE.

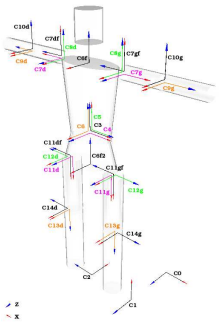


FIG. 2 – Modèle du corps humain et repères associés à chaque membre.

au cas où les partitions du vecteur d'état sont échantillonnées selon des fonctions d'importance plus générales que leur loi de dynamique.

3 Modélisation des membres corporels

3.1 Modèle cinématique

Le modèle complet utilisé pour le suivi est présenté figure 2. Il comprend 22 degrés de liberté, mais nous limitons actuellement le suivi à la moitié supérieure du modèle (le torse, la tête et les bras), qui constitue un ensemble de deux chaînes cinématiques ouvertes à 14 degrés de liberté :

– 6 affectant la position et l'orientation globale (translations t_i et rotations $r_i, i \in \{1,2,3\}$),
– une liaison rotule pour chaque épaule (2×3 rotations r_{bdi} et $r_{bgi}, i \in \{1,2,3\}$),
– 2×1 rotations pour les coudes (r_{abd} et r_{abg}).

Nous ne prenons pas en compte les degrés de liberté relatifs au poignet, et nous relions rigidement la tête et le torse. Les 6 paramètres de placement et d'orientation globale fixent ainsi la configuration du torse et de la tête dans l'espace. Nous utilisons les conventions de Denavit-Hartenberg modifiées pour la paramétrisation du modèle, qui comporte ainsi 3 liaisons prismatiques et 11 liaisons rotoïdes dont les plages de valeurs sont limitées aux intervalles physiquement compatibles avec les mouvements humains. Le vecteur de configuration du modèle dans l'espace articulaire à un instant k s'écrit :

$$\mathbf{q}_k = [t_1, t_2, t_3, r_1, r_2, r_3, r_{bd1}, r_{bg1}, r_{bd2}, r_{bg2}, r_{bd3}, r_{bg3}, r_{abd}, r_{abg}]'.$$

3.2 Modèle géométrique

Notre modèle 3D articulé est constitué de cônes tronqués. Ces primitives géométriques restent assez simples à manipuler tandis que la géométrie projective en permet, de façon élégante, la projection image et la gestion des occultations. Enfin, ces primitives 3D restent assez proches géométriquement des structures 3D modélisées. Chaque cône tronqué, relativement aux membres supérieurs du corps humain, est représenté par quatre paramètres : les rayons r_a et r_b de la face elliptique de base, la longueur l du cône tronqué, le rapport α de l'ellipse extrémité par rapport à l'ellipse de base.

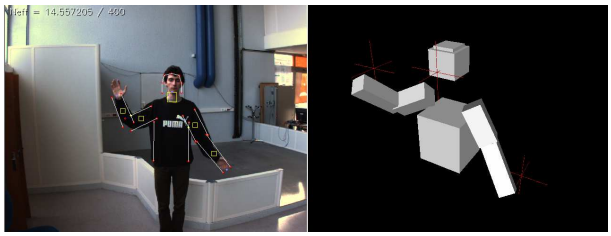
Des régions d'intérêt (ROIs) sont également associées à des points surfaciques du modèle 3D. Ces ROIs sont par exemple des distributions locales de couleur ou de texture, constituant des marqueurs naturels à suivre dans le flot vidéo.

3.3 Projection du modèle et gestion des occultations

Rappelons que notre approche consiste à recaler au mieux les limbes projetées du modèle sur les contours réels pour inférer la position 3D. Sous hypothèse du modèle sténopé, ces limbes sont projetées sous forme de segments (figure 3-(a)) selon la méthode utilisée par Azad dans [3].

La représentation du modèle sur la figure 3-(b) par des parallépipèdes est uniquement dédiée à l'animation afin de mieux visualiser les angles de rotation de chacun des corps (en particulier les épaules).

Les différentes parties du modèle peuvent donner lieu à des occultations partielles ou totales. Il faut donc



(a) (b)

FIG. 3 – Limbes du modèle et ROIs projetées (en jaune) (a), configuration 3D du modèle associée (b).

vérifier la visibilité des segments de droite associés obtenus par projection perspective. Le choix de primitives géométriques de type conique prend ici tout son sens car la complexité de l'algorithme de gestion des parties cachées dépend alors uniquement du nombre de ces primitives et non de la taille image des parties projetées ni de la précision requise pour leur projection [17]. Le lecteur pourra se référer à [11] pour plus de détails sur cet algorithme.

4 Modèle d'observation

Le modèle d'observation global fusionne des mesures liées à l'apparence (contours image, distributions locales de couleur/texture) mais aussi des mesures/contraintes 3D obtenues par triangulation entre les deux vues. Sous l'hypothèse d'indépendance de ces sources de mesures, le modèle d'observation à l'instant k s'écrit :

$$p(z_k^f, z_k^c, z_k^{3d}|x_k) = p(z_k^f|x_k).p(z_k^c|x_k).p(z_k^{3d}|x_k)$$

où les trois densités d'observation élémentaires sont caractérisées ci-dessous.

4.1 Modèle de mesure sur les contours

Le principe est d'échantillonner uniformément en N_p points $p_j, j \in \{1, \ldots, N_p\}$ les segments de droite correspondant aux limbes projetées du modèle pour une configuration x_k. Nous définissons alors une image de distance I_{DT} à partir des contours image extraits. Celle-ci permet d'évaluer à faible coût le critère pour chacune des N particules ($N \gg 100$). Ce modèle de mesure s'écrit :

$$p(z_k^f|x_k) \propto \exp\left(-\frac{D^2}{2\sigma_f^2}\right), \; D = \frac{1}{N_p}\sum_{j=1}^{N_p} I_{DT}(p_j),$$

où σ_f est un paramètre prédéfini. Cette mesure reste très sensible aux conditions d'éclairement *a priori* quelconques et surtout peu discriminante en présence de scènes encombrées.

4.2 Modèle de mesure sur la couleur

Ce modèle est relatif aux N_m ROIs attachées au modèle 3D. Ces ROIs sont caractérisées après projection du modèle par des distributions locales de couleur dans l'image. Le modèle de mesure colorimétrique $p(z_k^c|x_k)$ s'écrit :

$$p(z_k^c|x_k) \propto \exp\left(-\frac{D^2}{2\sigma_c^2}\right), D = \frac{1}{3N_m}\sum_c \sum_{j=1}^{N_m} D_B(h_{x_k,j}^c, h_{ref,j}^c)$$

où $c \in \{R, G, B\}$, σ_c est un paramètre prédéfini, et D_B est la distance de Bhattacharyya [1] utilisée pour comparer les histogrammes normalisés $(h_{ref,j}^c, h_{x_k,j}^c)$, $j = 1, \ldots, N_m$. L'apparence de ces ROIs, *i.e.* les histogrammes $h_{ref,j}^c$, est apprise sur la première image de la séquence. Notons enfin que ce modèle est transposable à des distributions locales de texture où l'index c référence alors le seul plan intensité.

4.3 Modèle de mesure sur le 3D

Ce modèle repose sur la segmentation en régions peau dans les deux vues, l'appariement entre ces régions extraites par les critères définis dans [14] (chapitre 4), enfin la triangulation pour caractériser des ellipsoïdes d'inertie censées inclure les mains et la tête dans l'espace.
Soient deux régions peau appariées entre vues gauche/droite et indicées par j ($j = 1, \ldots, N_r$). Chaque ellipsoïde 3D associée est paramétré par : (i) son centre de gravité $\widehat{P}_j = (X_j, Y_j, Z_j)'$ obtenu par la fonction de triangulation $f(.)$, (ii) sa matrice de covariance $\widehat{Cov}_j$ telle que :

$$\widehat{Cov}_j = F_j^g.cov_j^g.F_j^{g'} + F_j^d.cov_j^d.F_j^{d'},$$

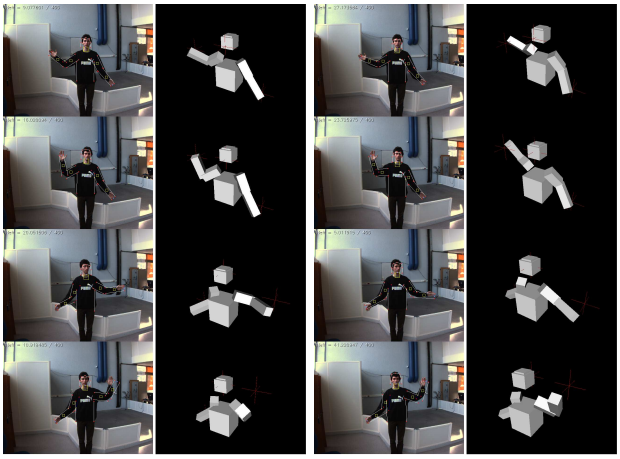
où cov_j^g et cov_j^d sont les matrices d'inertie image déduites des moments d'ordre 2, F_j^g et F_j^d sont les Jacobiennes relatives à la fonction $f(.)$. Le modèle de mesure 3D s'écrit :

$$p(z_k^{3d}|x_k) \propto \exp\left(-\frac{D^2}{2\sigma_{3d}^2}\right), D = \sum_{i=1}^{3} D_M(P_{x_k,i}, \widehat{P}_{j_i})$$

où $D_M(P_{x_k,i}, \widehat{P}_{j_i})$ représente la distance de Mahalanobis entre le centre de gravité $\widehat{P}_{j_i}$ de l'ellipsoïde j_i ($j_i \in \{1, \ldots, N_r\}$) et $P_{x_k,i}$ la position spatiale d'une main ou de la tête du modèle sous l'hypothèse x_k. L'association entre i et j_i est basée sur une heuristique simple mettant en jeu la position des ellipsoïdes 3D.

5 Implémentation du filtre à échantillonnage partitionné

Le but de notre filtre à échantillonnage partitionné est d'estimer les 14 degrés de liberté constituant $\mathbf{q}_k$ à partir de chaque paire d'images dans les flots vidéo. Le système stéréoscopique est étalonné hors-ligne. Concernant le modèle de dynamique $p(x_k|x_{k-1})$, les mouvements spatiaux du sujet observé sont difficiles



(a) (b)

FIG. 4 – Déroulement d'une exécution d'un filtre SIR (a) et d'un filtre PARTITIONNE (b) sur une même séquence.

à caractériser *a priori*. Nous supposons ici que les composantes de l'état évoluent suivant des marches aléatoires indépendantes, soit

$$p(x_k|x_{k-1}) = \mathcal{N}(x_k; x_{k-1}, \Sigma)$$

où $\mathcal{N}(.;\mu,\Sigma)$ est une distribution Gaussienne de moyenne μ et de covariance $\Sigma = diag(\sigma_i^2, \ldots)$ telle que :

- $\sigma_i = 0.05\ m$ si $i \in \{1,2,3\}$ (translation de base)
- $\sigma_i = 0.002\ rad$ si $i \in \{4,5,6\}$ (rotation de base)
- $\sigma_i = 0.2\ rad$ sinon (autres rotations)

Le vecteur d'état global à estimer dans le filtre est donc : $x_k = \mathbf{q}_k$. La stratégie partitionnée consiste à décomposer le vecteur d'état x_k en sous-vecteurs, soit :

$$\begin{aligned} x_k^1 &= [t_1, t_2, t_3, r_1, r_2, r_3]^T \\ x_k^2 &= [r_{bd1}, r_{bg1}, r_{bd2}, r_{bg2}, r_{bd3}, r_{bg3}, r_{abd}, r_{abg}]^T \\ x_k &= [x_k^{1T}, x_k^{2T}]^T \end{aligned}$$

Ainsi, le principe est donc de positionner le torse *via* x_k^1 puis les bras et avant-bras *via* x_k^2. Nous utilisons cinq ROIs de distribution de couleur relativement aux avant-bras, au bras et au col. Les paramètres relatifs aux équations de dynamique et de mesure du filtre sont donnés dans la table 3. Une exécution de chaque filtre est présenté figure 4. De manière empirique, nous constatons une meilleure stabilité de la position du torse avec la stratégie PARTITIONNE ainsi qu'une configuration estimée plus proche de ce que l'on attendrait au regard de la projection image.

6 Étude comparative

6.1 Protocole d'évaluation

Indépendamment de ces évaluations qualitatives, nous avons évalué quantitativement notre filtre sur des séquences tests acquises depuis le robot Jido. Les performances de la stratégie partitionnée sont comparées à celles de la CONDENSATION classique.

Symbole	Définition	Valeur
(w,h)	résolution des images	$(640, 480)$
σ_f	paramètre dans $p(z_k^f \| x_k)$	3.3
σ_c	paramètre dans $p(z_k^c \| x_k)$	0.5
σ_{3d}	paramètre dans $p(z_k^{3d} \| x_k)$	3
N_m	nombre de ROIs colorimétriques	5
N	nombre de particules	$100 - 1000$

TAB. 3 – Valeurs des paramètres utilisés dans les filtres SIR et PARTITIONNE.

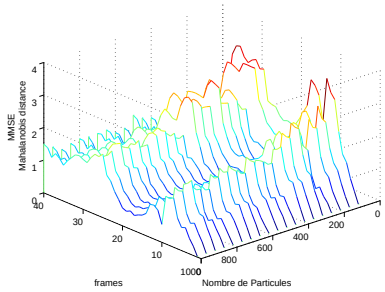


FIG. 5 – Erreur entre l'estimé fourni par la CONDENSATION et la vérité terrain : $d(x_{SIR,k}, x_{v,k})$.

Une base de séquences est dépouillée afin de caractériser le vecteur d'état pour chaque image et ainsi constituer une "vérité terrain" pour les évaluations ultérieures. La démarche est soit manuelle, soit semi-automatique. Sur ces séquences, nous évaluons alors les deux stratégies de filtrage, selon les critères définis dans [18] :

1. **la précision** *via* la distance de Mahalanobis $d(x_{f,k}, x_{v,k})$ entre l'état estimé $x_{f,k}$ par chaque filtre f à l'instant k et la vérité de terrain $x_{v,k}$;
2. **le nombre de particules efficaces** estimé N_{eff}, qui permet de caractériser la dispersion de la moyenne *a posteriori* calculée par le filtre ;
3. **le temps de traitement** car il représente un critère essentiel pour nos plateformes embarquant des ressources CPU limitées.

Les filtres sont initialisés à partir d'un tir gaussien autour de la vérité de terrain précédemment établie. Le nombre N de particules influence grandement les performances d'un filtre particulaire. Aussi, chaque stratégie est évaluée pour $100 < N < 1000$. Le caractère aléatoire du filtrage particulaire ne permettant pas de baser son évaluation sur une seule de ses réalisations, une étude statistique du comportement moyen du filtre est effectuée. Ainsi les critères sont calculés pour chaque stratégie à partir de 50 réalisations sur chaque séquence. La dispersion de N_{eff} est également évaluée afin de garantir les conclusions établies.

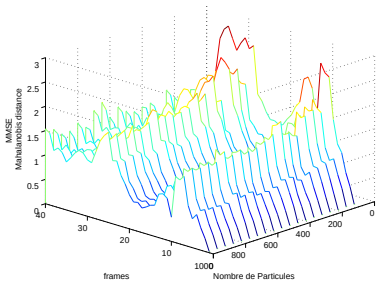


FIG. 6 – Erreur entre l'estimé fourni par le PARTITIONNE et la vérité terrain : $d(x_{Partitionne,k}, x_{v,k})$.

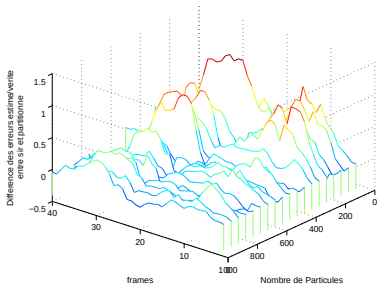


FIG. 7 – Différence entre les erreurs des estimés de chacun des filtres : $d(x_{SIR,k}, x_{v,k}) - d(x_{Partitionne,k}, x_{v,k})$.

6.2 Précision

Les figures 5 et 6 représentent respectivement les erreurs entre la verité terrain et les estimés fournis par les filtres particulaires CONDENSATION et PARTITIONNE pour la séquence présentée figure 4. D'une manière générale, l'augmentation du nombre de particules pour un type de filtre améliore sa précision ce qui est cohérent avec la diminution asymptotique du biais établie en théorie.
La différence $d(x_{SIR,k}, x_{v,k}) - d(x_{Partitionne,k}, x_{v,k})$ entre les erreurs d'estimation des deux filtres (figure 7) montre que le filtre PARTITIONNE est plus précis que la CONDENSATION pour un nombre de particules faible ($N \ll 400$ dans notre cas). Pour $N \gg 400$, les deux stratégies donnent logiquement lieu à des précisions sensiblement équivalentes. Dans notre contexte applicatif où les ressources CPU sont limitées, il est opportun de privilégier une stratégie de filtrage limitant autant que possible le nombre de particules, et donc le filtre PARTITIONNE.

6.3 Nombre de particules efficaces

L'observation de l'évolution du rapport des N_{eff} de chaque filtre (figure 8) nous permet de voir que le filtre partitionné place mieux les particules dans l'espace d'état ($N_{eff,Partitionne} \approx 1.5 \times N_{eff,SIR}$). Ainsi, pour un nombre de particules équivalent, la dispersion de l'estimé est moindre.

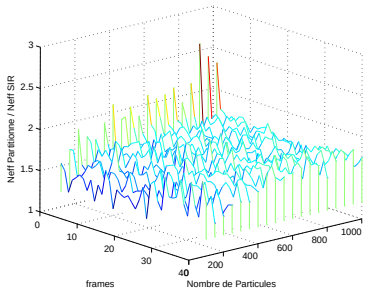


FIG. 8 – $\frac{N_{eff,Partionne}}{N_{eff,SIR}}$.

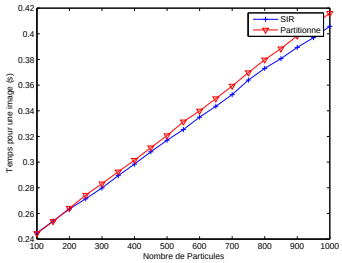


FIG. 9 – Temps de traitement image des filtres SIR et PARTITIONNE *vs.* nombre de particules.

6.4 Temps de traitement

La stratégie PARTITIONNE nécessite des temps de traitement sensiblement équivalents à ceux de la stratégie CONDENSATION (figure 9) pour un nombre donné de particules. Les tests se sont déroulés sur un *Pentium IV Centrino* 1.8 GHz, la fréquence de traitement étant de 4 Hz pour $N = 100$ particules et de 3.3 Hz pour $N = 400$ particules. Le prétraitement des images (débayérisation, redimensionnement, image de probabilité peau, image de distance) consomme la majeure partie du temps des ressources, soit 220 ms par image. Nous pensons pouvoir augmenter la cadence du filtre en réduisant la résolution des images traitées. Cette évaluation montre que les temps sont équivalents pour un même nombre de particules. Mais nous avons montré ci-dessus que l'on a besoin de moins de particules pour une stratégie PARTITIONNE.

7 Conclusion

Dans cet article, nous avons présenté une approche dédiée au suivi par vision multi-oculaire de structures 3D articulées représentant les membres corporels. Elle repose sur une stratégie de filtrage particulaire à échantillonnage partitionné pour estimer les différents degrés de liberté modélisant la cinématique de la structure. Notre modèle de mesure en fusionnant à la fois des informations d'apparence et géométriques (3D) en est une illustration. Il reste suffisamment discriminant malgré les environnements variés rencontrés par notre

robot mobile. Les premières évaluations sont prometteuses. Elles montrent notamment que la stratégie de filtrage PARTITIONNE est plus performante dans un contexte applicatif où le nombre de particules utilisable doit rester compatible avec les ressources CPU embarquées sur le robot.

Pour compléter ces évaluations, il conviendrait d'évaluer d'autres stratégies, telles l'"Annealed Particle Filter" ou l'ICONDENSATION. Cette dernière, en prenant en compte des mesures dans la fonction d'importance, permet de placer les particules dans les zones pertinentes de l'espace d'état. De plus, cette stratégie, répandue en suivi 2D, offre la capacité de se ré-initialiser automatiquement après décrochage du filtre, ce qui constituerait un intérêt majeur dans notre contexte applicatif.

Références

[1] F. Aherne, N. Thacker, and P. Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 32(4) :1–7, 1997.

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online non-linear/non-gaussian bayesian tracking. *Trans. on Signal Processing*, 2(50) :174–188, 2002.

[3] P. Azad, A. Ude, R. Dillman, and G. Cheng. A full body human motion capture system using particle filtering and on-the-fly edge detection. In *Proceedings of IEEE-RAS/RSJ International Conf. on Humanoid Robots (Humanoids 2004)*, Los Angeles, 11 2004.

[4] Q. Delamarre and O. Faugeras. 3D articulated models and multi-view tracking with physical forces. *Computer Vision and Image Understanding (CVIU'01)*, 81 :328–357, 2001.

[5] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'00)*, pages 126–133, 2000.

[6] A. Doucet, N. De Freitas, and N. J. Gordon. *Sequential Monte Carlo Methods in Practice*. Series Statistics For Engineering and Information Science. Springer-Verlag, New York, 2001.

[7] A. Doucet, S. J. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3) :197–208, 2000.

[8] M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. *Int. Journal on Computer Vision (IJCV'98)*, 29(1) :5–28, 1998.

[9] G. Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1) :1–25, 1996.

[10] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *European Conf. on Computer Vision (ECCV'00)*, pages 3–19, 2000.

[11] P. Menezes, F. Lerasle, J. Dias, and R. Chatila. A single camera motion capture system dedicated to gestures imitation. In *Int. Conf. on Humanoid Robots (HUMANOID'05)*, pages 430–435, Tsukuba, 2005.

[12] V. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction : A review. *Trans. On Pattern Analysis and Machine Intelligence (PAMI'97)*, 19(7) :677–695, 1997.

[13] P. Pérez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. *Proc. IEEE*, 92(3) :495–513, 2004.

[14] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. PhD thesis, Institut National Polytechnique de Grenoble, 1996.

[15] H. Sidenbladh, M.J. Black, and D.J. Fleet. Stochastic tracking of 3D human figures using 2D image motion. In *European Conf. on Computer Vision (ECCV'00)*, pages 702–718, 2000.

[16] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling. *Int. Journal on Robotic Research (IJRR'03)*, 6(22) :371–393, 2003.

[17] B. Stenger, P. R. S. Mendonça, and R. Cipolla. Model-based hand tracking using an unscented kalman filter. In *British Machine Vision Conf. (BMVC'01)*, volume 1, pages 63–72, September 2001.

[18] P. Torma and C. Szepesvári. Sequential importance sampling for visual tracking reconsidered. In *AI and Statistics*, pages 198–205, 2003.

[19] R. Urtasum and P. Fua. 3D human body tracking using deterministic temporal motion models. In *European Conf. on Computer Vision (ECCV'04)*, 2004.

[20] Y. Wu, J.Y. lin, and T.S. Huang. Capturing natural hand articulation. In *Int. Conf. on Computer Vision (ICCV'01)*, pages 426–432, 2001.