

# Stratégies de filtrage particulaire pour le suivi visuel de personnes : description et évaluation

L. Brèthes<sup>†</sup>, P. Danès<sup>†‡</sup>, F. Lerasle<sup>†‡</sup>

<sup>†</sup> LAAS-CNRS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4

<sup>‡</sup> Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cédex

Mél : {lbrethes, danes, lerasle}@laas.fr

## Résumé

*Cet article traite du suivi visuel de personnes à partir d'une caméra embarquée sur un robot mobile en environnement humain, a priori encombré. Nous décrivons diverses mesures visuelles reposant sur la couleur, la forme ou le mouvement, ainsi que diverses stratégies de filtrage particulaire prenant en compte tout ou partie de ces mesures. Ces stratégies permettent éventuellement de combiner les mesures visuelles afin de définir une fonction d'importance selon laquelle les particules sont échantillonnées, ou de fusionner plusieurs primitives à l'intérieur d'un modèle de mesure servant de base à la définition des poids.*

*Les performances de ces filtres sont comparées aux travers de quelques évaluations quantitatives ou qualitatives. L'association de certains algorithmes de filtrage et de certaines primitives visuelles montre des gains non négligeables en robustesse et précision du suivi. De plus, les diverses mesures combinées et les stratégies de filtrage proposées répondent aux scénarii envisagés pour notre robot (interaction proximale, proche, surveillance,...). Les extensions de ces travaux sont discutées en final.*

## Mots-Clef

vision couleur, filtrage particulaire, fusion de données visuelles, détection et suivi.

## Abstract

*This paper deals with visual tracking of people from a camera mounted on a mobile robot in a human — a priori cluttered— environment. Several visual cues are described, relying on color, shape or motion, together with several particle filtering strategies taking into account all or part of these measurements. These strategies enable the combination of visual cues into an importance function from which the particles are sampled, or the fusion of several primitives into a measurement model serving as the basis for the definition of weights.*

*The filter performances are compared through some quantitative and qualitative evaluations. Some associations of filtering strategies and visual cues show a significant increase in the tracking robustness and precision. Moreover, the combined measurement cues together with the proposed filtering strategies meet the requirements of the envisaged scenarii for our robot (mean-range interaction with an user, monitoring,...) The extensions of this work are discussed at the end.*

## Keywords

color vision, particle filtering, visual data fusion, detection and tracking.

## 1 Introduction

De nombreux travaux dans la communauté Vision portent sur le suivi visuel de personnes. Ces travaux visent notamment des applications telles que les interfaces homme-machine [9], la compression vidéo [16], la surveillance [6], ou la robotique [3]. Un défi majeur de la robotique est sans doute celui du robot personnel, avec la perspective de permettre à un robot mobile de naviguer en présence de public, voire d'interagir avec lui. Le robot évolue dans des environnements *a priori* encombrés et sujets à des changements d'illumination. Il est alors opportun de gérer à chaque instant plusieurs hypothèses sur les paramètres à estimer et d'exploiter plusieurs sources de mesures.

Le filtrage particulaire, abondamment référencé dans la littérature, semble très adapté à ce contexte. Il permet de s'affranchir de toute hypothèse restrictive



FIG. 1 – Le robot Rackham

quant aux distributions de probabilité entrant en jeu dans la caractérisation du problème. De plus, ce formalisme permet de combiner/fusionner aisément différents types de mesures. Malgré ce constat, la fusion de données par filtrage particulaire nous semble assez peu exploitée dans notre contexte et souvent confinée à un nombre restreint de primitives visuelles. Pérez *et al.* ont largement abordé ce problème dans [14]. Ils proposent un algorithme de filtrage hiérarchisé, où les mesures sont successivement consommées de façon à positionner efficacement les particules. Il nous semble cependant intéressant de combiner plus largement les mesures afin d'augmenter leur pouvoir discriminant dans la pondération des particules. Nous souhaitons également évaluer quelques-unes des autres nombreuses stratégies de filtrage proposées dans la littérature de façon à déterminer les combinaisons des primitives visuelles et des algorithmes qui répondent au mieux aux scénarii envisagés pour notre robot Rackham censé guider et interagir avec les visiteurs d'une exposition.

L'article est structuré comme suit. La section 2 rappelle sommairement le formalisme bien connu du filtrage particulaire et décrit quelques variantes permettant la fusion de données dans un contexte de suivi. La section 3 spécifie quelques mesures visuelles reposant sur la couleur, la forme ou le mouvement image de la cible observée. La section 4 décrit la mise en œuvre du suivi. Différentes stratégies de filtrage sont ensuite proposées et évaluées dans notre contexte applicatif. Enfin, la section 5 résume notre contribution et propose quelques extensions envisagées.

## 2 Algorithmes de filtrage particulaire pour la fusion de données

### 2.1 Un algorithme générique

Les techniques de filtrage particulaire sont des méthodes de simulation séquentielles de type Monte Carlo permettant l'estimation du vecteur d'état d'un système Markovien non nécessairement linéaire soumis à des excitations aléatoires possiblement non Gaussiennes [1, 4, 5]. En tant qu'estimateurs Bayésiens, leur but est d'estimer récursivement la densité de probabilité *a posteriori*  $p(x_k|z_{1:k})$  du vecteur d'état  $x_k$  à l'instant  $k$  conditionné sur l'ensemble des mesures  $z_{1:k} = z_1, \dots, z_k$ , une connaissance *a priori* de la distribution du vecteur d'état initial  $x_0$  pouvant être également prise en compte. À chaque instant  $k$ , la densité  $p(x_k|z_{1:k})$  est approximée au moyen de la distribution ponctuelle

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

exprimant la sélection d'une valeur – « particule » –  $x_k^{(i)}$  avec la probabilité – ou « poids » –  $w_k^{(i)}$ ,  $i = 1, \dots, N$ . Les moments conditionnels de  $x_k$ , tels que l'estimateur du minimum d'erreur quadratique moyenne  $E[x_k|z_{1:k}]$ , peuvent alors être approchés par ceux de la variable aléatoire ponctuelle de densité de probabilité (1).

Les particules  $x_k^{(i)}$  évoluent stochastiquement dans le temps. Elles sont échantillonnées selon une *fonction d'importance* visant à explorer adaptativement les zones « pertinentes » de l'espace d'état.

L'algorithme générique de filtrage particulaire est présenté Table 1-gauche. Son initialisation consiste en la définition d'un ensemble de particules pondérées décrivant la distribution *a priori*  $p(x_0)$ , e.g. en affectant des poids identiques  $w_0^{(i)} = \frac{1}{N}$  à des échantillons  $x_0^{(1)}, \dots, x_0^{(N)}$  indépendants identiquement distribués (i.i.d.) selon  $p(x_0)$ .

À chaque instant  $k$ , disposant de la mesure  $z_k$  et de la description particulaire  $\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}$  de  $p(x_{k-1}|z_{1:k-1})$ , la détermination de l'ensemble de particules pondérées  $\{x_k^{(i)}, w_k^{(i)}\}$  associé à la densité *a posteriori*  $p(x_k|z_{1:k})$  procède en deux étapes. Dans un premier temps, les  $x_k^{(i)}$  sont échantillonnés selon la fonction d'importance  $q(x_k|x_{k-1}, z_k)$  évaluée en  $x_{k-1} = x_{k-1}^{(i)}$ , cf. l'équation (2). Les poids  $w_k^{(i)}$  sont ensuite mis à jour de façon à assurer la cohérence de l'approximation (1). Ce calcul obéit à (3), où  $p(x_k|x_{k-1})$  rend compte de la dynamique du processus d'état sous-jacent, et la vraisemblance  $p(z_k|x_k)$  d'un état possible  $x_k$  vis à vis de la mesure  $z_k$  est évaluée à partir de la densité de probabilité relative au lien état-observation.

Toute méthode de simulation séquentielle de type Monte Carlo souffre du phénomène de dégénérescence, au sens où après quelques itérations, les poids non négligeables tendent à se concentrer sur une seule particule. Afin de limiter ce phénomène, une étape de rééchantillonnage peut être insérée en fin de chaque cycle, cf. l'item 9 de l'algorithme SIR (*Sampling Importance Resampling*) Table 1. Ainsi,  $N$  nouvelles particules  $\tilde{x}_k^{(i)}$  sont obtenues par rééchantillonnage avec remise dans l'ensemble  $\{x_k^{(j)}\}$  selon la loi  $P(\tilde{x}_k^{(i)} = x_k^{(j)}) = w_k^{(j)}$ . Les particules associées à des poids  $w_k^{(j)}$  élevés sont dupliquées, au détriment de celles, faiblement pondérées, qui disparaissent, de sorte que la séquence  $\tilde{x}_k^{(1)}, \dots, \tilde{x}_k^{(N)}$  est i.i.d. selon  $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ .

Cette étape de redistribution peut soit être appliquée systématiquement, soit être déclenchée seulement lorsqu'un critère d'efficacité du filtre passe en deçà d'un certain seuil [5, 1]. Le calcul des moments de (1) doit de préférence faire intervenir l'ensemble de particules pondérées avant rééchantillonnage.

$\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^N = \text{SIR}(\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1}^N, z_k)$	$\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^N = \text{CONDENSATION}(\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1}^N, z_k)$
1: <b>SI</b> $k = 0$ ( <b>INITIALISATION</b> ) <b>ALORS</b>	1: <b>SI</b> $k = 0$ ( <b>INITIALISATION</b> ) <b>ALORS</b>
2: Échantillonner $x_0^{(1)}, \dots, x_0^{(N)}$ i.i.d. selon $p(x_0)$ , et poser $w_0^{(i)} = \frac{1}{N}$ , $i = 1, \dots, N$	2: Échantillonner $x_0^{(1)}, \dots, x_0^{(N)}$ i.i.d. selon $p(x_0)$ , et poser $w_0^{(i)} = \frac{1}{N}$ , $i = 1, \dots, N$
3: <b>FIN SI</b>	3: <b>FIN SI</b>
4: <b>SI</b> $k \geq 1$ <b>ALORS</b>	4: <b>SI</b> $k \geq 1$ <b>ALORS</b>
5: <b>POUR</b> $i = 1, \dots, N$ , <b>FAIRE</b>	5: <b>POUR</b> $i = 1, \dots, N$ , <b>FAIRE</b>
6: « Propager » la particule $x_{k-1}^{(i)}$ en simulant de manière indépendante	6: « Propager » la particule $x_{k-1}^{(i)}$ en simulant
$x_k^{(i)} \sim q(x_k   x_{k-1}^{(i)}, z_k) \quad (2)$	$x_k^{(i)} \sim p(x_k   x_{k-1}^{(i)}) \quad (4)$
7: Mettre à jour le poids $w_k^{(i)}$ selon l'équation	7: Mettre à jour le poids $w_k^{(i)}$ selon l'équation
$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k   x_k^{(i)}) p(x_k^{(i)}   x_{k-1}^{(i)})}{q(x_k^{(i)}   x_{k-1}^{(i)}, z_k)} \quad (3)$	$w_k^{(i)} \propto w_{k-1}^{(i)} p(z_k   x_k^{(i)}) \quad (5)$
préalablement à une étape de normalisation assurant que $\sum_i w_k^{(i)} = 1$	préalablement à une étape de normalisation assurant que $\sum_i w_k^{(i)} = 1$
8: <b>FIN POUR</b>	8: <b>FIN POUR</b>
9: Rééchantillonner $\{x_k^{(i)}, w_k^{(i)}\}$ selon $P(\tilde{x}_k^{(i)} = x_k^{(j)}) = w_k^{(j)}$ , ce qui conduit à un ensemble de particules pondérées $\{\tilde{x}_k^{(i)}, \frac{1}{N}\}$ tel que $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ et $\frac{1}{N} \sum_{i=1}^N \delta(x_k - \tilde{x}_k^{(i)})$ approximent $p(x_k   z_{1:k})$ ; affecter $x_k^{(i)}$ et $w_k^{(i)}$ avec $\tilde{x}_k^{(i)}$ et $\frac{1}{N}$	9: Rééchantillonner $\{x_k^{(i)}, w_k^{(i)}\}$ selon $P(\tilde{x}_k^{(i)} = x_k^{(j)}) = w_k^{(j)}$ , ce qui conduit à un ensemble de particules pondérées $\{\tilde{x}_k^{(i)}, \frac{1}{N}\}$ tel que $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ et $\frac{1}{N} \sum_{i=1}^N \delta(x_k - \tilde{x}_k^{(i)})$ approximent $p(x_k   z_{1:k})$ ; affecter $x_k^{(i)}$ et $w_k^{(i)}$ avec $\tilde{x}_k^{(i)}$ et $\frac{1}{N}$
10: <b>FIN SI</b>	10: <b>FIN SI</b>

TAB. 1 – Algorithme générique de filtrage particulaire (SIR) et CONDENSATION

## 2.2 Échantillonnage relativement à la dynamique : la CONDENSATION

L'algorithme de CONDENSATION – pour “Conditional Density Propagation” [7] – peut être vu comme le cas particulier de l'algorithme SIR où la fonction d'importance est relative à la dynamique du processus d'état, cf. (4). Ceci confère à la CONDENSATION une structure prédiction/mise à jour comparable à celle du filtre de Kalman. En effet, la densité ponctuelle  $\sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$  approxime la prédiction  $p(x_k | z_{1:k-1})$ . En outre, la mise à jour des poids selon (5) rappelle la formule de Bayes sous-jacente à l'étape de mise à jour de l'estimé de Kalman.

Dans un contexte de suivi visuel, l'algorithme de CONDENSATION original définit les vraisemblances des particules à partir de primitives visuelles de type contour, mais d'autres primitives ont été également envisagées [13, 11].

## 2.3 Échantillonnage guidé par la mesure

Le rééchantillonnage utilisé seul ne suffit pas à limiter efficacement le phénomène de dégénérescence évoqué en fin du § 2.1. En outre, il peut conduire à une perte de diversité dans l'exploration de l'espace d'état, du fait que la description particulaire de la densité *a posteriori* risque de contenir de nombreuses particules identiques. La définition de la fonction d'importance  $q(x_k | x_{k-1}, z_k)$  – selon laquelle les particules sont distribuées – doit donc également faire l'objet d'une attention particulière [1].

En suivi visuel, les modes des fonctions de vraisem-

blance  $p(z_k | x_k)$  relativement à  $x_k$  sont généralement très marqués. Il s'en suit que les performances de la CONDENSATION sont souvent assez médiocres. Du fait que les particules sont positionnées selon la dynamique du processus d'état et « en aveugle » par rapport à la mesure  $z_k$ , un sous-ensemble important d'entre elles peut être affecté d'une vraisemblance très faible par l'équation (5), dégradant ainsi significativement les performances de l'estimateur.

Une alternative peut donc consister à échantillonner les particules à l'instant  $k$  – ou bien seulement certaines de leurs composantes – selon une fonction d'importance  $q(x_k | z_k)$  définie à partir de l'image courante. Ainsi, l'exploration de l'espace d'état peut être guidée par des fonctionnalités de détection visuelle telles que blobs couleur [8], visage (§3), ou bien toute autre primitive intermittente, qui, malgré un caractère sporadique, est très discriminante lorsqu'elle est présente [14] : mouvement, son, etc. Une telle stratégie sera référencée par MSIR pour *Measurement SIR*.

Cependant, rien n'empêche qu'une particule  $x_k^{(i)}$ , dont tout ou partie des composantes sont positionnées à partir de l'image courante, soit incompatible avec sa particule prédécesseur  $x_{k-1}^{(i)}$  du point de vue de la dynamique du processus d'état. Du fait que  $p(x_k^{(i)} | x_{k-1}^{(i)})$  prend de faibles valeurs, une telle particule est alors faiblement pondérée dans (3).

Dans le contexte du suivi visuel, la première stratégie prenant en compte la mesure dans la fonction d'im-

portance est l'ICONDENSATION [8]. Une proportion des particules demeure toutefois échantillonnée suivant l'algorithme de CONDENSATION ou selon la densité initiale  $p(x_0)$ .

Une alternative très intéressante, proposée dans [15], considère des modèles dynamiques dont le vecteur d'état  $x_k = (u'_k, v'_k, h'_k)'$  obéit à une dynamique de la forme  $p(x_k | x_{k-1}) = p(u_k, v_k | x_{k-1}) \delta(h_k - f(x_{k-1}))$ , le symbole  $'$  désignant l'opérateur de transposition. En d'autres termes, le sous-vecteur  $(u'_k, v'_k)'$  – appelé « partie innovation » de  $x_k$  – satisfait une équation stochastique en  $x_{k-1}$  alors que le sous-vecteur complémentaire  $h_k$  – qualifié de « partie historique » – est une fonction déterministe  $f(x_{k-1})$ . Les auteurs considèrent une fonction d'importance de la forme  $q(x_k | x_{k-1}, z_k) = \pi(u_k | z_k) p(v_k | u_k, x_{k-1}) \delta(h_k - f(x_{k-1}))$ , i.e. le sous-vecteur  $u_k^{(i)}$  de la « partie innovation » de la particule  $x_k^{(i)}$  est échantillonné à partir de la seule connaissance de l'observation  $z_k$  selon  $\pi(u_k | z_k)$ ,  $v_k^{(i)}$  suit la loi  $p(v_k | u_k^{(i)}, x_{k-1}^{(i)})$ , et la « partie historique »  $h_k^{(i)}$  est trivialement définie à partir de la particule prédécesseur  $x_{k-1}^{(i)}$  par  $h_k^{(i)} = f(x_{k-1}^{(i)})$ . La densité de probabilité de la mesure conditionnellement au vecteur d'état est supposée satisfaire  $p(z_k | x_k) = p(z_k | u_k, v_k)$ .

Ce contexte est particulièrement adapté à la problématique du suivi visuel. En effet, l'écriture des modèles de dynamique linéaire auto-régressifs, dont l'utilisation est courante, sous forme d'équation d'état fait intervenir un vecteur d'état se décomposant naturellement en une « partie innovation » et une « partie historique ». Le lien état-mesure ne fait alors pas intervenir la « partie historique », conformément à l'hypothèse ci-dessus. L'hypothèse relative à la fonction d'importance permet de différencier les composantes de la « partie innovation » pouvant facilement être échantillonnées à partir de la seule connaissance de l'observation – e.g. positions horizontale et verticale d'un « template » à partir du résultat d'un processus de détection de blobs couleur – de celles qu'il est préférable d'échantillonner selon la dynamique *a priori*.

Les auteurs définissent des mécanismes permettant d'éviter toute contradiction entre  $(u_k^{(i)'}, v_k^{(i)'})'$  et le passé  $x_{k-1}^{(i)}$ . Il peut s'agir d'un rééchantillonnage des « parties innovations » guidé par leurs probabilités d'occurrence, suivi, pour chacune des sous-particules ainsi obtenues, de l'échantillonnage d'un passé – et donc d'une partie historique – plausible.

Une version « Rao-Blackwellisée » est également développée, où seul un passé plausible correspondant à chaque sous-particule  $u_k^{(i)}$  est rééchantillonné, la suppression du rééchantillonnage de la partie innovation permettant de réduire la variance de l'esti-

mateur. Dans cette veine, la Table 2 présente l'algorithme RBSSHSSIR (*Rao-Blackwellised Subspace History Sampling SIR*) de [15, Table 4]. On remarque que l'étape 7 consiste, pour chaque sous-particule  $u_k^{(i)}$  positionnée à partir de  $z_k$ , en le rééchantillonnage d'une particule prédécesseur – et par conséquent en la « partie historique »  $h_k^{(i)}$  de  $x_k^{(i)}$  – qui soit à la fois vraisemblable par rapport à  $u_k^{(i)}$  du point de vue de la dynamique et dotée d'un poids significatif. L'algorithme RBSSHSSIR diffère de l'ICONDENSATION du fait même de la présence de cette étape, pourtant nécessaire sous peine que l'ensemble de particules pondérées  $\{x_k^{(i)}, w_k^{(i)}\}$  ne soit pas une description cohérente de la densité *a posteriori*  $p(x_k | z_{1:k})$  à l'instant  $k$ . Enfin, bien que la démonstration sorte du cadre de cet article, on montre que l'algorithme s'applique également lorsque  $u_k$  et  $v_k$  obéissent à des dynamiques du premier ordre indépendantes, auquel cas il suffit simplement de supprimer la composante  $f(x_{k-1})$  dans  $x_k$ , cf. [2].

## 2.4 Vers le cas « adapté » : l'algorithme de filtrage à particules auxiliaires

Il peut être montré [5] que la fonction d'importance et l'équation correspondante de définition des poids

$$x_k^{(i)} \sim q^*(x_k | x_{k-1}^{(i)}, z_k) \triangleq p(x_k | x_{k-1}^{(i)}, z_k), \quad (6)$$

$$w_k^{*(i)} \propto w_{k-1}^{*(i)} p(z_k | x_{k-1}^{(i)}) \quad (7)$$

conduisent à la stratégie optimale, ou « adaptée », au sens où elle permet de limiter au mieux la dégénérescence de l'estimateur. Les formules (6)–(7) sont cependant d'un intérêt pratique limité, au sens où il est souvent impossible d'échantillonner selon  $p(x_k | x_{k-1}^{(i)}, z_k)$  ni de calculer la vraisemblance  $p(z_k | x_{k-1}^{(i)}) = \int p(z_k | x_k) p(x_k | x_{k-1}^{(i)}) dx_k$ .

Une propriété remarquable du cas « adapté » (6)–(7) tient à ce que chaque particule  $x_k^{(i)}$  est affectée d'un poids  $w_k^{*(i)}$  ne dépendant que de sa particule prédécesseur  $x_{k-1}^{(i)}$ . Dès lors, l'étape (7) peut être réalisée avant l'étape (6). L'efficacité globale de l'algorithme peut encore être augmentée par l'introduction d'un rééchantillonnage auxiliaire préalablement à la « propagation » au travers de (6) de l'ensemble de particules pondérées  $\{x_{k-1}^{(i)}, w_k^{*(i)}\}$ , qui représente  $p(x_{k-1} | z_{1:k})$ .

Même si seule une approximation  $\hat{p}(z_k | x_{k-1}^{(i)})$  de  $p(z_k | x_{k-1}^{(i)})$  est disponible, il demeure possible de se rapprocher du cas optimal présenté ci-dessus. Il suffit de remplacer  $w_k^{*(i)}$  par une variable  $\lambda_k^{(i)}$ , définie à partir de  $\hat{p}(z_k | x_{k-1}^{(i)})$  et servant de base à un rééchantillonnage auxiliaire. Un filtre particulaire « à particules auxiliaires » (APF = Auxiliary Particle Filter), développé dans cette veine contemporanément

$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N = \text{RBSSHSSIR}([\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^N, z_k)$	$[\{x_k^{(i)}, w_k^{(i)}\}]_{i=1}^N = \text{APF}([\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^N, z_k)$
1: <b>SI</b> $k = 0$ ( <b>INITIALISATION</b> ) <b>ALORS</b> 2: Échantillonner $x_0^{(1)}, \dots, x_0^{(N)}$ i.i.d. selon $p(x_0)$ , et poser $w_0^{(i)} = \frac{1}{N}$ , $i = 1, \dots, N$ 3: <b>FIN SI</b> 4: <b>SI</b> $k \geq 1$ <b>ALORS</b> 5: <b>POUR</b> $i = 1, \dots, N$ , <b>FAIRE</b> 6: Échantillonner $u_k^{(i)} \sim \pi(u_k   z_k)$ 7: Échantillonner dans $(1, \dots, N)$ l'indice $I_k^{(i)}$ de la particule prédécesseur de $u_k^{(i)}$ selon les poids $(w_{k-1}^{(1)} p(u_k^{(1)}   x_{k-1}^{(1)}), \dots, w_{k-1}^{(N)} p(u_k^{(N)}   x_{k-1}^{(N)}))$ 8: Échantillonner $v_k^{(i)} \sim p(v_k   u_k^{(i)}, x_{k-1}^{(I_k^{(i)})})$ 9: Poser $x_k^{(i)} = \begin{pmatrix} u_k^{(i)} \\ v_k^{(i)} \\ f(x_{k-1}^{(I_k^{(i)})}) \end{pmatrix}$ 10: Mettre à jour les poids, préalablement à leur normalisation, en posant $w_k^{(i)} \propto \frac{p(z_k   u_k^{(i)}) \sum_{l=1}^N w_{k-1}^{(l)} p(u_k^{(i)}   x_{k-1}^{(l)})}{\pi(u_k^{(i)}   z_k)}$ , de sorte que $p(x_k   z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ 11: <b>FIN POUR</b> 12: <b>FIN SI</b>	1: <b>SI</b> $k = 0$ ( <b>INITIALISATION</b> ) <b>ALORS</b> 2: Échantillonner $x_0^{(1)}, \dots, x_0^{(N)}$ i.i.d. selon $p(x_0)$ , et poser $w_0^{(i)} = \frac{1}{N}$ , $i = 1, \dots, N$ 3: <b>FIN SI</b> 4: <b>SI</b> $k \geq 1$ <b>ALORS</b> 5: <b>POUR</b> $i = 1, \dots, N$ , <b>FAIRE</b> 6: À partir de l'approximation $\hat{p}(z_k   x_{k-1}^{(i)})$ , calculer les pondérations auxiliaires $\lambda_k^{(i)} \propto w_{k-1}^{(i)} \hat{p}(z_k   x_{k-1}^{(i)})$ 7: <b>FIN POUR</b> 8: Rééchantillonner $\{x_{k-1}^{(i)}, \lambda_k^{(i)}\}$ de façon à obtenir l'ensemble de particules pondérées équivalent $\{\tilde{x}_{k-1}^{(i)}, \frac{1}{N}\}$ ; $\sum_{i=1}^N \lambda_k^{(i)} \delta(x_{k-1} - x_{k-1}^{(i)})$ et $\frac{1}{N} \sum_{i=1}^N \delta(x_{k-1} - \tilde{x}_{k-1}^{(i)})$ décrivent $p(x_{k-1}   z_{1:k})$ 9: <b>POUR</b> $i = 1, \dots, N$ , <b>FAIRE</b> 10: « Propager » les particules en échantillonnant $x_k^{(i)} \sim q(x_k   \tilde{x}_{k-1}^{(i)}, z_k)$ 11: Mettre à jour les poids, préalablement à leur normalisation, en posant $w_k^{(i)} \propto \frac{p(z_k   x_k^{(i)}) p(x_k^{(i)}   \tilde{x}_{k-1}^{(i)})}{\hat{p}(z_k   \tilde{x}_{k-1}^{(i)}) q(x_k^{(i)}   \tilde{x}_{k-1}^{(i)}, z_k)}$ , de sorte que $p(x_k   z_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)})$ 12: <b>FIN POUR</b> 13: <b>FIN SI</b>

TAB. 2 – Filtre particulaire avec rééchantillonnage d'un sous-ensemble de la partie innovation (RBSSHSSIR= Rao-Blackwellised Subspace History Sampling SIR) – Filtre particulaire « à particules auxiliaires » (APF= Auxiliary Particle Filter)

à l'apparition de l'ICONDENSATION par [12], est résumé Table 2-droite. Il admet pour fonction de vraisemblance auxiliaire  $\hat{p}(z_k | x_{k-1}^{(i)}) = p(z_k | x_{k-1}^{(i)})$ , où  $x_{k-1}^{(i)}$  caractérise la distribution de  $x_k$  conditionné sur  $x_{k-1}^{(i)}$  – e.g.  $x_{k-1}^{(i)} = \mathbb{E}[x_k | x_{k-1}^{(i)}]$  ou  $x_{k-1}^{(i)} \sim p(x_k | x_{k-1}^{(i)})$ , et pour fonction d'importance  $q(x_k | \tilde{x}_{k-1}^{(i)}, z_k) = p(x_k | \tilde{x}_{k-1}^{(i)})$ .

Le filtre particulaire « à particules auxiliaires » permet de mieux orienter les particules vers les zones pertinentes de l'espace d'état. Dans un contexte de suivi visuel, il est tout à fait possible de définir les vraisemblances auxiliaires à partir de primitives visuelles différentes de celles entrant en jeu dans le calcul des poids définitifs. La limitation principale de cet algorithme est un mauvais fonctionnement lorsque la dynamique est très bruitée ou lorsque les densités d'observation admet des modes très étroits.

Il convient enfin de signaler que la validité de l'algorithme RBSSHSSIR peut être démontrée selon des arguments analogues à ceux sous-tendant l'APF [2].

### 3 Fonctions d'importance et modèles de mesure

Comme indiqué précédemment, les informations visuelles peuvent être prises en compte dans le modèle de mesure et/ou dans la fonction d'importance. La fonction d'importance permet de placer les particules conformément à des mesures image pertinentes mais possiblement intermittentes dans le flot vidéo, e.g. le mouvement. La mise à jour des poids s'effectue par le

modèle de mesure; ces mesures sont ici persistentes dans le flot mais hélas pas toujours suffisamment discriminantes en présence d'environnements encombrés. Il est alors judicieux de combiner/fusionner plusieurs mesures de natures différentes, typiquement la couleur, la forme ou le mouvement. Nous spécifions ci-après quelques unes des mesures utilisées dans nos filtres.

#### 3.1 Couleur

**Fonction d'importance.** La couleur est souvent caractéristique des cibles à suivre et peut être exploitée comme mesure dans le filtre. Nous avons ainsi proposé dans [3] une méthode de segmentation non supervisée de régions peau correspondant aux membres corporels. Plus globalement, il existe dans la littérature de nombreuses approches traitant du problème. Leur efficacité est souvent liée à l'encombrement des scènes observées et aux variations d'illuminations associées. Ainsi, une sur-exposition (robot dans un hall d'entrée) ou une sous-exposition (robot dans un couloir) rendent souvent la segmentation incertaine tandis que l'encombrement de la scène peut induire de fausses détections. Cette information doit être utilisée avec précaution et combinée avec d'autres mesures visuelles. Soit  $B$  le nombre de régions détectées définies par leurs centroïdes  $b_k$ . La fonction d'importance pour les paramètres de position image  $x_{pos}$  s'écrit comme une mixture de Gaussiennes [8] i.e. :

$$g^C(x_{pos}) = \sum_{k=1}^B \delta_k \mathcal{N}(b_k, \quad B) \quad (8)$$

où  $b_k = b'_k + \bar{X}b$ . Les paramètres  $\bar{X}b$  et  $B$ , appris hors ligne, représentent respectivement la moyenne et la covariance du décalage entre la position réelle de la cible et le centroïde du *blob* couleur extrait sur plusieurs séquences.

**Modèle de mesure.** Une stratégie pour exploiter la couleur dans la fonction de mesure est de comparer les distributions de couleur de régions d'intérêt à un modèle de référence. Pour représenter celles-ci, nous utilisons ici des histogrammes normalisés indépendants dans l'espace RGB. Soit  $h_{ref}^c = (h_{1,ref}^c, \dots, h_{N_{bi},ref}^c)$  l'histogramme de référence sur  $N_{bi}$  cellules où  $c \in \{R, G, B\}$ . La distribution de couleur  $h_x^c = (h_{1,x}^c, \dots, h_{N_{bi},x}^c)$  d'une région d'intérêt  $B_x$  correspondant à l'état  $x$  est donnée par

$$h_{j,x}^c = c_H \sum_{u \in B_x} \delta_j(b_u^c), j = 1, \dots, N_{bi},$$

où  $b_u^c \in \{1, \dots, N_{bi}\}$  indexe la cellule de l'histogramme correspondant au niveau pour le canal  $c$  du pixel  $u$  et  $c_H$  est un terme de normalisation tel que  $\sum_{j=1}^{N_{bi}} h_{j,x}^c = 1$ .

Pour comparer les deux distributions  $h_{j,x}^c$  et  $h_{ref}^c$ , nous utilisons la distance de Bhattacharyya [14] notée ici  $D$ . Le modèle de mesure est alors défini par la relation (9), cf. Figure 2 :

$$p(z^C|x) \propto \exp\left(-\sum_{c \in \{R,G,B\}} D^2(h_x^c, h_{ref}^c)/2 \cdot \frac{2}{C}\right). \quad (9)$$

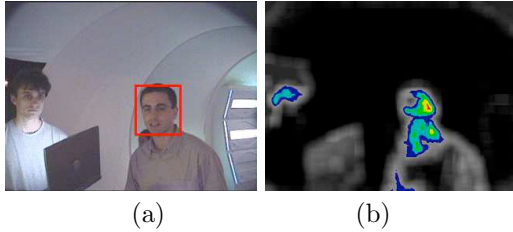


FIG. 2 – (a) Région d'intérêt, (b) Modèle de mesure couleur pour chaque position image (échelle fixe)

Il peut être intéressant de considérer plusieurs régions d'intérêt ayant des couleurs distinctes, typiquement le visage et les vêtements d'une personne. En posant

$$B_x = \bigcup_{p=1}^{N_R} B_{p,x}, \text{ le modèle de mesure (9) devient}$$

$$p(z^C|x) \propto \exp\left(-\sum_{c \in \{R,G,B\}} \sum_{p=1}^{N_R} D^2(h_{p,x}^c, h_{p,ref}^c)/2 \cdot \frac{2}{C}\right). \quad (10)$$

Enfin, dans notre contexte, les changements d'apparence du sujet observé, de par ses mouvements réels ou les variations d'illumination, impliquent une réactualisation du modèle à chaque instant image  $k$ . Cette mise à jour est donnée par [11]

$$h_{ref,k} = (1 - \alpha) \cdot h_{ref,k-1} + \alpha \cdot h_{E[x_k]}$$

où les indices  $c$  et  $j$  sont omis pour plus de clarté, et  $\alpha$  pondère l'influence de l'histogramme  $h_{E[x_k]}$  correspondant à l'état moyen  $E[x_k]$  dans la réactualisation.

### 3.2 Forme

**Modèle de mesure.** Les membres corporels à suivre ont une forme caractéristique que nous pouvons exploiter dans notre contexte applicatif. Ainsi, la silhouette d'une tête est modélisée par une spline dont l'état est estimé par un filtre particulaire.

Les particules sont classiquement pondérées à partir des observations constituées des contours image suivant les directions orthogonales à la spline aux points de contrôle [9]. A l'instant  $k$ , ce modèle de mesure, conditionné sur la particule  $x_k^{(i)}$  s'écrit, en omettant les indices  $i$  et  $k$  pour simplifier l'écriture,

$$p(z^F|x) \propto \exp\left(-K \sum_{j=0}^{N_p} \phi(j)\right) \quad (11)$$

avec  $d(j) = |x(j) - z(j)|$  tandis que  $\phi(j)$  vaut

$$\phi(j) = \begin{cases} d(j)^2 & \text{si } d(j) < \delta \\ \rho & \text{sinon,} \end{cases} \quad (12)$$

où  $x(j)$  et  $z(j)$ ,  $j = 1, \dots, N_p$ , désignent respectivement le  $j^{\text{ème}}$  point de contrôle sur la spline et le point de contour le plus proche associé sur la normale à la spline (figure 3). Enfin  $K$ ,  $\rho$  et  $\delta$  sont des constantes prédéfinies. Cette mesure,

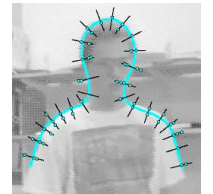


FIG. 3 – Modèle de mesure

bien que persistente, reste très sensible aux conditions d'éclairage *a priori* quelconques et surtout peu discriminante en présence de scènes encombrées.

**Fonction d'importance.** Nous exploitons ici le détecteur multi-échelle de visages introduit par Viola *et al.* [17]. Ce détecteur bien connu repose sur des masques de Haar (figure 4-(a)) mesurant des contrastes relatifs entre les yeux, le nez, les joues. Ces contrastes sont liés à la forme caractéristique du visage humain. Une phase d'apprentissage sélectionne les masques les plus discriminants au travers d'une cascade de classifieurs isolant, en final, les sous-fenêtres pertinentes dans l'image (figure 4-(b)). Ces régions d'intérêt sont exploitables dans un module de détection dont la fonction d'importance  $g^V(x_{pos})$  est définie par la relation (8) à partir des  $B$  régions correspondant aux visages détectés.

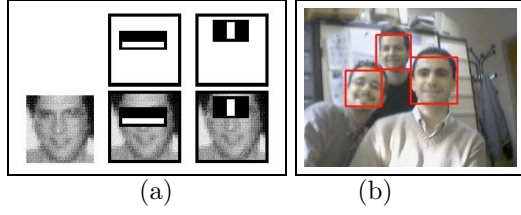


FIG. 4 – (a) Masques de Haar et images d'apprentissage associées, (b) exemples de détection

### 3.3 Mouvement

**Fonction d'importance.** La détection d'objets par le mouvement est largement abordée dans la littérature [10]. Pour une caméra statique (robot à l'arrêt), une simple différence entre deux images successives aux instants  $k$  et  $k - 1$  permet de déduire aisément une fonction d'importance. Comme précédemment, nous définissons pour la seule composante intensité un histogramme de référence  $h_{ref}^m$  caractérisant une région en mouvement. Un bon compromis est de poser [14]

$$h_{i,ref}^m = \frac{1}{N_{bi}}, \quad i = 1, \dots, N_{bi}$$

Soit  $B_x$  une région d'intérêt dont la taille varie en fonction de l'échelle courante (à l'instant  $k$ ) et positionnée à intervalles réguliers dans l'image. Nous retenons les  $B$  positions (régions) qui vérifient  $D^2(h_{x_{pos}}^m, h_{ref}^m) > seuil$  où  $h_{x_{pos}}^m$  est l'histogramme associé à chaque région définie par sa position  $x_{pos}$ . La fonction d'importance  $g^m(x_{trans})$  associée est caractérisée par un mélange de  $B$  Gaussiennes à partir de ces positions  $x_{pos}$  (relation 8). La figure 5 illustre cette fonction d'importance pour un exemple d'image de différence.

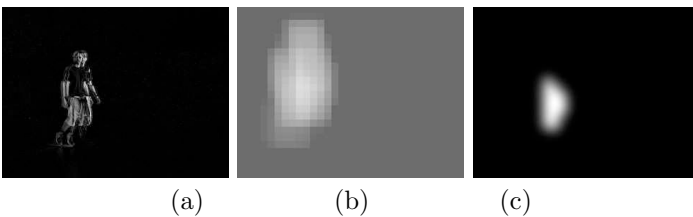


FIG. 5 – (a) Image de différence absolue ( $z^M$ ), (b) Image de similarité avec un histogramme de référence uniforme ( $1 - D^2(h_{x_{pos}}^m, h_{ref}^m)$ ), (c) Fonction d'importance associée.

**Modèle de mesure.** Il est possible de considérer le mouvement dans le modèle de mesure, notamment de privilégier les pixels en mouvement dans la vraisemblance calculée. Soit  $f(z(j))$  le vecteur flot optique au pixel  $z(j)$ , l'expression (12) devient

$$\phi(j) = \begin{cases} d(j)^2 + \rho\gamma(z(j)) & \text{si } d(j) < \delta \\ 2\rho & \text{sinon} \end{cases}$$

avec

$$\gamma(z(j)) = \begin{cases} 0 & \text{si } |\vec{f}(z(j))| \neq 0 \\ 1 & \text{sinon.} \end{cases}$$

On remarquera qu'il s'agit ici d'une pondération donc le modèle de mesure  $p(z^M|x)$  reste cohérent avec un mouvement intermittent de la cible dans le temps.

## 4 Suivi de personnes

Le but est de recalculer le *template* associé à la personne à suivre sur chaque image dans le flot vidéo et donc d'estimer ses coordonnées image  $(x, y)$ , son facteur d'échelle  $s$ , voire son orientation lorsque le *template* est relatif à la forme.

Concernant le modèle de dynamique  $p(\mathbf{x}_k|\mathbf{x}_{k-1})$  du vecteur d'état  $\mathbf{x}_k$  à l'instant  $k$ , les mouvements image des personnes observées sont difficiles à caractériser. Nous supposons ici que les composantes de l'état évoluent suivant des modèles de marche aléatoire indépendants, soit

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k|\mathbf{x}_{k-1}, \Sigma)$$

où  $\mathcal{N}(\cdot | \mu, \Sigma)$  est une distribution Gaussienne de moyenne  $\mu$  et covariance  $\Sigma = \text{diag}(\frac{\sigma_x^2}{2}, \frac{\sigma_y^2}{2}, \dots)$ . Le vecteur d'état estimé dans le filtre est donc :  $\mathbf{x}_k = [x_k, y_k, s_k, \theta_k]'$ .

Notre mécanisme de fusion, pour le modèle d'observation, des mesures visuelles décrites précédemment est présenté ci-après puis intégré dans diverses stratégies de filtrage avec ou sans fonctions d'importance. Des évaluations quantitatives (précision, robustesse, temps de calcul) et qualitatives sont également proposées sur des scénarii propres à notre contexte robotique.

### 4.1 Fusion de données

Dans notre contexte robotique où l'environnement est variable et *a priori* encombré, il est nécessaire d'adopter une stratégie de fusion des données visuelles dans l'observation afin de robustifier le suivi. La figure 6 montre le résultat du calcul de la vraisemblance en tous points d'une même image pour plusieurs attributs visuels. Cette image est extraite d'une séquence de suivi où la personne de droite (cible) est mobile et la personne de gauche reste immobile.

L'utilisation du seul attribut forme n'est manifestement pas assez discriminant comme l'illustre la figure 6-(a) où de nombreuses régions liées à l'arrière-plan sont vraisemblables. Si on ne considère que la distribution de couleur (figure 6-(b)) la vraisemblance, élevée pour les deux individus, ne permet pas de différencier la cible. La fusion de (a) et (b) dans la figure 6-(c) permet d'éliminer la majeure partie des régions du fond et réduit considérablement la largeur des régions vraisemblables sans toutefois différencier les deux personnes. Dans le cas d'une cible mobile, la fusion de l'attribut forme avec l'attribut mouvement



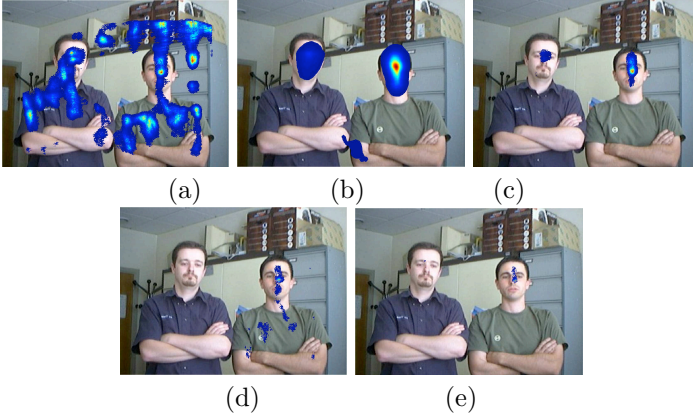


FIG. 6 – Attributs : (a) forme ; (b) couleur ; (c) fusion de (a)+(b) ; (d) fusion de (a)+mouvement, (e) fusion de (b)+(d)

(figure 6-(d)) permet d’isoler la cible mais aussi certaines régions non désirées. Le problème majeur de cette combinaison est lié à l’intermittence du mouvement, un arrêt de la cible nous ramenant alors au cas (a). Finalement, la fusion des trois attributs (figure 6-(e)) permet de distinguer la cible.

Afin d’observer l’effet produit par le choix de ces différentes stratégies de mesure sur le filtrage particulière, nous avons réalisé une expérimentation de suivi d’une personne évoluant dans un environnement encombré. L’estimation à chaque instant image  $k$  des paramètres du vecteur d’état  $\mathbf{x}_k$  est réalisée par un filtre particulière de type CONDENSATION pour un nombre  $N$  de particules variant de 10 à 900. Pour chacun des attributs précédents, les erreurs moyennes d’estimation – pour l’estimateur de la moyenne *a posteriori*  $E(x_k|z_{1:k})$  – et les taux moyens d’échec sont calculés et reportés respectivement sur les figures 7 et 8 à partir de réalisations sur un ensemble de séquences représentatives de notre contexte robotique.

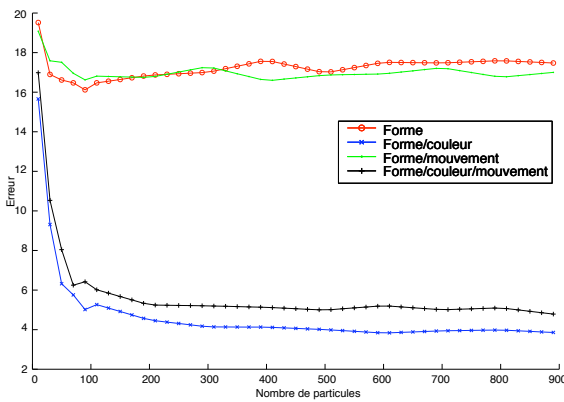


FIG. 7 – Erreur d’estimation pour chaque stratégie de mesure *vs* nombre de particules

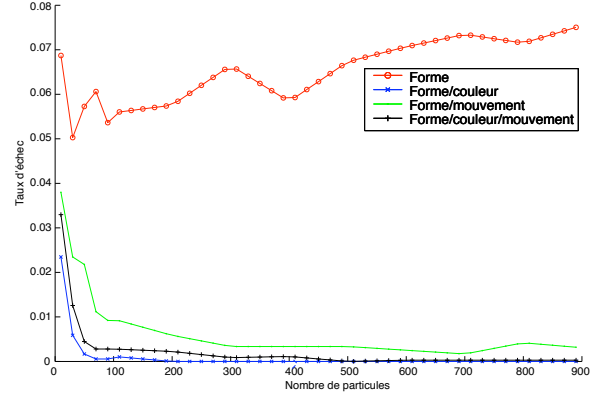


FIG. 8 – Taux d’échec pour chaque stratégie de mesure *vs* nombre de particules

La figure 7 montre que la combinaison des attributs forme et couleur est la stratégie la plus efficace pour les séquences traitées. L’erreur est nettement inférieure à celle générée par l’utilisation de la forme seule. La fusion des attributs forme et mouvement n’apporte pas d’amélioration significative sur la précision tandis que le taux d’échec est sensiblement diminué (figure 8). En effet, l’attribut mouvement pénalise les contours immobiles de l’image, ce qui permet de diminuer le nombre de fausses mesures. L’influence des contours statiques n’étant cependant pas complètement éliminée, la précision est de l’ordre de celle obtenue avec l’attribut forme seul. La stratégie de fusion des trois attributs visuels permet un gain en robustesse. On note cependant que la stratégie de fusion des deux seuls attributs forme et distribution de couleur donnent des performances très similaires. Nous privilégions donc cette dernière stratégie dans la mise en œuvre de nos filtres.

## 4.2 Stratégies de filtrage

Diverses stratégies de filtrage particulière sont ici considérées pour notre contexte applicatif. Nous considérons des stratégies basées sur des fonctions d’importance relatives respectivement à la dynamique (notées FID) ou à la mesure (notées FIM). Pour ces fonctions d’importance, les stratégies de filtrage présentées au §2 sont mises en œuvre puis évaluées. Notons que la stratégie APF est classée dans les FID malgré le fait que les mesures interviennent dans le rééchantillonnage intermédiaire. Ces évaluations s’appuient sur trois modalités d’interaction pour notre « robot-guide » Rackham :

1. Interaction active proximale ( $d_{H/R} < 1m$ ) : le robot suit visuellement son interlocuteur situé à environ  $1m$ , qui interagit avec lui *via* ses interfaces pour notamment sélectionner un stand à visiter.
2. Mission de guidage ( $d_{H/R} \in [1m, 3m]$ ) : le robot



suit visuellement les visiteurs durant sa mission de guidage sur un lieu prédéfini en #1.

3. Surveillance longue distance ( $d_{H/R} > 3m$ ) : le robot immobile et en recherche d'interaction, surveillance *via* la caméra située en haut du mât, les visiteurs entrant sur le site afin de les interpellier.

Les évaluations des trois fonctions de suivi associées ne peuvent être détaillées ici pour des raisons de place. Nous nous focalisons sur la mission de guidage qui est la modalité principale. Dans ce contexte, la stratégie de filtrage mise en œuvre doit rester robuste (*i*) aux conditions de prises de vue (scènes encombrées, variations d'illumination), (*ii*) aux occultations ou changements d'apparence de la cible, (*iii*) aux sauts dans ses mouvements apparents induits par les déplacements relatifs H/R durant la mission.

L'évaluation porte ici sur une base de 20 séquences acquises depuis le robot et représentatives des artefacts mentionnés ci-dessus. 20 réalisations de suivi sont effectuées par séquence. Pour chaque filtre et un nombre variable de particules, des statistiques sur la précision et la robustesse sont alors calculées, *via* respectivement les erreurs moyennes et taux d'échec.

La fonction de mesure de nos filtres fusionne classiquement forme (relation 11) et distributions de couleur sur plusieurs régions (relation 10) comme illustré sur la figure 9. La stratégie auxiliaire considère ces distributions de couleur comme vraisemblance auxiliaire et la forme comme mesure terminale. Enfin, pour les stratégies FIM, nous utilisons le détecteur de visage vu au §3.2.

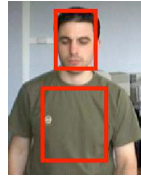


FIG. 9 – Régions d'intérêt du *template*

Les évaluations, illustrées par les figures 10, montrent que les stratégies FIM sont plus précises et plus robustes que les stratégies FID car elles peuvent « raccrocher » la cible grâce à leur fonction d'importance basée sur la mesure. La figure 11 illustre ce constat sur une séquence incluant deux personnes avec occultations mutuelles. Les stratégies FID ne permettent pas logiquement de « raccrocher » la cible après occultation, contrairement aux stratégies FIM.

La comparaison des résultats pour des stratégies de même type montre dans le cas des stratégies FID que le filtre basé sur une mesure auxiliaire donne une meilleure précision et un taux d'échec plus faible que la CONDENSATION. Concernant les stratégies FIM on note que le filtre RBSSHSSIR donne une précision légèrement meilleure que le filtre MSIR mais un taux d'échec similaire.

La modalité #3 implique l'utilisation d'attributs visuels différents. Ainsi, la fonction d'importance repose

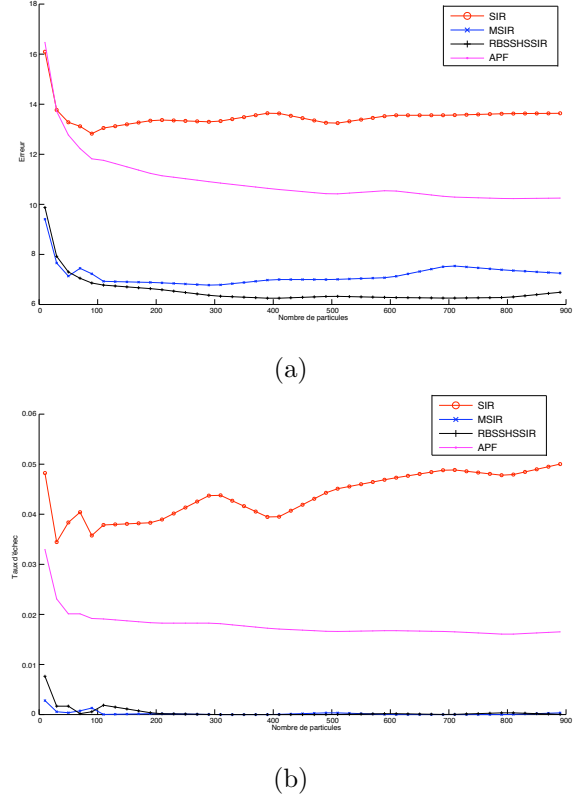


FIG. 10 – (a) Erreur d'estimation et (b) Taux d'échec, *vs* nombre de particules

sur le détecteur de mouvement vu au §3.3 tandis que la fonction de mesure est basée sur la distribution de couleur associée au rectangle englobant la cible. Les évaluations effectuées pour cette modalité montrent, ici encore, la pertinence des stratégies FIM. La Figure 12 montre un exemple de suivi dans cette modalité incluant une occultation par une autre personne.

Les évaluations relatives à la modalité #1 montrent que les stratégies FID sont au moins aussi performantes que les stratégies FIM lorsqu'on considère la fonction de mesure combinant forme et mouvement du §3.3. Pour cette modalité, des stratégies FID seront donc implémentées sur le robot.

## 5 Conclusion et perspectives

Cet article traite du suivi visuel de personnes à partir d'une caméra embarquée sur un robot mobile censé guider les visiteurs d'une exposition et interagir avec eux. Les environnements rencontrés par le robot sont *a priori* encombrés et évolutifs. Le repositionnement des particules par la fonction d'importance puis l'association d'informations hétérogènes dans le modèle de mesure augmentent la robustesse et la précision du suivi. Les diverses mesures combinées et les stratégies de filtrage proposées répondent aux modalités d'interaction envisagées pour notre « robot-guide ».

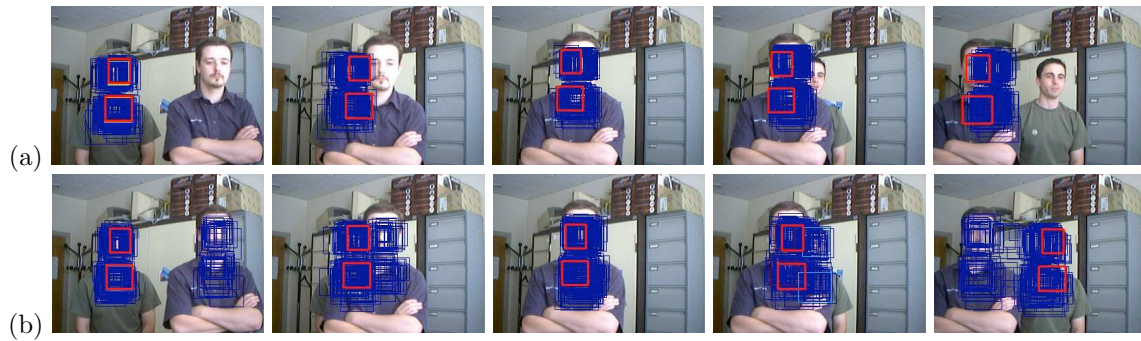


FIG. 11 – Fonction d'importance relative (a) à la dynamique (CONDENSATION), (b) à la mesure (MSIR)



FIG. 12 – Exemple de suivi dans la modalité #3 incluant une occultation

Les fonctions de suivi propres à ces modalités sont en cours d'intégration sur la plateforme Rackham. Leur but est ici de permettre au robot d'interagir et de guider les visiteurs d'un musée. Plus globalement, elles doivent faciliter/sécuriser les déplacements d'un robot mobile en présence de public.

D'autres stratégies de filtrage seront implémentées et évaluées. Des critères quantitatifs relatifs à l'efficacité de l'estimateur seront calculés et discutés. Une réflexion portera également sur la définition d'un mécanisme de détection d'un « décrochage » du filtre. Nous prévoyons d'intégrer des primitives issues d'un système de localisation acoustique. Cette information plutôt intermittente sera prise en compte dans la fonction d'importance. Enfin, à terme, notre évaluation sera étendue à des approches non particulières.

Toutes les évaluations et réalisations de suivi associées aux trois modalités d'interaction sont disponibles à l'URL [www.laas.fr/~lbrethes/rfia06/](http://www.laas.fr/~lbrethes/rfia06/).

## Références

- [1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Trans. on Signal Processing*, 50(2) :174–188, February 2002.
- [2] L. Brèthes. *Suivi visuel par filtrage particulière. Application à l'interaction Homme-Robot*. PhD thesis, Université Paul Sabatier, LAAS-CNRS, en cours de publication, 2005.
- [3] L. Brèthes, P. Menezes, F. Lerasle, and M. Briot. Segmentation couleur et condensation pour le suivi et la reconnaissance de gestes humains. In *Reconnaissance des Formes et Intelligence Artificielle*, volume 2, pages 967–975, 2004.
- [4] A. Doucet, N. de Freitas, and N.J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Series Statistics for Engineering and Information Science. Springer-Verlag, New York, May 2001.
- [5] 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.
- [6] I. Haritaogly, D. Harwood, and L. Davis. W4 : Real-time surveillance of people and their activities. *IEEE Trans. on Pattern Analysis Machine Intelligence*, 8(22) :809–830, 2000.
- [7] M.A. Isard and A. Blake. CONDENSATION-Conditional Density Propagation for Visual Tracking. *Int. Journal of Computer Vision (IJCV'98)*, 29(1) :5–28, 1998.
- [8] M.A. Isard and A. Blake. Icondensation : Unifying low-level and high-level Tracking in a Stochastic Framework. In *European Conf. on Computer Vision*, pages 893–908, 1998.
- [9] M.A. Isard and A. Blake. Visual Tracking by Stochastic Propagation of Conditional Density. In *European Conf. on Computer Vision*, pages 343–356, Cambridge, April 1996.
- [10] J. Konrad. *Motion Detection and Estimation*. Handbook of Image and Video Processing, 2000.
- [11] K. Nummiaro, E. Koller-Meier, and L. Van Gool. Object Tracking with an Adaptive Color-Based Particle Filter. In *Symp. for Pattern Recognition of the DAGM*, pages 353–360, 2002.
- [12] Michael K. Pitt and Neil Shephard. Filtering via simulation : Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446) :590–599, 1999.
- [13] P. Pérez, J. Vermaak C. Hue, and M. Gangnet. Color-based probabilistic tracking. In *European Conf. on Computer Vision*, pages 661–675, 2002.
- [14] P. Pérez, J. Vermaak, and A. Blake. Data Fusion for Visual Tracking with Particles. *IEEE*, 92(3) :495–513, 2004.
- [15] P. Torma and C. Szepesvári. Sequential importance sampling for visual tracking reconsidered. In *AI and Statistics*, pages 198–205, 2003.
- [16] W. Vieux, K. Schwerdt, and J. Crowley. Face Tracking and Coding for Video Compression. In *Int. Conf. on Computer Vision Systems*, pages 151–160, January 1999.
- [17] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2001.