Utilisation du mouvement visuel en suivi par filtrage particulaire

On the use of visual motion in particle filter tracking

Jean-Marc Odobez[1], Daniel Gatica-Perez[1] et Sileye Ba[1]

[1]Idiap Research Institute, Martigny, Switzerland

Manuscrit reçu en janvier 2004(*)

Résumé et mots clés

Le filtrage par méthode de Monte-Carlo séquentielle (MCS) est l'une des méthodes les plus populaires pour effectuer du suivi visuel. Dans ce contexte, une hypothèse importante faite généralement stipule que, étant donnée la position d'un objet dans des images successives, les observations extraites de ces dernières sont indépendantes. Dans cet article, nous soutenons que, au contraire, ces observation sont fortement corrélées et que la prise en compte de cette corrélation permet d'améliorer le suivi. Par ailleurs, un choix relativement fréquent consiste à utiliser le modèle dynamique *a priori* comme fonction de proposition. Par conséquent la génération des échantillons à l'instant courant se fait en aveugle, c'est-à-dire sans exploiter d'information liée à l'image courante. Il en résulte que la variance du bruit dans le modèle dynamique doit être fixée à une valeur importante afin de pouvoir appréhender de rapides changements de trajectoire. De ce fait de nombreuses particules sont générées inutilement dans des régions de faible vraisemblance, ce qui réduit l'efficacité de l'échantillonnage, ou sont propagées sur des ambiguïtés voisines de la vraie trajectoire, ce qui, ultérieurement, peut conduire à des erreurs de suivi. Dans cet article, nous proposons d'utiliser le mouvement visuel afin de remédier aux deux problèmes soulevés. Des mesures de mouvement explicites sont utilisées pour diriger l'échantillonnage vers les nouvelles régions intéressantes de l'image, tandis que des mesures implicites et explicites sont introduites dans la distribution de vraisemblance afin de modéliser la corrélation entre données temporelles. Le nouveau modèle permet d'appréhender des mouvements brusques et de lever des ambiguïtés visuelles tout en gardant des modèles d'objet simples basés sur des contours ou des distribution de couleurs, comme le montrent les résultats obtenus sur plusieurs séquences et comparés à la méthode classique de CONDENSATION.

Suivi d'objets, séquence, corrélation, mouvement, estimation, filtrage particulaire, CONDENSATION, méthode de Monte-Carlo séquentielle.

Abstract and key words

Particle filtering is now established as one of the most popular methods for visual tracking. Within this framework, a basic assumption is that the data are temporally independent given the sequence of object states. In this paper, we argue that in general the data are correlated, and that modeling such dependency should improve tracking robustness. Besides, the choice of using the transition prior as proposal distribution is also often made. Thus, the current observation data is not taken into account in the generation of the new samples, requesting the noise process of the prior to be large enough to handle abrupt trajectory changes between the previous image data and the new one. Therefore, many particles are either wasted in low likelihood area, resulting in a low efficiency of the sampling, or, more importantly, propagated on near distractor regions of the image, resulting in tracking failures. In this paper, we propose to handle both issues using motion. Explicit motion measurements are used to drive the sampling process towards the new interesting regions of the image, while implicit motion measurements are introduced in the likelihood evaluation to model the data correlation term. The proposed model allows to handle abrupt motion changes and to filter out visual distractors when tracking objects with generic models based on shape or color distribution representations. Experimental results compared against the CONDENSATION algorithm have demonstrated superior tracking performance.

Object tracking, sequence, visual motion, estimation, particle filter, CONDENSATION, sequential Monte-Carlo.

1. Introduction

Le suivi d'objets est un problème important en vision par ordinateur, avec des applications en téléconférence, surveillance visuelle, reconnaissance de gestes, et définition d'interfaces visuelles [5]. Néanmoins, bien qu'étant étudié de façon intensive, cela reste un problème difficile en présence d'ambiguïtés (*e.g.* lors du suivi d'un objet en présence d'objets de cette même classe), de bruit dans les mesures (*e.g.* les problèmes d'illumination), de variabilité de la classe d'objets considérée.

Dans ce domaine, les méthodes de suivi par filtrage particulaire – aussi appelé filtrage par méthode de Monte-Carlo Séquentielle (MCS) – [10, 11, 8, 3, 7] ont prouvé leur validité. Le filtrage particulaire est une approche bayesienne dans laquelle la probabilité de la configuration d'un objet (*e.g.* sa position, son échelle), en tenant compte des observations, est représentée par un ensemble d'échantillons pondérés appelés particules. Cette représentation permet de maintenir plusieurs hypothèses de suivi simultanément [11, 21], contrairement aux algorithmes basés sur une représentation unique [6] cherchant à maximiser un critère en utilisant des méthodes d'optimisation itératives, et qui sont de ce fait plus sensibles à des erreurs ponctuelles dues à la présence d'ambiguïtés ou de mouvements rapides ou erratiques. Notons ici que, récemment, des travaux [27, 28] ont montré l'intérêt de combiner les avantages respectifs (hypothèses multiples/optimisation locale) de ces deux approches.

Dans cet article, nous abordons deux aspects importants du suivi par filtrage particulaire. Le premier concerne l'hypothèse d'indépendance des observations conditionellement à la séquence d'états. Le second aspect traité est celui du choix d'une distribution de proposition qui, pour être efficace, doit prendre en compte les mesures extraites de l'image courante. Pour aborder ces deux aspects, nous proposons une méthode de suivi particulaire exploitant le mouvement visuel apparent. Notre approche repose sur un nouveau modèle graphique permettant d'introduire naturellement des mesures de mouvement implicites ou explicites dans la fonction de vraisemblance des observations, ainsi que sur l'emploi de mesures de mouvement explicites dans la distribution de proposition. Les paragraphes suivants décrivent plus en détail les aspects évoqués ainsi que notre approche et ses bénéfices.

La définition de la distribution de vraisemblance des observations, qui modélise la probabilité de l'observation courante étant donnée la configuration courante de l'objet, est certainement l'un des points les plus important en suivi particulaire. Elle repose sur la représentation de l'objet choisie. Celle-ci correspond à tout ce qui, implicitement ou explicitement, caractérise l'objet : sa position, sa couleur, son apparence, etc. Par exemple, des modèles de contour paramétrisés comme des splines [5] ou des ellipses [32], ainsi que des distributions de couleur [6, 20, 23, 32] sont souvent utilisés. Un inconvénient de ces représentations est qu'elles sont peu spécifiques, ce qui les rend sensibles aux ambiguïtés locales. La fusion de représentations de bas niveau comme les contours et la couleur [32] permet de réduire le nombre d'ambiguïtés.

La forme standard du modèle de vraisemblance repose sur une hypothèse classique en filtrage MCS visuel, à savoir l'indépendance des données étant donné la séquence d'états [2, 4, 5, 12, 24, 30, 32]. Dans cet article, nous soutenons que cette hypothèse n'est pas toujours appropriée. Afin de la généraliser, nous proposons un modèle dans lequel l'observation courante dépend non seulement de l'état présent mais aussi de l'observation et de l'état précédents. Nous montrons que cette nouvelle hypothèse, plus générale, permet d'obtenir un algorithme de filtrage particulaire qui possède des équations similaires à celles reposant sur l'hypothèse traditionnelle. À notre connaissance, ceci n'avait jamais été montré auparavant et constitue la première contribution de l'article. La nouvelle hypothèse permet alors d'introduire naturellement des informations implicites ou explicites de mouvement dans le terme de vraisemblance. L'introduction d'une telle corrélation entre images successives transforme ainsi des suiveurs[1] génériques basés sur des contours et histogrammes, en suiveurs plus spécifiques.

La fonction de proposition, c'est-à-dire la fonction dont on tire les nouvelles configurations échantillons où sera évaluée la vraisemblance *a posteriori* des données, est une autre distribution importante d'un filtre particulaire. De manière générale, un choix optimal [8, 3] consiste à tirer les échantillons dans les régions les plus probables de l'espace d'état, en tenant compte à la fois du modèle dynamique, qui caractérise l'information *a priori* sur le séquence d'états, et de la vraisemblance des nouvelles observations. Malheureusement, les modèles de vraisemblance les plus courants ne permettent pas de simuler des échantillons suivant cette loi. Un choix fréquent consiste alors à utiliser le modèle dynamique p_{prior} comme fonction de proposition. Dans ce cas, la variance du bruit dans le processus dynamique définit implicitement l'espace de recherche des nouveaux échantillons. La difficulté de modélisation de ce terme provient alors de deux aspects contradictoires : d'un côté, on souhaite que l'espace de recherche soit suffisament grand pour pouvoir appréhender des changements brusques de mouvement ; de l'autre, on souhaite le restreindre pour éviter au suivi d'être perturbé par des ambiguïtés locales proches de la véritable configuration de l'objet, ce qui est susceptible de se produire lors de l'utilisation de modèles d'objets peu spécifiques. De plus, une telle fonction de proposition ne prend pas en compte l'image courante. Les particules générées ainsi auront probablement une faible vraisemblance, ce qui conduit à une faible efficacité du processus d'échantillonnage. De manière générale, un tel filtre risque d'être fortement perturbé par des ambiguïtés dans le fond de l'image. Pour remédier à ces problèmes, nous proposons d'exploiter des mesures de mouvement explicites dans la fonction de proposition, ce qui permettra de mieux appréhender les mouvements abruptes de l'objet et d'augmenter ainsi l'efficacité de l'échantillonnage. En conjonction avec le nouveau modè-

1. Nous utiliserons ce terme en remplacement de « tracker » en anglais.

le de vraisemblance, une telle approche réduira la sensibilité de l'algorithme de suivi à la définition numérique des paramètres de variance de bruit dans les distributions de proposition et *a priori* puisque, lors de l'utilisation de variances plus élevées, les ambiguïtés potentielles seront éliminées par le terme de corrélation introduit. Par ailleurs, cette approche permet de mettre en œuvre l'idée intuitive selon laquelle les configurations les plus vraisemblables vis-à-vis d'un modèle d'objet se « déplacent » en accord avec le mouvement apparent.

Le reste de l'article est organisé de la façon suivante : dans la section à venir, nous établissons un état de l'art des méthodes se rapportant à notre travail ; dans la section 3, nous introduisons le filtre particulaire générique ; nous motivons notre approche dans la section 4 et décrivons plus en détail notre modèle dans la section 5. Les expériences et les résultats sont présentés dans la section 6. Enfin, la section 7 conclut l'article.

2. État de l'art

Dans cet article, la première contribution concerne l'introduction d'un nouveau modèle graphique pour le suivi particulaire. Ce modèle permet de modéliser les dépendances temporelles entre observations. En pratique, celui-ci nous a conduit à introduire naturellement des observations de mouvement dans la vraisemblance des données.

L'utilisation du mouvement pour le suivi n'est pas une idée nouvelle. Les suiveurs basés sur le mouvement, principalement déterministes, intègrent le déplacement estimé entre deux images au cours du temps. Néanmoins, en l'absence de modèle d'objet, il s'avère difficile d'éviter une dérive après quelques secondes de suivi. L'utilisation de modèles d'apparences, notamment des prototypes [2, 28, 29], conduit à des algorithmes plus robustes. L'inconvénient de ce type d'approches est de n'autoriser que de faibles variations d'apparence dans la séquence.

Pour appréhender des changements d'apparences, une adaptation souvent difficile du modèle s'avère nécessaire [13,15]. Une alternative consiste à employer des modèles d'apparence plus complexes (*e.g.* par espace propre [4] ou par ensemble d'exemples [22, 30]), ce qui pose le problème de leur apprentissage, hors-ligne [4, 30] ou en ligne [22]. Par exemple, dans [13] les auteurs proposent d'utiliser un modèle génératif où la vraisemblance des données courantes est représentée par un mélange de lois faisant intervenir un prototype à long terme, l'image précédente, et une composante de bruit non-gaussien. L'adaptation se fait au travers de l'estimation des paramètres optimaux (comprenant la configuration de l'objet et son protoype), estimation réalisée à l'aide d'un algorithme « expectation-maximisation » (EM) identifiant au passage les parties stables du prototype long terme. Une approche similaire est exploitée dans [15], où le niveau de gris de chaque pixel du prototype est mis-à-jour à l'aide d'un filtre de Kalman. L'adaptation est bloquée lorsque l'innovation est trop grande. Dans ces deux cas, si des occultations partielles – voire totales – peuvent être appréhendées, rien n'empêche véritablement, à long terme, la dérive évoquée plus haut. Cette dérive se produit lorsque le mouvement apparent 2D ne correspond pas à l'évolution réelle des paramètres d'états. Le cas problématique d'une tête tournant sur elle-même mais restant à la même position dans l'image est notamment rapporté dans [13] ; dans [15], les objets à suivre présentent peu de variations de pose. Une approche intéressante est proposée dans [31], où, dans un cadre statistique, un modèle de couleur est adapté en ligne par l'intermédiaire d'un module de détection de mouvement supposant une caméra statique selectionnant les instants d'adaptation les plus appropriés, ce qui conduit à de bons résultats.

Dans cet article, le modèle proposé n'exploite pas un prototype d'apparence *de référence* (*cf.* discussion en 5.3). La mise en œuvre de notre modèle vise à évaluer l'accord entre le mouvement mesuré, implicitement ou explicitement, entre deux images, et le mouvement apparent induit par le changement d'état entre les deux instants. Notre approche est donc différente des approches évoquées ci-dessus, et s'apparente plutôt aux méthodes proposées dans [25, 26]. C'est en particulier dans [25], qui aborde le difficile problème du suivi de personnes avec des modèles articulés, que l'utilisation des mesures de mouvement répond mieux au modèle graphique que nous proposons qu'à l'hypothèse courante d'indépendance conditionnelle faite par les auteurs. Il est intéressant de noter que, dans un contexte non lié au suivi d'objets, l'utilisation de chaînes de Markov dites couplées ont été proposées pour prendre en compte la dépendance entre états et observations à des instants consécutifs [18].

Nous avons soulevé dans l'introduction les problèmes liés au choix du modèle dynamique comme fonction de proposition. Différentes approches ont été proposées dans la littérature pour les résoudre : par exemple, si elles sont disponibles, des informations auxiliaires générées à partir de la couleur [12, 31, 21], de la détection de mouvement [21], ou de signaux audios synchronisés dans le cas de suivi de locuteur [9, 21] peuvent être exploitées pour tirer des nouveaux échantillons. La fonction de proposition s'exprime alors comme un mélange entre la loi *a priori* et une partie de la vraisemblance des données. Un avantage important de cette approche est de permettre une (ré)initialisation automatique du suivi. Dans [19], les auteurs proposent un autre type de filtre auxiliaire. L'idée est d'utiliser la vraisemblance d'un premier ensemble d'échantillons prédits à l'instant $t+1$ pour rééchantillonner les échantillons d'origine à l'instant t, et d'appliquer aux nouveaux échantillons la méthode de propogation standard. La prise en compte de la nouvelle image agit par l'intermédiaire de l'augmentation ou de la diminution des descendants d'un échantillon en fonction de sa vraisemblance future. Un tel schéma, cependant, ne fonctionne bien que si la variance de la dynamique est faible comparativement à celle de la vraisemblance, ce qui est rarement le cas en suivi visuel.

Une alternative proposée dans [24] consiste à utiliser le filtre à particules « unscented » pour estimer les distributions de vrai-

semblance. Bien qu'attrayante, cette approche ne résout pas vraiment le problème des mouvements rapides. De plus, elle nécessite de convertir les évaluations de vraisemblances en mesures dans l'espace des configurations, ce qui est difficile dans le cas de vraisemblances basées sur des distributions de couleur et pour certains paramètres d'état (échelle, rotation). Dans [24], la configuration à estimer est réduite à la simple position de l'objet dans l'image. Dans [2,1], l'ensemble des probabilités du filtre est conditionné par rapport aux images. Ceci permet aux auteurs d'utiliser le mouvement estimé entre deux images comme modèle dynamique, de préférence à un modèle auto-regressif, afin d'améliorer la prédiction. De plus, dans leur application (le suivi de points), l'utilisation d'un modèle d'observation linéaire permet d'exploiter la fonction de proposition optimale. Cependant, comme dans [24], des mesures dans l'espace de configuration sont nécessaires. Par ailleurs, bien que leur usage du mouvement soit similaire à ce que nous proposons, celui-ci est introduit différemment (par l'intermédiaire du modèle dynamique plutôt que du terme de vraisemblance) et réduit dans la pratique à des translations.

3. Filtre particulaire

Le filtre à particules est une technique permettant d'implémenter un filtre bayesien récursif à l'aide de simulations de Monte-Carlo. Le lecteur peut consulter les articles [3,8] pour avoir une description plus détaillée de ces méthodes. L'idée consiste à représenter la distribution *a posteriori* $\mathrm{p}(c_{0:k}|z_{1:k})$ de la séquence d'états $c_{0:k} = (c_l, l = 0,\ldots,k)$ jusqu'à l'instant k conditionnellement aux observations $z_{0:k}$ jusqu'à ce même instant par un ensemble d'échantillons pondérés $\{c^i_{0:k}, w^i_k\}_{i=1}^{N_s}$. Chaque échantillon (ou particule) $c^i_{0:k}$ représente une trajectoire possible de la séquence d'états, et w^i_k représente la vraisemblance de cette trajectoire obtenue à l'aide des observations mesurées jusqu'à l'instant k. Les poids sont normalisées ($\sum_i w^i_k = 1$) afin d'obtenir l'approximation discrète suivante de la densité *a posteriori* :

$$\mathrm{p}(c_{0:k}|z_{1:k}) \approx \sum_{i=1}^{N_s} w^i_k \; \delta(c_{0:k} - c^i_{0:k}) \tag{1}$$

Une telle représentation permet alors de calculer l'espérance d'une fonction f par rapport à cette distribution à l'aide d'une somme pondérée :

$$\int f(c_{0:k})\mathrm{p}(c_{0:k}|z_{1:k})\mathrm{d}c_{0:k} \approx \sum_{i=1}^{N_s} w^i_k f(c^i_{0:k}) \tag{2}$$

En particulier, la trajectoire moyenne de l'état caché peut-être estimée en considérant le moment d'ordre 1 (*i.e.* $f(x) = x$). Dans le cas idéal les échantillons sont tirés de la loi *a posteriori* et les poids sont alors tous égaux à $\frac{1}{N_s}$; malheureusement, en pratique, il est souvent très difficile de tirer de tels échantillons. Une alternative consiste à utiliser le principe de l'échantillonnage pondéré (ou « Importance Sampling » (IS) en anglais), en tirant les échantillons selon une fonction de proposition, et en introduisant un facteur de correction (le poids) pour tenir compte de la différence entre la loi d'échantillonnage et la loi que l'on souhaite approcher. Plus précisément, si l'on note $q(c_{0:k}|z_{1:k})$ la fonction de proposition, alors les poids conduisant à une approximation correcte de la loi *a posteriori* (équation (1)) s'expriment par :

$$w^i_k \propto \frac{\mathrm{p}(c^i_{0:k}|z_{1:k})}{q(c^i_{0:k}|z_{1:k})} \qquad (\text{ et } \sum_i w^i_k = 1)\,. \tag{3}$$

Le but de l'algorithme par filtre particulaire consiste à propager récursivement les échantillons et leurs poids associés au fur et à mesure de l'arrivée des nouvelles données. Pour cela, en utilisant la règle de Bayes, on détermine l'équation de récursivité suivante de la loi *a posteriori* :

$$\mathrm{p}(c_{0:k}|z_{1:k}) = \frac{\mathrm{p}(z_k|c_{0:k}, z_{1:k-1})\mathrm{p}(c_k|c_{0:k-1}, z_{1:k-1})}{\mathrm{p}(z_k|z_{1:k-1})} \times \mathrm{p}(c_{0:k-1}|z_{1:k-1}) \tag{4}$$

En supposant que fonction de proposition possède une forme factorisée similaire ($q(c_{0:k}|z_{1:k}) = q(c_k|c_{0:k-1}, z_{1:k})\; q(c_{0:k-1}|z_{1:k-1})$), on obtient l'équation de remise à jour des poids suivante [3,7] :

$$w^i_k \propto w^i_{k-1} \; \frac{\mathrm{p}(z_k|c^i_{0:k}, z_{1:k-1})\mathrm{p}(c^i_k|c^i_{0:k-1}, z_{1:k-1})}{q(c^i_k|c^i_{0:k-1}, z_{1:k})}\,. \tag{5}$$

Afin de simplifier cette expression générale, les dépendances conditionnelles entre variables sont généralement modélisées selon le modèle graphique de la figure 1a, ce qui correspond aux hypothèses suivantes :

H1 : Les observations $\{z_k\}$, conditionnellement à la séquence d'états, sont indépendantes. Ceci conduit à $\mathrm{p}(z_{1:k}|c_{0:k}) = \prod_{i=1}^{k} \mathrm{p}(z_k|c_k)$, nécessitant la définition des vraisemblances individuelles $\mathrm{p}(z_k|c_k)$. De ce fait, $\mathrm{p}(z_k|c^i_{0:k}, z_{1:k-1}) = \mathrm{p}(z_k|c^i_k)$;

H2 : La séquence d'états $c_{0:k}$ suit un modèle markovien d'ordre 1, caractérisé par la définition de la dynamique $\mathrm{p}(c_k|c_{k-1})$. Ainsi, $\mathrm{p}(c^i_k|c^i_{0:k-1}, z_{1:k-1}) = \mathrm{p}(c^i_k|c^i_{k-1})$.

On obtient alors l'équation de remise à jour des poids suivante :

$$w^i_k \propto w^i_{k-1} \; \frac{\mathrm{p}(z_k|c^i_k)\mathrm{p}(c^i_k|c^i_{k-1})}{q(c^i_k|c^i_{0:k-1}, z_{1:k})} \qquad (\text{ et } \sum_i w^i_k = 1)\,. \tag{6}$$

L'ensemble $\{c^i_{0:k}, w^i_k\}_{i=1}^{N_s}$ est alors approximativement distribué selon $\mathrm{p}(c_{0:k}|z_{0:k})$. Malheureusement, l'échantillonnage d'importance est connu pour être inefficace avec des vecteurs de grande dimension [8], ce qui est le cas de l'espace d'état $c^i_{0:k}$ lorsque k augmente. En pratique, cela conduit à une augmentation continuelle de la variance des poids due à la concentration de la masse de probabilité sur quelques (éventuellement une seule) particules uniquement. Pour éviter cette dégénérescence, un rééchantillonnage est nécessaire de temps à autre. Il a pour effet d'éliminer les particules dont le poids est très faible et de multiplier celles qui sont les plus vraisemblables. Il existe diffé-

Figure 1. Modèle graphique pour le suivi visuel (a) modèle standard (b) modèle proposé.

rentes méthodes de rééchantillonnage [8] ; dans notre implémentation, nous utilisons l'algorithme décrit à la figure 3. Par ailleurs, ce rééchantillonnage peut être effectué systématiquement ou uniquement lorsqu'un critère de qualité en indique la nécessité (par exemple, lorsque la variance des poids dépasse un certain seuil). Lorsque cet échantillonnage est effectué à chaque instant, on obtient l'algorithme de filtrage particulaire décrit à la figure 2.

1. Initialisation :
- pour $i = 1,\ldots,N_s$, échantillonner $c_0^i \sim \mathrm{p}(c_0)$ et poser $k = 1$.

2. Diffusion/propagation :
- pour $i = 1,\ldots,N_s$, échantillonner $\tilde{c}_k^i \sim q(c_k|c_{0:k-1}^i, z_{1:k})$.

3. Remise à jour des poids :
- pour $i = 1,\ldots,N_s$, évaluation des poids w_k^i avec l'équation (5)

4. Selection : rééchantillonnage avec remplacement de N_s particules (*cf* fig. 3)
- $\{c_k^j, \frac{1}{N_s}\} \longleftarrow$Rééchantillonnage $(\{\tilde{c}_k^i, w_k^i\})$
- poser $k = k+1$ et retour à l'étape 2.

Figure 2. L'algorithme de filtrage particulaire.

1. Entrée : un ensemble d'échantillons $\{\tilde{c}_k^i, w_k^i\}$ avec $\sum_i w_k^i = 1$
2. Sortie : un ensemble d'échantillons $\{c_k^i, w_k^i = \frac{1}{N_s}\}$
3. Calcul de la densité cumulée : $b_0 = 0$
- pour $i = 1,\ldots,N_s$; $\quad b_i = b_{i-1} + w_k^i$
4. Échantillonnage : $u \sim$ Uniforme $([0, \frac{1}{N_s}])$; $i = 1$
- pour $j = 1,\ldots,N_s$
• tant que $b_i < u$; $\quad i = i+1$;
• $c_k^j = \tilde{c}_k^i$ et $u = u + \frac{1}{N_s}$

Figure 3. Algorithme de rééchantillonnage.

L'efficacité de l'algorithme de filtrage particulaire dépend en grande partie de la fonction de proposition. Une condition importante que doit respecter celle-ci est de produire une fonction de poids (équation (3) ou (6)) bornée [8]. Sous cette condition plusieurs choix sont possibles. Une stratégie locale optimale consiste à choisir la fonction de proposition qui minimise la variance des poids des échantillons à l'instant k, conditionnellement aux trajectoires $c_{1:k-1}^i$ et aux observations $z_{1:k}$. On peut montrer alors que la fonction de proposition optimale est [8] :

$$q(c_k|c_{k-1}^i, z_k) = \mathrm{p}(c_k|c_{k-1}^i, z_k) \propto \mathrm{p}(z_k|c_k)\mathrm{p}(c_k|c_{k-1}^i) \tag{7}$$

ce qui conduit à l'équation de remise à jour :

$$w_k^i \propto w_{k-1}^i \ \mathrm{p}(z_k|c_{k-1}^i) \tag{8}$$

En pratique cependant, l'échantillonnage de $\mathrm{p}(c_k|c_{k-1}^i, z_k)$ et l'évaluation de $\mathrm{p}(z_k|c_{k-1}^i)$ ne sont réalisables que dans des cas bien particuliers, impliquant par exemple des bruits gaussiens et des modèles d'observation linéaires [3, 8, 1]. Un choix simple généralement fait consiste à employer la distribution *a priori* comme fonction d'importance, ce qui mène à l'équation de remise à jour :

$$w_k^i \propto w_{k-1}^i \ \mathrm{p}(z_k|c_k^i) \qquad (\text{ et } \sum_i w_k^i = 1) \tag{9}$$

Bien que ce modèle soit intuitif et simple à implémenter, ce choix, qui ne prend pas en compte les observations courantes, présente un certain nombre d'inconvénients, notamment pour des vecteurs d'état de grande dimension et des modèles de vraisemblance très étroits.

4. Approche et motivations

Dans cette partie, nous proposons une nouvelle approche de suivi exploitant des mesures caractéristiques du flux optique dans le filtre à particules. Le premier principe utilisé par notre approche consiste à incorporer une mesure de corrélation temporelle dans la fonction de vraisemblance. D'un point de vue théorique, ceci peut se justifier en modifiant le modèle graphique standard (figure 1a), en faisant dépendre la vraisemblance de l'observation courante non seulement de l'état courant mais également de l'état et de l'observation à l'instant précédent (*cf* figure 1b). Par ailleurs, nous proposons d'utiliser des mesures de mouvement explicites afin d'obtenir une meilleure fonction de proposition. Dans cette partie, nous justifions notre approche.

4.1 Révision des hypothèses du filtre standard

L'équation de remise à jour (9) repose sur le modèle graphique probabiliste standard présenté à la figure 1a, en accord avec les hypothèses H1 et H2 de la section précédente.
Si la seconde de ces hypothèses est relativement raisonnable, la première est en revanche rarement vérifiée en pratique dans le cas du suivi visuel. Si l'on conserve seulement deux instants par simplicité, l'hypothèse implique que pour tout couple d'états c_{k-1}, c_k et pour tout couple de données z_{k-1}, z_k :

$$\mathrm{p}(z_k, z_{k-1}|c_k, c_{k-1}) = \mathrm{p}(z_k|c_k, c_{k-1})\mathrm{p}(z_{k-1}|c_k, c_{k-1}) \tag{10}$$

Ceci est une hypothèse forte. En pratique, il existe des trajectoires d'intérêt (*e.g.* la « vraie » trajectoire, ou plus généralement les trajectoires proches de la trajectoire moyenne) pour lesquelles les mesures sont corrélées, et donc pour lesquelles l'hypothèse standard n'est pas valide.
La considération précédente peut s'illustrer de la façon suivante. dans la plupart des modèles de suivi, la configuration de l'objet inclut des paramètres d'une transformation géométrique $\mathcal{T}$. Cette dernière permet d'extraire explicitement ou implicitement de l'image courante l'imagette $\tilde{z}_{c_t}$ associée à l'objet considéré selon :

$$\tilde{z}_{c_k}(\mathbf{r}) = z_k(\mathcal{T}_{c_k}\mathbf{r}),\ \forall\, \mathbf{r} \in R, \tag{11}$$

où $\mathbf{r}$ désigne une position, R est une région de référence fixée, et $\mathcal{T}_{c_k}\mathbf{r}$ correspond à l'application de la transformation $\mathcal{T}$ paramétrée par c_k au pixel $\mathbf{r}$. La vraisemblance des données est alors calculée à partir de cette imagette : $\mathrm{p}(z_k|c_k) = \mathrm{p}(\tilde{z}_{c_k})$. Or, si c_{k-1} et c_k sont deux états successifs d'un objet, on peut faire l'hypothèse suivante :

$$\tilde{z}_{c_k}(\mathbf{r}) = \tilde{z}_{c_{k-1}}(\mathbf{r}) + \eta(\mathbf{r}) \qquad \forall\, \mathbf{r} \in R \tag{12}$$

où η est une variable aléatoire à moyenne nulle et de variance faible. Ce point est illustré sur la figure 4. L'équation (12) est à la base de tous les algorithmes de compensation de mouvement et de compression comme MPEG. Ainsi, d'aprés cette équation, l'indépendance des données conditionnellement à la séquence d'états n'est pas valide. Plus précisément :

$$\mathrm{p}(z_k|z_{1:k-1}, c_{1:k}) \neq \mathrm{p}(z_k|c_k)\,, \tag{13}$$

ce qui signifie que l'on ne peut pas réduire le terme de gauche à celui de droite, comme il est fait usuellement. Compte tenu de (12), un modéle plus approprié pour le suivi visuel est représenté sur la figure 1b. Le nouveau modéle peut être incorporé dans le filtre MCS. Tous les calculs conduisant à l'équation (5) sont généraux et ne dépendent pas des hypothéses H1 et H2. Partant de là, remplacer H1 par le nouveau modéle donne :

$$\mathrm{p}(z_k|z_{1:k-1}, c_{1:k}) = \mathrm{p}(z_k|z_{k-1}, c_k, c_{k-1}). \tag{14}$$

En conservant l'hypothéses H2, l'équation de remise à jour des poids devient alors :

$$w_k^i \propto w_{k-1}^i\ \frac{\mathrm{p}(z_k|z_{k-1}, c_k^i, c_{k-1}^i)\mathrm{p}(c_k^i|c_{k-1}^i)}{q(c_k^i|c_{0:k-1}^i, z_{1:k})}$$

$$(\text{et } \sum_i w_k^i = 1)\,. \tag{15}$$

en remplacement de l'équation (6).

4.2 Fonction de proposition

Compte tenu de notre nouveau modéle graphique, il est possible de montrer, en suivant les mêmes arguments exposés dans [3, 8], que conditionnellement aux particules de l'état passé, la fonction de proposition optimale et l'équation de remise à jour des poids sont données par :

$$q(c_k|c_{k-1}^i, z_{1:k}) = \mathrm{p}(c_k|z_k, z_{k-1}, c_{k-1}^i)$$

$$\propto \mathrm{p}(z_k|z_{k-1}, c_k, c_{k-1}^i)\mathrm{p}(c_k|c_{k-1}^i) \text{ et}$$

$$w_k^i \propto w_{k-1}^i \mathrm{p}(z_k|z_{k-1}, c_{k-1}^i)\,.$$

Comme leurs homologues (7) et (8), ces équations sont difficilement utilisables en pratique. Une premiére possibilité consiste alors à utiliser la distribution *a priori* comme fonction de proposition. Ce choix souffre des inconvénients évoqués en intro-

Figure 4. Images aux temps t et t + 3. Les deux imagettes locales correspondant à la tête sont fortement corrélées.

duction, et, dans le cas du suivi visuel, de la non spécificité des variations d'état qui tend à favoriser l'emploi de modéles structurels simples (à vitesse ou accélération constante). De plus, la faible fréquence d'échantillonnage temporel et la présence de mouvements rapides et inattendus, dus à des mouvements de caméra ou d'objets, rendent la détermination des paramétres de bruit difficile.

Une alternative réside dans l'utilisation d'un mélange de lois entre modéle *a priori* et loi dérivée de la vraisemblance. Par exemple, dans [21], des mesures de détection de mouvement sont utilisés dans la distribution de vraisemblance, et des mesures similaires sont exploitées pour construire la fonction de proposition. Dans cette article, nous adoptons une approche similaire : des mesures de mouvement sont estimées et servent à la fois à l'élaboration de notre nouveau modèle de vraisemblance $\mathrm{p}(z_k|c_k, z_{k-1}, c^i_{k-1})$ et à la construction de notre fonction de proposition.

Le reste de cette section décrit les expériences illustrant l'intérêt de l'utilisation d'une fonction de proposition basée sur des mesures de mouvement. En effet, du fait de l'utilisation de l'image courante lors de l'estimation du mouvement, ces mesures s'avérent plus adaptées pour modéliser les changements d'états et prédire les nouvelles configurations qu'un modéle dynamique basé uniquement sur les valeurs d'état. Néanmoins, nous souhaitons insister sur le fait que ces mesures ne viennent pas se substituer au modèle dynamique et ne nous dispensent donc pas de la modélisation de ce dernier. Considérons comme état c la position horizontale de la tête de la personne au premier plan dans la séquence de la figure 9. Notons c^v la vrai valeur obtenue par annotation manuelle de cette position dans 200 images. De plus, notons par ξ_k l'erreur de prédiction définie par :

$$\xi_k = c^v_k - \hat{c}_k \,, \tag{16}$$

où $\hat{c}_k$ représente une prédiction de l'état à l'instant k obtenue à partir des valeurs de cet état aux instants précédents, et éventuellement des images. Cet état prédit peut être calculé de deux maniéres. La premiére exploite le modéle auto-régressif simple suivant[2] :

$$\hat{c_k} = c^v_{k-1} + \dot{c}_{k-1} \;\text{ et }\; \dot{c}_{k-1} = c^v_{k-1} - c^v_{k-2} \,, \tag{17}$$

où $\dot{c}$ représente la dérivée temporelle de la variable d'état et modélise l'évolution de celui-ci. La seconde méthode exploite le mouvement inter-images pour prédire le nouvel état :

$$\hat{c}_k = c^v_{k-1} + \dot{c}^{mvt}_{k-1} \tag{18}$$

où $\dot{c}^{mvt}_{k-1}$ est calculé à partir des coefficients d'un modéle de mouvement affine estimés de façon robuste sur la région définie par c^v_{k-1} (*cf* partie 5.2).

La figure 5a présente l'erreur de prédiction obtenue avec le modéle AR. Comme on peut le constater, la prédiction est trés bruitée. De plus, il y a des pics d'erreurs importants (juqu'à 30 % de la taille de la tête). Pour pouvoir négocier ces pics d'erreurs dus à des changements brusques de régime, la variance du bruit dans la dynamique doit être surestimée afin de ne pas trop défavoriser les futurs états proches des valeurs recherchées. Sinon, seules les particules à proximité des états prédits risquent de survivre à l'étape de rééchantillonnage. Cependant, une variance élévée entraîne une perte de particules dans des régions de faible vraisemblance, ou la propagation de ces particules sur des ambiguïtés locales, ce qui peut conduire ultérieurement à des erreurs de suivi. En revanche, si l'on utilise le mouvement visuel inter-images, on obtient une réduction conjointe de la variance du bruit et de l'amplitude des pics (figure 5b).

Il y a un second avantage à utiliser le mouvement visuel. Remarquons tout d'abord que les valeurs de l'état aux instants précédents (c_{k-1}, c_{k-2} dans notre exemple) dont on dispose pour effectuer la prédiction de c_k sont bruitées en pratique. Dans l'approche standard avec modéle AR, dans l'équation (17), aussi bien l'état c_{k-1} que l'estimée de la dérivée $\dot{c}_{k-1}$ sont affectés par ce bruit, ce qui conduit à une amplification du bruit (*cf.* figure 5c). Lorsque l'on exploite le mouvement, l'estimation de $\dot{c}$ est peu affectée par le bruit (dont l'effet est de modifier légérement le support de la région utilisée pour estimer le mouvement), comme l'illustre la figure 6, ce qui résulte là encore en un processus de prédiction moins bruité (figure 5d).

Ainsi, bien que nécessitant plus de calculs, les prédictions exploitant le mouvement visuel sont souvent plus précises que les modéles auto-régressifs pour prédire la valeur future des paramétres de transformations géométriques. Par conséquent, elles sont un meilleur choix pour définir des fonctions de proposition. Ces observations sont confirmées par des expériences sur d'autres paramétres (position verticale, facteur d'échelle) et d'autres séquences. Enfin, ces conclusions illustrées en considérant uniquement la séquence d'états recherchée peut aussi s'appliquer à un ensemble de particules. Si celles-ci sont placées sur les modes d'une distribution liée à des observations visuelles, leurs prédictions en conformité avec le mouvement seront en général situées à proximité des nouveaux modes de la distribution dans l'image courante.

5. Mise en œuvre du modéle proposé

Le modéle graphique de la figure 1b est générique. La mise en œuvre spécifique que nous avons effectuée est décrite plus en détail dans cette partie.

2. Des modéles d'ordre supérieurs ont été employés également. Si ceux-ci diminuent la variance d'estimation, ils ont en revanche tendance à accentuer les pics d'erreur (amplitude, durée).

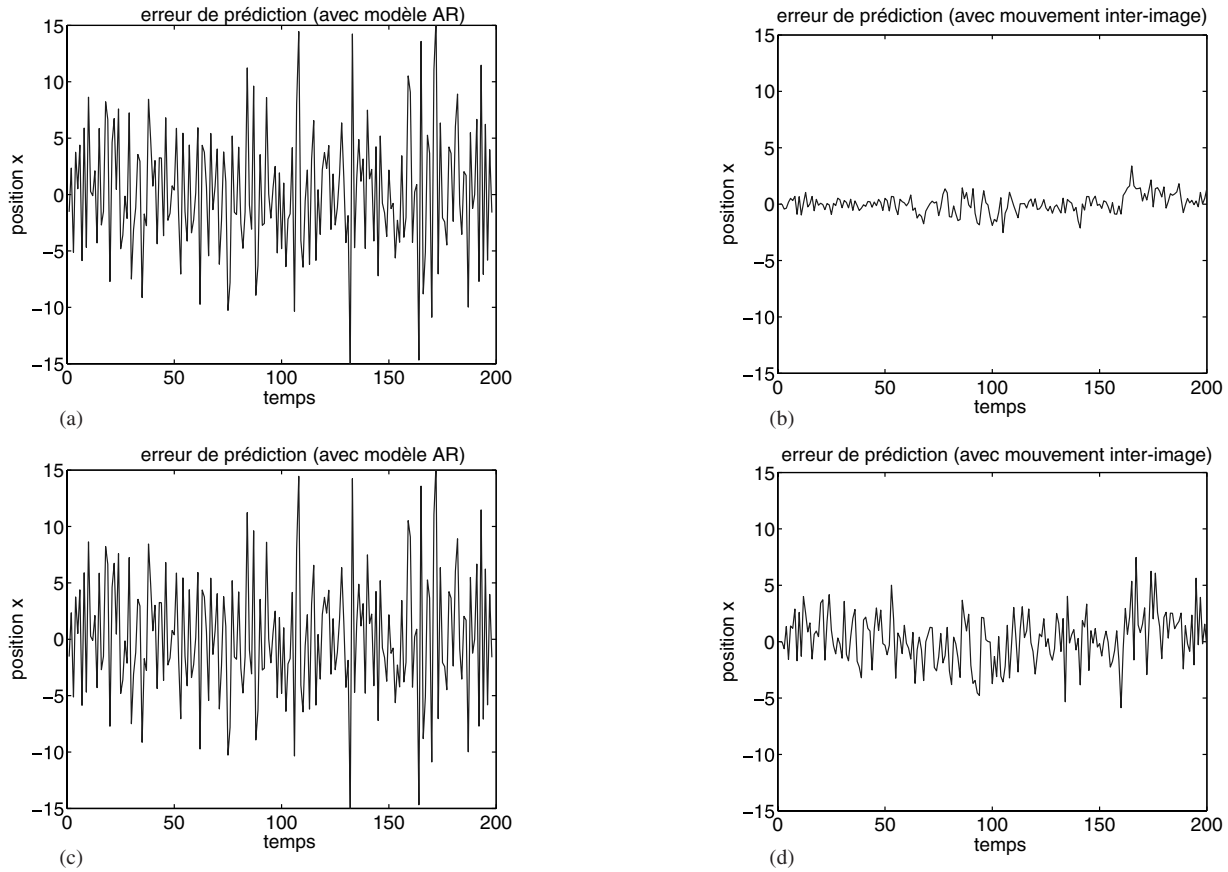

Figure 5. a) Erreur de prédiction de la position horizontale x, en utilisant un modéle AR d'ordre 2 (écart type estimé $\widehat{\sigma}_\xi = 2.7$). b) Erreur de prédiction, mais en exploitant l'estimation du mouvement inter-images ($\widehat{\sigma}_\xi = 0.83$). c) et d), pareil que a) et b), mais en ajoutant un bruit gaussien (d'écart type 2 pixels) sur les mesures x utilisées pour effectuer la prédiction. Avec le modéle AR d'ordre 2, (fig. c), aussi bien l'état précédent que l'estimation de la dérivée temporelle sont affectés par le bruit ($\widehat{\sigma}_\xi = 5.6$), alors qu'avec le mouvement visuel (fig. d), le bruit n'affecte principalement que la valeur de l'état à l'instant précédent ($\widehat{\sigma}_\xi = 2.3$).

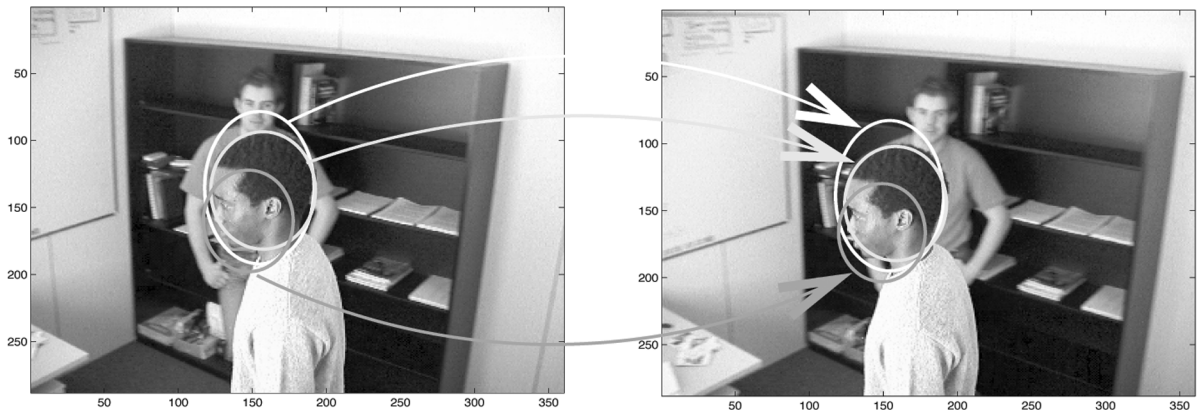

Figure 6. Exemple d'estimation du mouvement entre deux images à partir d'états bruités. Les trois ellipses correspondent à des valeurs d'état différentes. Bien que le support ne couvre qu'une partie de la tête et contienne une partie de fond texturé avec un mouvement différent, l'estimation du mouvement de la tête (ici, principalement une translation) est correcte.

5.1 Représentation de l'objet et espace d'état

Pour représenter l'objet, nous suivons une approche 2D, où l'objet est représenté par une région R et son contour. Ce dernier est caractérisé par une forme paramétrique, dans notre cas une ellipse. La transformation géométrique, auquel l'objet est sujet, est composée d'une translation $\mathbf{T}$, d'un facteur d'échelle s, et d'un rapport d'aspect e. La transformation est alors spécifiée par :

$$\mathcal{T}_\alpha \mathbf{r} = \begin{pmatrix} \mathbf{T}_x + x s_x \\ \mathbf{T}_y + y s_y \end{pmatrix}, \tag{19}$$

où $\mathbf{r} = (x, y)$ représente la position d'un point dans un repére de référence, $\alpha = (\mathbf{T}, s, e)$, et :

$$s = \frac{s_x + s_y}{2}, e = \frac{s_x}{s_y}, s_x = \frac{2es}{1+e} \text{ et } s_y = \frac{2s}{1+e} \tag{20}$$

L'état du filtre est défini par le modéle augmenté $c_k = (\alpha_k, \alpha_{k-1})$.

5.2 Estimation de mouvement

Comme évoqué dans la section précédente, nous utilisons des estimations de mouvement inter-images à la fois comme observations et pour échantillonner des nouvelles valeurs d'état. Plus précisément, un modéle de déplacement affine $\vec{d}_\Theta$ paramétrisé par $\Theta = (a_i)_{i=1..6}$ est calculé à l'aide d'une méthode d'estimation multirésolution décrite dans [16, 17][3]. $\vec{d}_\Theta$ est défini par :

$$\vec{d}_\Theta \mathbf{r} = \begin{pmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{pmatrix}. \tag{21}$$

La méthode combine les avantages d'une approche multirésolution et d'un schéma incrémental de type Gauss-Newton. La robustesse est assurée en minimisant un critére de M-estimateur :

$$\widehat{\Theta}(c_{k-1}) = \text{argmin}_\Theta \sum_{\mathbf{r} \in R(c_{k-1})} \rho\left(\text{DFD}_\Theta(\mathbf{r})\right)$$

$$avec \quad \text{DFD}_\Theta(\mathbf{r}) = z_k(\mathbf{r} + \vec{d}_\Theta \mathbf{r}) - z_{k-1}(\mathbf{r}), \tag{22}$$

où z_k et z_{k-1} représentent les images à deux instants consécutifs et $\rho(x)$ est un estimateur robuste (borné pour de grandes valeurs de x). Étant donné la robustesse de l'estimateur, une imprécision sur la définition de la région support $R(c_{k-1})$ due à une valeur bruitée de l'état affecte peu l'estimation (*cf.* figure 6). À partir des paramétres estimés, nous pouvons calculer une estimation $\widehat{\dot{\alpha}}_k$ de l'évolution entre deux instants des paramétres de la transformation choisie. En supposant que les coordonnées dans l'équation (21) soient exprimées par rapport au centre de l'objet courant (localisé aux coordonnées $\mathbf{T}$ dans l'image), nous obtenons en premiére approximation :

$$\begin{cases} \dot{\mathbf{T}}_x = a_1 \\ \dot{\mathbf{T}}_y = a_4 \end{cases} \text{ et } \begin{cases} \dot{s}_x = a_2 s_x \\ \dot{s}_y = a_6 s_y \end{cases} \text{ et } \begin{cases} \dot{s} = \frac{s}{1+e}(a_2 e + a_6) \\ \dot{e} = e(a_2 - a_6) \end{cases}. \tag{23}$$

Grâce à ces paramétres, on peut donc prédire la valeur que prendront a à de l'état à l'instant k s'obtient par :

$$\widehat{\alpha}_k(\alpha_{k-1}) = \alpha_{k-1} + \widehat{\dot{\alpha}}_{k-1}. \tag{24}$$

3. Nous utilisons le code disponible sur le site http://www.irisa.fr/vista.

Dans la suite de l'article, nous ne rappellerons pas systématiquement la dépendance implicite de $\widehat{\dot{\alpha}}_k$ vis-à-vis de α_{k-1}. Par ailleurs, bien qu'elle ne soit pas utilisé dans les expériences présentées, la matrice de covariance des paramétres estimés peut également être calculée. Dans des approches avec un espace d'états plus grand, elle pourrait s'avérer utile pour tenir compte des incertitudes et des problémes d'optimisation sous-contrainte.

5.3 Modélisation de la vraisemblance des données

Pour appliquer le nouveau modéle, nous supposons que les données z_k sont de deux types : des mesures orientées objets z_k^o (contours), et les mesures d'intensités z_k^g. Ensuite, nous considérons la vraisemblance de données suivante :

$$\text{p}(z_k | z_{k-1}, c_k, c_{k-1}) = \text{p}(z_k^o, z_k^g | z_{k-1}^o, z_{k-1}^g, c_k, c_{k-1}) \tag{25}$$

$$= \text{p}(z_k^o | z_k^g, z_{k-1}^o, z_{k-1}^g, c_k, c_{k-1}) \text{p}(z_k^g | z_{k-1}^o, z_{k-1}^g, c_k, c_{k-1}) \tag{26}$$

$$= \text{p}_{\text{sh}}(z_k^o | c_k) \text{p}(z_k^g | z_{k-1}^g, c_k, c_{k-1}). \tag{27}$$

La derniére équation résulte de deux hypothéses : la premiére suppose que les mesures de contour sont indépendantes des informations de niveaux de gris conditionellement à la séquence d'états. Ce choix découple la modélisation de la corrélation existant entre deux images consécutives d'un même objet, dont le but implicite est de s'assurer que la trajectoire de l'objet suit le flux optique, de la modélisation de la forme de l'objet. Comme l'évaluation de la vraisemblance de l'objet fait intervenir des données sur le pourtour de l'objet alors que le terme de corrélation s'applique essentiellement à l'intérieur de l'objet, cette hypothése est valide. La seconde hypothése utilisée est celle de l'indépendance temporelle des mesures de contours. Elle est raisonable dans la mesure où la fonction d'autocorrélation temporelle pour ce type de mesure est trés pointue. Nous décrivons maintenant les deux modéles d'observations adoptés.

Mesures de contour

Ce terme suit le modéle présenté dans [5], où des mesures de contours sont calculées le long de L lignes normales à une ellipse hypothèse supposée située sur un fond bruité. Pour chaque ligne l, on obtient un vecteur $\{\nu_m^l\}$ de points de contour candidats détectés et relatifs au point ν_0^l se trouvant sur le contour hypothése. Les hypothèses standards [5] conduisent à :

$$\text{p}_{\text{sh}}(z_k | c_k) \propto \prod_{l=1}^{L} \max\left(K_s, \exp(-\frac{\|\hat{\nu}_m^l - \nu_0^l\|^2}{2\sigma_s^2})\right), \tag{28}$$

où $\hat{\nu}_m^l$ est le point de contour le plus proche de ν_0^l, et K_s une constante introduite quand aucun contour n'est détecté. Dans les expériences, nous avons utilisé $L = 16$ lignes de recherche, l'espace de recherche des contours candidats est de 10 pixels

vers l'intérieur et l'extérieur de l'ellipse, σ_s a été fixé à la moitié de l'espace de recherche (*i.e.* 5), et $K_s = \exp^{-2}$.

Mesure de corrélation temporelle

Pour modéliser ce terme, deux choix se présentent :
- le premier consiste à extraire des mesures dans l'espace d'états. Habituellement, celles-ci s'obtiennent par seuillage et/ou extraction de maxima locaux d'une fonction d'intérêt [1, 21]. Un avantage de cette approche est de produire des vraisemblances qui se comportent bien (au sens où elles impliquent uniquement quelques modes). Un inconvénient est que le temps d'extraction des observations peut être long.
- dans le second cas, les régions de niveaux de gris sont directement comparées, aprés recalage en fonction de la valeur de l'état (voir l'équation (11)). L'avantage est de fournir des vraisemblances plus détaillées qui peuvent se calculer directement à partir des données.

Dans cet article, nous exploitons les deux options, en supposant que les observations se composent des paramétres $\widehat{\alpha}_k$ obtenus grâce au modèle de mouvement estimé (*cf.* équation 24), et des mesures d'intensité locales $\tilde{z}^g_{c_k}$. En supposant ces mesures indépendantes, nous modélisons alors le terme de correlation par (on rappelle que $c_k = (\alpha_k, \alpha_{k-1})$) :

$$\mathrm{p}(z^g_k | z^g_{k-1}, c_k, c_{k-1}) \propto \mathrm{p}_{c1}(\widehat{\alpha}_k, \alpha_k)\mathrm{p}_{c2}(\tilde{z}^g_{c_k}, \tilde{z}^g_{c_{k-1}}) \tag{29}$$

avec :

$$\mathrm{p}_{c1}(\widehat{\alpha}_k, \alpha_k) = \mathcal{N}(\widehat{\alpha}_k : \alpha_k, \Lambda_{\xi_p}) \tag{30}$$

$$\mathrm{p}_{c2}(\tilde{z}^g_{c_k}, \tilde{z}^g_{c_{k-1}}) = \mathrm{Z}^{-1}\exp^{-\lambda_c \mathrm{d}_c^2(\tilde{z}^g_{c_k}, \tilde{z}^g_{c_{k-1}})} \tag{31}$$

où $\mathcal{N}(. : \mu, \Lambda)$ représente une distribution gaussienne de moyenne μ et de variance Λ, d_c une distance entre imagettes, $\Lambda_{\xi_p} = diag(\sigma^2_{\xi_p,j})$ la matrice de covariance des mesures. Z est une constante de normalisation qui dépend de la distance choisie. Dans le cas où une distance L2 est utilisée, ce qui correspond à l'hypothèse d'un bruit gaussien dans l'équation 12, $\lambda_c = 1/(2\sigma^2_\eta)$ et $\mathrm{Z} = (2\pi\sigma^2_\eta)^{-N/2}$, où N est le nombre de pixels d'une imagette. Pour d'autres distances, Z peut se calculer par apprentissage, par exemple comme :

$$\mathrm{Z} = \int_{z',z''} \exp^{-\lambda_c \mathrm{d}_c^2(z',z'')} \mathrm{d}z' \mathrm{d}z'' \tag{32}$$

où l'intégrale porte sur des paires d'imagettes consécutives associées au même objet et extraites de séquences d'apprentissage [30]. En pratique cependant, nous avons supposé que celle-ci était constante pour toutes les imagettes d'objet, ce qui suppose implicitement que toutes les imagettes $\tilde{z}$ sont équiprobables.

Le premier terme de vraisemblance (équation 30) compare les paramétres prédits à l'aide du mouvement estimé avec ceux de l'échantillon et repose sur un modéle gaussien dans l'espace des paramètres. Pour introduire une composante non gaussienne, la vraisemblance, nous proposons d'utiliser un second terme basé sur la distance de similarité d_c entre les imagettes extraites à l'aide de c_k et c_{k-1}. La motivation de ce terme est illustrée par les trois configurations dessinées sur l'image de gauche de la figure 6. Alors que les trois configurations prédites qui leur sont associées (image de droite) ont la même vraisemblance selon le terme p_{c1}, le second terme p_{c2} attribuera une vraisemblance plus faible aux deux configurations dont le support recouvre en partie le fond de l'image, ce dernier ayant un mouvement différent de celui de la tête.

La définition de p_{c2} repose sur la spécification d'une distance entre imagettes. De nombreuses distances de ce type ont été définies dans la littérature [28, 30]. Le choix de cette distance doit prendre en compte les considérations suivantes :
1. la distance doit modéliser l'information de mouvement sous-jacente, *i.e.* la distance doit augmenter si l'erreur de prédiction augmente ;
2. l'aspect aléatoire du processus de prédiction dans le filtre MCS produit rarement des configurations correspondant à la mise en correspondance optimale. Ceci est particulièrement vrai dans le cas où peu de particules sont utilisées ;
3. quand l'objet et le fond ont des mouvements différents, des particules dont la région associée couvre à la fois l'objet et le fond doivent avoir une faible vraisemblance.

Pour ces raisons, nous avons trouvé qu'il était préférable de ne pas utiliser de distance robuste, comme une distance L1 saturée ou une distance de Haussdorf [30]. De plus, il fallait éviter des distances qui pourraient favoriser *a priori* des imagettes de contenus spécifiques. C'est le cas par exemple de la distance quadratique L2 (qui correspond à un bruit Gaussien additif dans l'équation (12)) qui génére généralement des distances plus faibles avec des imagettes possédant des régions uniformes importantes. Pour éviter cet effet, nous utilisons le coefficient de corrélation croisé normalisé, défini par :

$$\mathrm{d}_c(\tilde{z}_1, \tilde{z}_2) = 1 - \frac{\sum_{\mathbf{r}\in R}\left(\tilde{z}_1(\mathbf{r}) - \bar{\tilde{z}}_1\right)\cdot\left(\tilde{z}_2(\mathbf{r}) - \bar{\tilde{z}}_2\right)}{\sqrt{\mathrm{Var}(\tilde{z}_1)}\sqrt{\mathrm{Var}(\tilde{z}_2)}} \tag{33}$$

où $\bar{\tilde{z}}_1$ représente la moyenne de $\tilde{z}_1$. La constance λ_c (équation 31) est fixé à 20.

Soulignons ici que la méthode ne fait pas de mise en correspondance de prototypes, comme dans [28]. Aucun prototype d'objet n'est défini hors-ligne ou au début de la séquence, et le suiveur ne maintient pas à jour un unique prototype. Ainsi, le terme de corrélation n'est pas spécifique de l'objet, excepté au travers de la définition de la région de référence R. Une particule localisée sur le fond de l'image peut recevoir une vraisemblance importante si le mouvement prédit est en accord avec le mouvement du fond. Néanmoins, la méthodologie pourrait être étendue en autorisant la région R à varier au cours du temps, ou en introduisant un bruit variant spatialement dans la définition de la corrélation.

5.4 Modéle dynamique

Nous utilisons un modéle auto-régressif d'ordre 2 standard (*cf.* équation 17) pour chacune des composantes de α. Néamoins, pour tenir compte des données aberrantes (*i.e.* changements de régime abrupts) et réduire la sensibilité du modéle *a priori* dans la queue de la distribution, nous modélisons le bruit avec une distribution de Cauchy $\rho_c(x,\sigma^2) = \frac{\sigma}{\pi(x^2+\sigma^2)}$. Cela conduit à :

$$\mathrm{p}(c_{k+1}|c_k) = \prod_{j=1}^{4} \rho_c\left(\alpha_{k+1,j} - (2\alpha_{k,j} - \alpha_{k-1,j}), \sigma^2_{\xi_{d,j}}\right). \quad (34)$$

où $\sigma^2_{\xi_{d,j}}$ désigne la variance de la $j^{\text{iéme}}$ composante. De plus, comme le mouvement estimé est plus fiable que le modèle *a priori*, nous avons fixé la valeur de $\sigma_{\xi_{d,j}}$ à trois fois la valeur de $\sigma_{\xi_{p,j}}$.

5.5 Distribution de proposition

Comme le motive la partie 4.2, la fonction de proposition exploite le mouvement visuel estimé. Le nouvel état $c_k = (\alpha_k, \alpha_{k-1})$ est échantillonné suivant la loi définie par :

$$q(c_k|c^i_{0:k-1}, z_{1:k}) \propto \delta(\alpha_{k-1} - \alpha^i_{k-1}) \times \mathcal{N}(\alpha_k : \widehat{\alpha}_k(\alpha^i_{k-1}), \Lambda_{\xi_p}) \quad (35)$$

où δ représente la fonction de Dirac. Cela signifie que les nouveaux échantillons sont tirés autour de la valeur prédite grace au mouvement visuel observé.

6. Résultats

Pour illustrer la méthode, nous avons considéré trois séquences de suivi de tête. Dans les trois cas, le suiveur est initialisé à la main. Pour distinguer l'effet des différents éléments du modéle, nous avons considéré les trois configurations suivantes :

- suiveur de forme M1 : ce cas correspond à l'algorithme CONDENSATION standard [5], dans lequel le même modéle AR avec un bruit gaussien est utilisé pour la dynamique et la fonction de proposition, et la vraisemblance est celle du modéle de contour uniquement.
- suiveur de forme+corrélation M2 : il s'agit de l'algorithme CONDENSATION, dans lequel le terme de vraisemblance de mouvement implicite est ajouté à la fonction de vraisemblance (égale à $\mathrm{p_{sh}}.\mathrm{p}_{c2}$). Cette méthode n'utilise pas de mesures de mouvement explicites.
- suiveur avec fonction de proposition basée sur le mouvement M3 : il s'agit du modéle complet. Les échantillons sont tirés à partir de la fonction de proposition définie par l'équation (35), et la remise à jour des poids se fait avec l'équation (5). Aprés simplification, la remise à jour devient :

$$w^i_k \propto w^i_{k-1}\ \mathrm{p_{sh}}(z_k|c^i_k)\mathrm{p}_{c2}(\tilde{z}_{c^i_k}, \tilde{z}_{c^i_{k-1}})\mathrm{p}(c^i_k|c^i_{k-1}) \quad (36)$$

Avec cette méthode, l'estimation de mouvement n'est pas effectuée pour chaque particule puisque l'algorithme d'estimation du mouvement est robuste vis-à-vis des variations de la région support. À chaque instant, les particules sont groupées en K groupes avec l'algorithme des K moyennes. Les paramétres de mouvement sont estimés à partir de la configuration moyenne de chaque groupe et servent pour la fonction de proposition de chaque particule du groupe. En pratique, nous utilisons $\max(20, N/10)$ groupes.

Actuellement, avec 200 particules, le suiveur de forme M1 fonctionne en temps réel (sur une machine avec un Pentium IV à 2.5GHz), le suiveur M2 à 20 images/s, et le modéle complet à 8 images/s environ. Dans les expériences, tous les paramètres communs sont identiques. Seuls le nombre d'échantillons N_s et la variance dans la fonction de proposition (et par conséquent dans le modéle dynamique) seront modifiés. On notera $\sigma_{\mathbf{r}}$ l'écart type du bruit des composantes de translation (*i.e.* $\sigma_{\xi_{p,1}} = \sigma_{\xi_{p,2}} = \sigma_{\mathbf{r}}$), et σ_s celui de la composante d'échelle. Le bruit lié au paramétre d'aspect est gardé fixe, à la valeur $\sigma_{\xi_{p,4}} = 0.01$.

Le premier exemple, figure 7, comprend 64 images de taille 240x320 et illustre l'apport de la méthode en cas d'ambiguïtés. Compte tenu de la présence d'une forte texture dans le fond, les observations sont clairement multimodales. Ainsi, quel que soit le nombre de particules utilisées et la variance dans le modéle dynamique, l'utilisation du seul modèle d'objet ne permet jamais d'effectuer un suivi correct au-delà de l'instant t_{12}. En revanche, avec des valeurs de bruit faibles dans la dynamique (deuxiéme ligne fig. 7), l'utilisation du suiveur M2 permet d'effectuer un suivi correct de la tête en dépit des mouvements de caméra et de la tête, de la variation d'apparence de la tête, et d'occlusions partielles. Pour des variances plus élevées, cependant, le suiveur M2 échoue, au contraire du suiveur M3 qui réussit dans tous les cas.

Le second exemple est une séquence de 12 secondes comprenant 330 images (*cf.* figure 8) extraite d'une vidéo amateur avec caméra tenue à la main. Le tableau 1 rapporte les performances

Tableau 1. Taux de suivis réussis (en %, à partir de 50 essais avec des graines différentes dans le générateur de nombres aléatoires) en fonction des paramétres de dynamique et du nombre d'échantillons. Dans les expériences D1 à D4, $N_s = 500$. Les paramétres de dynamique $(\sigma_{\mathbf{r}}, \sigma_s)$ sont : D1 (2,0.01), D2 (3,0.01), D3(5,0.01) D4(8,0.02). Dans les expériences S1 et S2, la dynamique est caractérisée par (5,0.01), et les nombres d'échantillons sont 250 (S1) et 100 (S2).

Tracker	D1	D2	D3	D4	S1	S2
CONDENSATION	88	60	2	0	0	0
M2	100	100	98	96	96	58
M3	90	100	100	100	100	100

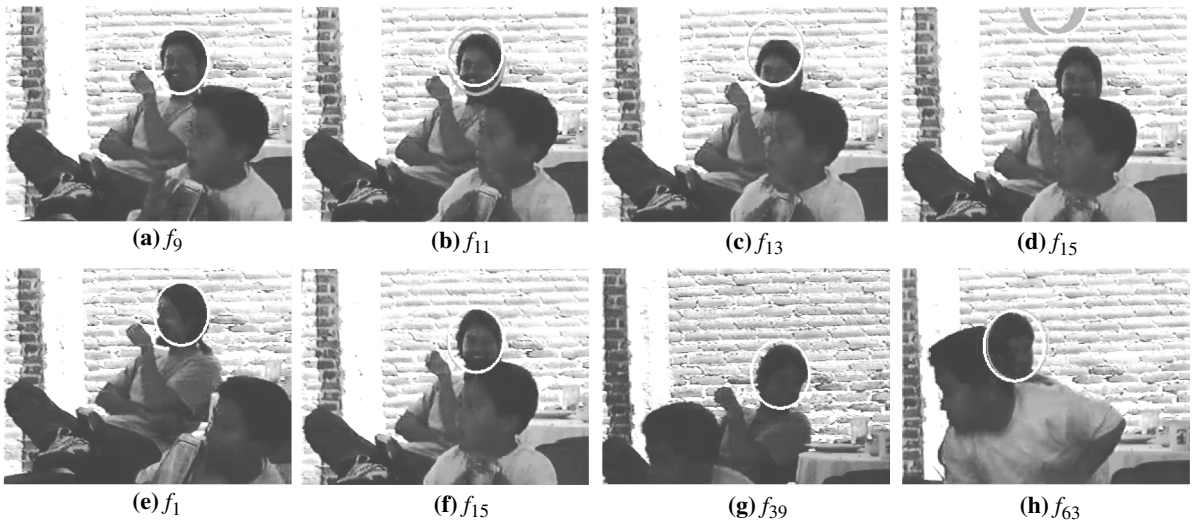

Figure 7. Suivi de tête 1 : premières ligne : suiveur de forme M1 (CONDENSATION). deuxième ligne : suiveur de forme M2, avec les paramètres $N_s = 100$, $(\sigma_{\mathbf{r}},\sigma_s) = (1.,0.005)$.

Figure 8. Suivi de tête 2 avec $N_s = 500$, $(\sigma_{\mathbf{r}},\sigma_s) = (5,0.01)$. *Première ligne :suiveur de forme M1. Seconde et troisième lignes : suiveur de forme M2. En rouge (sombre), configuration moyenne. En jaune (clair), particules les plus vraisemblables.*

de suivi des trois suiveurs pour différents paramétres de dynamique et nombre d'échantillons. Une erreur de suivi se produit lorsque le suiveur se perd sur une partie de l'image différente de la tête. Comme on peut le constater, alors que l'algorithme CONDENSATION donne des résultats convenables avec une variance de bruit adaptée (D1), ses performances se dégradent trés vite, même pour de petites augmentations du niveau de bruit dans la dynamique (D2 à D4). La premiére ligne de la figure 8 illustre une erreur de suivi typique, due à la petite taille de la tête au début de la séquence, à un faible contraste à la gauche de cette tête, et aux ambiguïtés de contour. Par contraste, le suiveur de forme et corrélation M2 effectue un suivi correct dans

Figure 9. Suiveur M3 avec fonction de proposition basée sur le mouvement, $N_s = 2\,000$, $(\sigma_{\mathbf{r}},\sigma_s) = (8,0.03)$. Suivi aux instants t_2, t_{40}, t_{85}, t_{100}, t_{130}, t_{145}, t_{170}, t_{195}, et t_{210}. En rouge (sombre), configuration moyenne ; en vert, mode ; en jaune (clair), particules les plus vraisemblables.

presque toutes les circonstances, démontrant sa robustesse vis-à-vis des ambiguïtés, des mesures partielles (vers l'instant t_{250}), et d'une occultation partielle (fin de la séquence). Ce n'est que lorsque le nombre d'échantillons est faible (100 dans S2) que les performances de ce suiveur se dégradent. Les échecs se produisent à différents instants de la séquence. Finalement, dans toutes les expériences, le suiveur M3 atteint une performance de suivi supérieure à 90 %, même avec 100 échantillons, ce qui démontre l'intérêt de l'utilisation du mouvement dans la fonction de proposition.

La dernière séquence (figure 9) illustre plus clairement le bénéfice de la fonction de proposition. Cette séquence de 24 secondes acquise à 12 trame/s est particuliérement difficile : la personne effectue plusieurs tours sur elle même[4], le mouvement de la caméra subit des variations brusques (translation, zoom avant et arrière), les variations d'échelle sont importantes, il y a une absence de contours de la tête lorsque celle-ci se trouve devant la bibliothèque. Pour ces raisons, le suiveur CONDENSATION se perd très rapidement. D'un autre coté, le suiveur M2 suit correctement la tête au début, mais se perd lorsque la personne se déplace devant les étagéres (trames t_{130}-t_{145}), à cause d'un manque de mesures de contours couplé avec un important facteur de zoom. Ce dernier probléme est résolu par la fonction de proposition basée sur le mouvement, qui permet de mieux appréhender les variations rapides d'état, et conduit à un suivi correct jusqu'à la fin de la séquence (instant t_{340}).

7. Discussion, conclusion

Dans cet article, nous avons présenté deux principes pour incorporer des informations sur le mouvement visuel dans un filtre à particules. Le premier repose sur un nouveau modéle graphique que nous avons proposé et qui permet de prendre en compte la corrélation temporelle qui existe entre deux images successives d'un même objet. Nous montrons d'une part que ce modéle s'insére aisément dans la méthodologie du filtrage de Monte-Carlo séquentiel et d'autre part que le nouveau terme de vraisemblance introduit peut être exploité pour modéliser le mouvement visuel de façon implicite ou explicite. Le second principe proposé tire profit de l'estimation de paramétres de mouve-

4. Ceci est un cas difficile pour la méthode, qui reçoit des informations contradictoires : l'intérieur de la tête indique un mouvement vers la droite alors que le contour extérieur de la tête reste statique.

ment afin de prédire plus précisément les nouvelles valeurs de la variable d'état. Cette approche permet de définir une meilleure fonction de proposition qui tient compte des nouvelles données image. Dans l'ensemble, l'approche proposée permet d'appréhender des changements de mouvement rapides et imprévus, de supprimer des ambiguïtés locales qui apparaissent lors de l'utilisation de modèles d'objets génériques basés sur la forme, et de réduire la sensibilité de l'algorithme vis-à-vis des différents paramétres du modéle *a priori*.

Les expériences que nous avons menées ont démontré les bénéfices de notre approche. Cependant, ceux-ci ne doivent pas masquer le fait que les performances de suivi dépendent du choix d'un modèle d'objet robuste. Par exemple, un suiveur exploitant un modèle de couleur [20], lorsque son utilisation est appropriée, fonctionne mieux en général que le modèle de contour présenté dans cet article. Néanmoins, même avec un tel suiveur, des expériences que nous avons menées par ailleurs montrent l'intérêt d'utiliser les méthodes présentées.

Plusieurs directions de recherche peuvent être suivies pour construire des modèles de vraisemblance plus précis ou adaptatifs. Une première direction passe par l'utilisation de méthodes d'apprentissage à l'aide d'exemples, surtout si l'on s'interesse à un objet spécifique (comme une tête). La fusion de mesures multimodales ou effectuées à différentes positions spatiales et temporelles en est une autre. Par ailleurs, le problème spécifique de la gestion d'occlusions très importantes ou complètes, non abordé dans le présent article, fait également l'objet de nombreuses recherches.

Enfin, nous avons montré que l'utilisation de mesures explicites de mouvement dans la fonction de proposition permet d'améliorer l'efficacité d'échantillonnage de l'algorithme. Cette approche est générale et peut par exemple s'appliquer au suivi d'objets déformables en intégrant les mesures de mouvement le long des contours d'une forme, ainsi qu'il est décrit dans [14]. Dans ce cas cependant, la robustesse de l'estimation des variations temporelles de paramètres de déformation très précis à partir des mesures de mouvement bas-niveau doit être démontrée. De manière alternative, l'utilisation d'un schéma hybride, dans lequel certains paramètres de l'espace d'état (translation, échelle, rotation,…) sont échantillonnés à partir d'une fonction de proposition reposant sur des mesures de mouvement visuel, et les autres sont tirés à partir d'un modéle AR, pourrait s'avérer plus appropriée.

Références

[1] E. ARNAUD et E. MÉNIN, Optimal importance sampling for tracking in image sequences : applications to point tracking. Dans *Proc. of 8th Eur. Conf. Computer Vision*, Prague, Czech Republic, Mai 2004.

[2] E. ARNAUD, E. MÉNIN et B. CERNUSHI-FRIAS, Filtrage conditionnel pour la trajectographie dans des séquences d'images – application au suivi de points. Dans *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, Toulouse, France, Janvier 2004.

[3] S. ATULAMPALAM, S. MASKELL, N. GORDON et T. CLAPP, A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking, *IEEE Trans. Signal Process*, 50(2): 100-107, 2001.

[4] M.J. BLACK et A.D. JEPSON, Eigentracking : robust matching and tracking of articulated objects using a view based representation, *Int. J. Computer Vision*, 21(1): 63-84, 1998.

[5] A. BLAKE et M. ISARD, *Active Contours*, Springer, 1998.

[6] D. COMANICIU, V. RAMESH et P. MEET, Real-time tracking of non-rigid objects using mean shift, Dans *Proc. IEEE. Conf. Computer Vision Pattern Recognition*, 142-151, 2000.

[7] A. DOUCET, N. de FREITAS, et N. GORDON, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.

[8] A. DOUCET, S. GODSILL, et C. ANDRIEU, On sequential Monte Carlo sampling methods for bayesian filtering, *Statistics and Computing*, 10(3): 197-208, 2000.

[9] D. GATICA-PEREZ, G. LATHOUD, I. McCOWAN, et J.-M. ODOBEZ, A Mixed-State I-Particle Filter for Multi-Camera Speaker Tracking. Dans *IEEE Int. Conf. on Computer Vision Workshop on Multimedia Technologies for E-Learning and Collaboration (ICCV-WONTEC)*, Nice, France, 2003.

[10] N. GORDON, D. SALMON, et A.F.M. SMITH, Novel approach to nonlinear/non-gaussian bayesian state estimation, *IEE Proc. F.*, 140(2) : p. 107-113, 1993.

[11] M. ISARD et A. BLAKE, Contour tracking by stochastic propagation of conditional density, Dans, *Proc. of 4th Eur. Conf. Computer Vision.*, vol. 1, 343-356, 1996.

[12] M. ISARD et A. BLAKE, Icondensation : Unifying low-level and high-level tracking in a stochastic framework. Dans *Proc. of 5th Eur. Conf. Computer Vision., Lect. Notes in Compter Sciences, vol. 1406*, vol. 1 p. 893-908, Freiburg, Allemagne, 1998.

[13] A.D. JEPSON., D.J. FLEET et T.F. El-MARAGHI, Robust on-line appearance models for visual tracking. *IEEE Trans. Pattern Anal Machine Intell*, 25(10): 661-673, Octobre 2003.

[14] C. KERVRANN, F. HEITZ et P. PÉREZ, Statistical model-based estimation and tracking of non-rigid motion. Dans *Proc. 13th Int. Conf. Pattern Recognition*, 244-248, Vienne, Autriche, Août 1996.

[15] H.T. NGUYEN, M. WORRING et R. VAN DEN BOOMGAARD, Occlusion robust adaptive template tracking. Dans *Proc. IEEE Int. Conf. Comp. Vision*, 678-683, Vancouver, Juillet 2001.

[16] J.-M. ODOBEZ et P. BOUTHEMY, Estimation robuste multi-échelle de modèles paramétrés de mouvement sur des scènes complexes, *Traitement du Signal*, Vol. 12 (N°2): 113-128, 1995.

[17] J.-M. ODOBEZ et P. BOUTHEMY, Robust multiresolution estimation of parametric motion models, *JI of Visual Com. and Image Representation*, 6(4): 348-365, Décembre 1995.

[18] W. PIECZYNSKI, Pairwise markov chains, *IEEE Trans. Pattern Anal. Machine Intell.*, 25(5): 634-640, Mars 2003.

[19] M.-K. PITT et N. SHEPHARD, Filtering via simulation : Auxiliary particle filters, *Journal of the American Statistical Association*, 94(446): 590-599, 1999.

[20] P. PÉREZ, C. HUE, J. VERMAAK and M. GANGNET, Color-based probabilistic tracking. Dans *Proc. of 7th Eur. Conf. Computer Vision, Lect. Notes in Compter Sciences, vol. 2350*, 661-675, Copenhague, Danemark, Juin 2002.

[21] P. PÉREZ, J. VERMAAK et A. BLAKE, Data fusion for visual tracking with particles. *Proc. IEEE*, 92(3): 495-513, 2004.

[22] A. RAHIMI, L.P. MORENCY et T. DARRELL, Reducing drift in parametric motion tracking. Dans. *Proc. IEEE Int. Conf. Comp. Vision*,15-322, Vancouver, Juillet 2001.

[23] Y. RAJA, S. McKENNA et S. GONG, Colour model selection and adaptation in dynamic scenes. Dans *Proc. of 5th Eur. Conf. Comp. Vision, Lect. Notes in Compter Science, vol. 1406*, 460-474, Freiburg, Allemagne, 1998.

[24] Y. RUI et Y. CHEN, Better proposal distribution : object tracking using unscented particle filter. Dans *Proc. IEEE Conf. Comp. Vision Pattern Recognition*, 786-793, Décembre 2001.

[25] H. SIDENBLADH et M.J. BLACK, Learning image statistics for bayesian tracking. Dans *Proc. 8th IEEE Conf. Comp. Vision*, vol. 2, 709-716, Canada, Juillet 2001.

[26] H. SIDENBLADH, M.J.. BLACK. et D.L. FLEET., Stochastic tracking of 3d human figures using 2d image motion. Dans *Proc. Eur Conf. Comp. Vision*, vol. 2, 702-718, Dublin, Ireland, Juin 2000.

[27] C. SMINCHISESCU et B. TRIGGS., Covariance scaled sampling for monocular 3d body tracking. Dans *Proc. IEEE Conf. Computer Vision Pattern Recognition*, vol. 1, 447-454, Décembre 2001.

[28] J. SULLIVAN et J. RITTSCHER., Guiding random particles by deterministic search. Dans *Proc. 8th IEEE Int. Conf. Comp. Vision*, 323-330, Vancouver, Juillet 2001.

[29] H. TAO, H.S. SAWHNEY et R. KUMAT., Object tracking with bayesian estimation of dynamic layer representations. *IEEE Trans. Pattern Anal. Machine Intell*, 24(1): 75-89, 2001.

[30] K. TOYAMA et A. BLAKE., Probabilistic tracking in a metric space. Dans. *Proc. 8th IEE Int. Conf. Comp. Vision*, vol. 2, 50-57, Vancouver, Juillet 2001.

[31] J. VERMAAK, P. PÉREZ, M. GANGNET et A. BLAKE., Towards improved observation models for visual tracking : Selective adaptation. Dans. *Proc. of 7th Eur. Conf. Comp. Vision, Lect. Notes in Compter Science, vol. 2350*, 645-660, Copenhague, Danemark, 2002.

[32] Y. WU et T. HUANG, A co-inference approach for robust visual tracking. Dans *Proc. 8th IEEE Int.. Conf. Comp. Vision*, vol. 2, 26-33, Vancouver, Juillet 2001.

(*) Cet article, accepté pour publication en 2007, est publié tardivement pour des raisons techniques indépendantes de la volonté de l'auteur.

J.-M. **Odobez**

Jean-Marc Odobez est né en France en 1968. Il a obtenu en 1990 le diplôme d'ingénieur de l'École Nationale Supérieure des Télécommunications de Bretagne (ENSTBr), et en 1994 son doctorat en Traitement du Signal de l'université de Rennes I. Sa thèse, réalisée à l'IRISA de Rennes, porte sur l'analyse du mouvement dans des séquences d'images à l'aide d'outils statistiques (estimateur robuste, modèles de Markov). Après un an passé au laboratoire GRASP de l'Université de Pennsylvanie, Philadelphie, USA, il a été entre 1996 et 2001 Maître de conférences à l'université du Maine au Mans. Depuis 2001, il occupe un poste de chercheur à l'IDIAP, institut de recherche suisse, où il travaille à la modélisation d'images ou de documents multimédia, ainsi qu'à la reconnaissance de l'activité des personnes et de leurs interactions.

D. **Gatica-Perez**

Daniel Gatica-Perez a obtenu le Bachelor of Science en Ingénierie Electronique de l'université Puebla de Mexico en 1993, le Master of Science de l'université nationale de Mexico en 1996, et le doctorat en ingénierie électrique de l'Université de Washington, Seattle, USA, en 2001. Il a rejoint l'institut de recherche IDIAP en janvier 2002, où il est chercheur. Sa recherche s'articule autour de trois axes : la vision par ordinateur, le traitement de signaux multimodaux, et la fouille d'informations dans des documents multimédias.

S. **Ba**

Sileye Ba a obtenu le DEA de mathématiques appliquées option traitement du signal de l'Université de Dakar en 2000, et le Dea en mathématiques, vision et apprentissage de l'ENS Cachan de Paris en 2002. Depuis octobre 2002, il est doctorant à l' institut de recherche IDIAP. L'objet de sa thèse est le suivi d'individus et la reconnaissance de leurs activités dans des séquences vidéos à l'aide de méthodes de Monte Carlo séquentielles.