

# Paramétrisation des fonctions de vraisemblance et choix de filtres particulières pour le suivi visuel

Mathias FONTMARTY<sup>1</sup>

Frédéric LERASLE<sup>1</sup>

Patrick DANÈS<sup>1</sup>

<sup>1</sup> CNRS ; LAAS ; 7 avenue du colonel Roche, F-31077 Toulouse, France  
Université de Toulouse ; UPS, INSA, INP, ISAE ; LAAS ; F-31077 Toulouse, France

{mfontmar, lerasle, danes}@laas.fr

## Résumé

Les filtres particulières (FP) sont largement utilisés dans la communauté Vision pour le suivi d'objets. De nombreuses stratégies sont proposées dans la littérature. En parallèle, la sélection et la paramétrisation des densités de probabilité d'observation, impliquées en tant que fonction de vraisemblance dans le calcul des poids des particules, sont souvent occultées. Ces paramètres ont pourtant un impact important sur les performances finales des algorithmes, particulièrement pour le suivi de mouvement humain, de par le nombre important de degrés de liberté et la présence d'extrema locaux dans l'espace d'état. Dans cet article, nous proposons une étude exhaustive afin de comparer l'impact du choix des mesures et des stratégies de filtrage sur le système de suivi final dans le but de choisir la meilleure association filtre/mesures. La méthode proposée est illustrée dans un contexte de suivi de mouvements humains et les évaluations sont faites par rapport à une vérité terrain fournie par un système commercial. Les observations menées montrent l'importance capitale des fonctions de mesures au regard de la stratégie de filtrage et aboutissent à un ensemble d'heuristiques destinées à l'utilisateur régulier de filtres particuliers afin de résoudre le difficile problème de la paramétrisation des fonctions de vraisemblance.

## Mots Clef

Suivi visuel, filtre particulière, fusion de données, paramétrisation.

## Abstract

Particle filters (PF) are widely used in the Vision literature for visual object tracking. Various schemes are proposed in the literature. Meanwhile, the selection and the tuning of the observation pdf (or likelihood function) involved in the particle weighting stage are often eclipsed. These considerations have a strong influence on the tracking performance, especially for human motion capture (HMC) due to the high number of degrees of freedom and the presence of local extrema in the state space. In this paper, we propose an exhaustive study in order to compare the impact of the

filtering strategy and the chosen measurements on the final system so that the optimal filter/measure association can be achieved. The proposed method is illustrated in the HMC context and the assessments are done w.r.t. a ground truth provided by a commercial HMC system. The presented observations show the major importance of the measurement functions in front of the filtering strategy. These insights lead to some heuristics dedicated to help the practitioner in tackling the difficult problem of the likelihood function tuning.

## Keywords

Visual tracking, particle filtering, data fusion, tuning.

## 1 Introduction

Le suivi visuel est un problème très présent dans la communauté Vision par ordinateur. En particulier, la capture de mouvement humain depuis plusieurs caméras a généré de nombreux travaux, et des approches très variées ont été proposées [14]. Le principe général consiste à inférer la configuration d'un modèle cinématique fruste de l'homme à partir de sa projection dans des séquences monoculaires [19] ou multi-oculaires [4, 8, 20]. Concernant le processus d'estimation, le cadre du filtrage particulière [5], introduit pour la première fois dans un contexte de suivi visuel sous la forme de la CONDENSATION [9], s'avère très adapté au suivi visuel d'objets [4]. L'idée principale est de représenter la distribution *a posteriori* par un ensemble d'échantillons –ou particules– auxquels sont associés des probabilités de sélection –ou poids. Ce nuage de particules est mis à jour récursivement au cours du temps en prenant en compte les données visuelles par le biais du modèle d'observation. Les filtres particuliers n'imposent aucune hypothèse sur les distributions de probabilité manipulées et permettent la fusion de données dans un cadre mathématiquement étayé.

Dès lors, l'idée de fusionner de nombreux indices visuels a été très exploitée dans la communauté Vision [4, 11, 15]. Toutefois, la sélection des indices visuels combinés, de même que la paramétrisation des fonctions de vraisemblances qui leur sont associées, sont très rarement justi-

fiées<sup>1</sup>. Devant cette difficulté de choix (de mesures, de filtres), nous proposons dans cette étude un comparatif entre différentes stratégies de filtrage et analysons l'importance de la conception et de la paramétrisation des fonctions de vraisemblance permettant la prise en compte d'indices visuels variés. Parmi l'ensemble des applications de suivi, la capture de mouvement humain est un domaine particulièrement adapté. En effet, (i) les approches "basées apparence" sont largement utilisées car elles permettent l'exploitation de mesures variées [19, 20], (ii) le réglage des paramètres du système impacte de manière décisive sur les performances du suivi de par la présence de nombreux *extrema* locaux qui peuvent piéger le filtre, (iii) une vérité terrain peut être établie grâce à un système de capture de mouvement commercial.

La section 2 rappelle tout d'abord le formalisme du filtrage particulière et présente plusieurs évolutions de l'algorithme pionnier. La section 3 décrit ensuite quelques mesures visuelles classiquement utilisées et que nous étudierons par la suite. L'ensemble du protocole expérimental mis en œuvre est présenté dans la section 4. La section 5 détaille nos évaluations sur la conception et la paramétrisation des mesures ainsi que sur le choix des stratégies de filtrage. De ces observations, nous tirons des heuristiques et règles générales pour l'implémentation efficace de filtres particulières pour le suivi visuel en section 6. Enfin, nous concluons et exposons quelques perspectives de ces travaux en section 7.

## 2 Filtrage particulière

### 2.1 Présentation

Nous posons le problème du suivi comme l'estimation du vecteur d'état d'un système dynamique stochastique markovien. Le modèle du corps humain utilisé est défini par 9 cônes tronqués et présente 22 degrés de liberté. À chaque instant  $k$ , le vecteur d'état  $\mathbf{x}_k$  à estimer est constitué des paramètres de configuration, et le vecteur de mesures  $\mathbf{z}_k$  est extrait de la projection du modèle du corps humain dans les images trinoculaires. Le système est caractérisé par sa dynamique  $p(\mathbf{x}_k|\mathbf{x}_{k-1})$  –ici une marche aléatoire– et par la densité  $p(\mathbf{z}_k|\mathbf{x}_k)$  des mesures conditionnellement à l'état, qui définit également la vraisemblance de l'état vis-à-vis de la mesure. On suppose connue la distribution de l'état à l'instant initial  $p_0(\mathbf{x}_0)$ . Le principe consiste à propager récursivement dans le temps une densité ponctuelle  $p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$ ,  $\sum_{i=1}^N w_k^{(i)} = 1$  approximant la densité *a posteriori* réelle [5, 9]. L'approximation de l'estimé du minimum d'erreur quadratique moyenne (MMSE)  $\mathbb{E}(\mathbf{x}_k|\mathbf{z}_{1:k}) = \sum_{i=1}^N w_k^{(i)} \mathbf{x}_k^{(i)}$  est alors obtenue immédiatement.

Le célèbre algorithme "Sampling Importance Resampling" (SIR) procède en trois étapes : les particules  $\mathbf{x}_k^{(i)}$ ,  $i = 1..N$  sont propagées selon une fonction d'importance  $q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ , choisie pour explorer les zones pertinentes de l'espace d'état. Ensuite, les poids  $w_k^{(i)}$  sont mis à jour en prenant en compte la fonction de vraisemblance  $p(\mathbf{z}_k|\mathbf{x}_k)$

<sup>1</sup>[12] n'entre pas dans le cadre de notre étude car seule est abordée la robustesse des mesures vis-à-vis du bruit ambiant.

$$\{(\mathbf{x}_k^{(i)}, w_k^{(i)})\}_{i=1}^N = \text{PARTITIONNEQRS}(\{(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)})\}_{i=1}^N, \mathbf{z}_k)$$

- 
- 1: **SI**  $k = 0$ , **ALORS** Échantillonner une séquence QMC randomisée de Sobol  $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(N)}$  selon  $\mathcal{U}_{[0,1]^d}(\mathbf{u})$ , la convertir en  $\mathbf{x}_0^{(1)}, \dots, \mathbf{x}_0^{(N)} \sim p_0(\mathbf{x}_0)$ , et poser  $w_0^{(i)} = \frac{1}{N}$ . **FIN SI**
  - 2: **SI**  $k \geq 1$  **ALORS**  $\{(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)})\}_{i=1}^N$  représente  $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$
  - 3: Poser  $\tau_0^{(i)} = w_{k-1}^{(i)}$  et  $\mathbf{x}_k^{0,(i)} = \mathbf{x}_{k-1}^{(i)}$  pour  $i = 1, \dots, N$
  - 4: **POUR**  $m = 1, \dots, M$ , **FAIRE**
  - 5: Sélectionner avec remise  $s^{(1)}, \dots, s^{(N)}$  dans  $\{1, \dots, N\}$  tels que  $P(s^{(i)} = j) = \frac{\tau_{m-1}^{(j)}}{\sum_{i=1}^N \tau_{m-1}^{(i)}}$
  - 6: Poser  $C_j = \text{card}(\{i|s^{(i)} = j\})$
  - 7: **POUR**  $j = 1, \dots, N$ , **FAIRE**
  - 8: Échantillonner une séquence QMC randomisée de Sobol  $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(C_j)}$  selon  $\mathcal{U}_{[0,1]^d}(\mathbf{u})$  et la convertir en  $\mathbf{x}_k^{m,(\sum_{i=1}^{j-1} C_i+1)}, \dots, \mathbf{x}_k^{m,(\sum_{i=1}^j C_i+C_j)} \sim p_m(\mathbf{x}_k^m|\mathbf{x}_k^{m-1,(j)})$
  - 9: **FIN POUR**
  - 10: Mettre à jour les poids via  $\tau_m^{(i)} \propto l_m(\mathbf{z}_k|\mathbf{x}_k^{m,(i)})$
  - 11: Normaliser les poids de sorte que  $\sum_i \tau_m^{(i)} = 1$
  - 12: **FIN POUR**
  - 13: Poser  $w_k^{(i)} = \tau_m^{(i)}$  et  $\mathbf{x}_k^{(i)} = \mathbf{x}_k^{m,(i)}$  pour  $i = 1, \dots, N$
  - 14: Calculer le MMSE  $\mathbb{E}_{p(\mathbf{x}_k|\mathbf{z}_{1:k})}[\mathbf{x}_k] = \sum_{i=1}^N w_k^{(i)} \mathbf{x}_k^{(i)}$
  - 15: **FIN SI**
- 

TAB. 1 – Algorithme PARTITIONNE QRS.

pour maintenir la cohérence de l'approximation ponctuelle. Enfin, lorsque la représentation a tendance à dégénérer, une étape de rééchantillonnage est appliquée. Notons enfin que le cadre théorique du SIR englobe l'algorithme pionnier CONDENSATION [9]. Ci-après, nous présentons quelques extensions de cet algorithme fréquemment retrouvées dans la littérature.

### 2.2 Filtre particulière PARTITIONNE

Contrairement à une croyance répandue, les temps de calcul affichés par un filtre particulière pour des problèmes classiques, bien que linéaires en le nombre de particules, sont exponentiels en l'ordre du système pour une erreur indépendante de la dimension [3]. Afin de réduire cette complexité, plusieurs algorithmes ont été proposés. Lorsque la dynamique du système peut s'écrire comme la composition de  $M$  évolutions partielles  $p_m(\mathbf{x}_k^m|\mathbf{x}_k^{m-1})$  du vecteur d'état  $\mathbf{x}_k^m$  à l'étape  $m$ , et que des vraisemblances intermédiaires  $l_m(\mathbf{x}_k^m|\mathbf{z}_k)$ ,  $m = 1..M$ , peuvent être évaluées après chacune de ces dynamiques partielles, on peut exploiter le filtre PARTITIONNE [13]. Grâce à ce partitionnement de l'espace d'état, la complexité devient alors linéaire en le nombre de partitions.

### 2.3 Techniques de Quasi Monte Carlo

L'échantillonnage purement stochastique mène généralement au problème « d'amas et de trous », ce qui dégrade la qualité de l'estimation fournie par le filtre, particulièrement dans les espaces de grande dimension. Remplacer l'échantillonnage purement aléatoire par des séquences à faible discrédance (déterministes ou randomisées) —dites

de « Quasi Monte Carlo (QMC) »— permet d’améliorer la convergence du filtre pour le même nombre de particules  $N$  [6], en réduisant la racine de l’erreur quadratique moyenne (RMSE) tout en améliorant la dispersion de l’estimé de 5% à 20% [17, 16].

Parmi les problèmes principaux liés à l’utilisation des filtres QMC, nous pouvons noter la difficulté de la mise en œuvre des séquences à faible discrétion malgré les étapes de rééchantillonnage, l’exploitation de la mesure courante au sein de ces séquences, et le compromis éventuel entre la réduction de la complexité quadratique et la rigueur mathématique des algorithmes. Une alternative QMC à la CONDENSATION, dénommé QRS (pour Quasi Random Sampling), est proposé dans [7]. Nous avons adapté cette idée au filtre PARTITIONNE proposé par [13]. L’algorithme final est décrit en Table 1. L’idée principale est ici de regrouper les étapes de propagation selon la dynamique et de rééchantillonnage. Cela permet de générer une séquence à faible discrétion en appliquant la dynamique à une particule à rééchantillonner, ce qui résulte en une exploration plus régulière de l’espace des configurations.

### 3 Fonctions de vraisemblance

L’implémentation de notre système repose sur l’utilisation des quatre indices visuels suivants. Leurs vraisemblances associées sont de la forme :

$$p(\mathbf{z}_k^{c,i} | \mathbf{x}_k) \propto \exp\left(-\frac{D_{c,i}^2}{2\sigma_i^2}\right), \quad (1)$$

où  $D_{c,i}$  désigne la distance de similarité entre l’hypothèse de configuration  $\mathbf{x}_k$  et l’image  $\mathbf{z}_k$  pour l’indice visuel  $i$  et la caméra  $c$  considérés, et  $\sigma_i$  en est l’écart-type *a priori*, dont la valeur est sélectionnée à la discrétion de l’utilisateur. On retrouve ce modèle dans de nombreux travaux [4, 13]. Ci-après sont présentées quelques-unes des distances de similarité associées aux mesures les plus classiquement utilisées en suivi visuel, et sur lesquelles nous mènerons notre étude :

**Silhouette** ( $D_{sil}$ ) - Dans la veine de [4], nous mettons en place une segmentation de la silhouette dans l’image courante par soustraction de fond afin d’obtenir un masque. Plusieurs points sont ensuite échantillonnés à l’intérieur des limbes du modèle projeté sous l’hypothèse de chaque particule. Il s’agit alors de caractériser la cohérence des limbes relatives à chaque particule avec la silhouette segmentée.

**Silhouette duale** ( $D_{sil2}$ ) - Dans le but de limiter l’effet du premier indice visuel, nous exploitons également l’idée proposée dans [19]. Le principe est symétrique, consistant à échantillonner la silhouette segmentée et à vérifier sa cohérence vis-à-vis de la projection de la particule courante.

**Probabilité de couleur peau** ( $D_{peau}$ ) - Afin d’améliorer la précision de la localisation, particulièrement pour les membres fins tels que les bras, une mesure additionnelle est utilisée, reposant sur une segmentation de la couleur peau. Une image de probabilité est calculée par rétroprojection d’un histogramme de couleur peau préalablement appris [10, 13]. Pour chaque hypothèse  $\mathbf{x}_k$ , nous calculons

la probabilité de couleur peau moyenne sur 3 points virtuels situés sur la tête et les mains du modèle sous l’hypothèse courante. Une probabilité moyenne de couleur peau élevée résulte en une distance de similarité faible.

**Distance aux blobs de couleur peau** ( $D_{peau\_dist}$ ) - Nous proposons également une variante de l’indice précédent. Des blobs sont extraits de l’image de probabilité de couleur peau précédemment construite. Pour chaque particule, nous calculons la distance entre la tête et les mains et le plus proche blob de couleur peau détecté. Les meilleures configurations sont celles qui présentent les plus faibles distances.

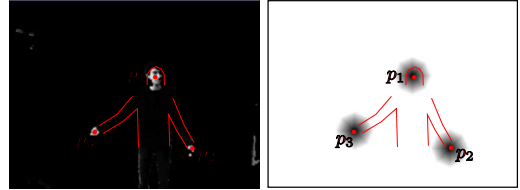


FIG. 1 – [gauche] : image de probabilité de couleur peau (utilisée pour  $D_{peau}$ ); [droite] : image de distance aux blobs de couleur peau (utilisée pour  $D_{peau\_dist}$ ). Le modèle est projeté sur chacune des images.

Les valeurs des distances de similarité peuvent être significativement différentes les unes des autres. Ainsi, dans notre implémentation,  $D_{sil} \in [0, 255]$ ,  $D_{sil2} \in [0, 1]$ ,  $D_{peau} \in [0, 255]$ ,  $D_{peau\_dist} \in [0, 50]$ . Les quatre vraisemblances ont une forme similaire à (1), avec  $\sigma_{sil}$ ,  $\sigma_{sil2}$ ,  $\sigma_{peau}$ ,  $\sigma_{peau\_dist}$  les écarts-types correspondants. Sous l’hypothèse de leur indépendance conditionnellement à l’état, la vraisemblance unifiée s’écrit alors :

$$p(\mathbf{z}_k | \mathbf{x}_k) = \prod_{c=1}^C \prod_i p(\mathbf{z}_k^{c,i} | \mathbf{x}_k),$$

avec  $i \in \{sil, sil2, peau, peau\_dist\}$ . La figure 2 illustre l’influence des paramètres libres  $\sigma_i$ . Selon leur valeur, nous pouvons observer un échec du suivi du bras gauche du sujet capturé.

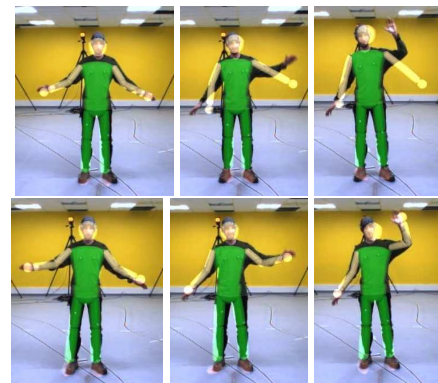


FIG. 2 – Suivi sur une séquence simple [haut] :  $(\sigma_{sil}, \sigma_{sil2}, \sigma_{peau}, \sigma_{peau\_dist}) = (130, 0.5, 20, 30)$ ; [bas] :  $(\sigma_{sil}, \sigma_{sil2}, \sigma_{peau}, \sigma_{peau\_dist}) = (20, 0.1, 60, 5)$ . L’avatar en surimpression représente la configuration estimée *via* le MMSE.

## 4 Protocole expérimental

### 4.1 Configuration expérimentale

La configuration expérimentale de notre étude met en oeuvre trois <sup>2</sup> caméras couleur IEEE 1394 “progressive scan” Flea 2 qui fournissent des images de taille  $640 \times 480$ . La position réelle des articulations du sujet suivi est donnée par un système de capture commercial à marqueurs de Motion Analysis [22]. La figure 3 présente la situation des deux systèmes (commercial et visuel). Ceux-ci sont calibrés et synchronisés logiciellement. Afin de dresser une étude stochastique du filtre, 30 exécutions sont réalisées sur chaque séquence d’images. Le comportement est évalué sur plusieurs séquences d’environ 20 s proposant des mouvements de marche et de gymnastique.

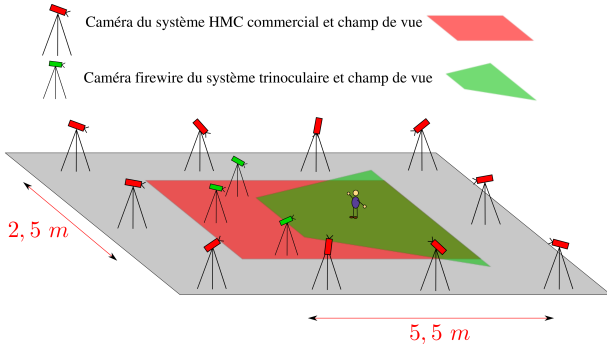


FIG. 3 – Configuration des deux systèmes de capture de mouvement humain.

### 4.2 Définition des métriques pour les évaluations

Quatre métriques sont définies, impliquant l’estimé du MMSE  $\widehat{\mathbf{x}}_{k,r}$  fourni à l’instant  $k$  par chaque exécution  $r \in 1..R$  du filtre et la vérité terrain  $\mathbf{t}_k$  issue du système commercial. Elles reposent sur la comparaison des positions 3D estimées  $\mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}}$  des articulations  $j \in 1..J$  et leurs positions réelles  $\mathbf{m}_j^{\mathbf{t}_k}$ . En effet, comparer des n-uplets d’angles ou des paramétrisations d’orientations constitue un mauvais choix car des valeurs très différentes peuvent correspondre à des poses du modèle humain quasiment identiques. Ainsi nous définissons :

**Racine de l’erreur quadratique moyenne par rapport à la vérité terrain (RMSE)**- Le premier critère est la moyenne sur toutes les articulations de la racine de l’erreur quadratique moyenne entre la position estimée et la position réelle calculée pour toutes les réalisations du filtres et pour toutes les images :

$$\frac{1}{J} \sum_{j=1}^J \sqrt{\frac{1}{K} \sum_{k=1}^K \frac{1}{R} \sum_{r=1}^R \|\mathbf{m}_j^{\mathbf{t}_k} - \mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}}\|_2^2}. \quad (2)$$

Cette métrique, nommée “RMSE” (pour Root Mean Square Error) est très classique dans la littérature [2, 20,

<sup>2</sup> [1] a montré que trois caméras étaient suffisantes pour proposer un suivi correct.

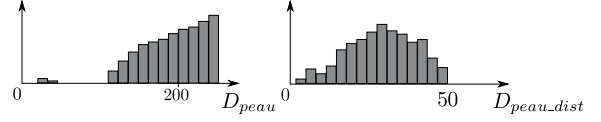


FIG. 5 – Histogrammes des distances de similarités pour un nuage de particules donné : [gauche]  $D_{peau}$  ; [droite]  $D_{peau\_dist}$ .

21].

**Biais** - Cette métrique permet de vérifier que plusieurs réalisations du filtre fournissent un estimé moyen centré sur la vérité terrain. Curieusement, ce critère est peu présent dans la littérature :

$$\frac{1}{J} \sum_{j=1}^J \left\| \frac{1}{K} \sum_{k=1}^K (\mathbf{m}_j^{\mathbf{t}_k} - \frac{1}{R} \sum_{r=1}^R \mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}}) \right\|_2. \quad (3)$$

**Dispersion de l’estimateur** - Même si ce critère est également peu utilisé [18], la variance du système de suivi doit être analysée sur l’ensemble des réalisations afin de s’assurer que, pour une séquence donnée, les estimés fournis sont suffisamment regroupés en dépit de leur nature stochastique :

$$\frac{1}{J} \sum_{j=1}^J \sqrt{\frac{1}{K} \sum_{k=1}^K \frac{1}{R} \sum_{r=1}^R \|\mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}} - \frac{1}{R} \sum_{r=1}^R \mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}}\|_2^2}. \quad (4)$$

**Taux d’échec du suivi** - Afin de compléter l’évaluation, une dernière métrique se focalise sur le taux d’échec. On considère que le système est mis en échec dès lors qu’une articulation présente une erreur supérieure à un certain seuil :

$$\frac{1}{R} \frac{1}{J} \frac{1}{K} \sum_{r=1}^R \sum_{j=1}^J \sum_{k=1}^K \text{fails}(\mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}}), \quad (5)$$

où  $\text{fails}(\mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}})$  vaut 1 si  $\|\mathbf{m}_j^{\mathbf{t}_k} - \mathbf{m}_j^{\widehat{\mathbf{x}}_{k,r}}\|_2 > T_{\text{Failure}}$ , 0 sinon.

## 5 Évaluations et résultats

La définition et la combinaison des mesures peuvent avoir des conséquences importantes sur le comportement du filtre. C’est pourquoi nous évaluons ci-après l’effet du paramètre  $\sigma_i$  dans l’équation (1). Pour ce faire, nous nous focalisons –pour l’instant– sur un filtre de type CONDENSATION (qui constitue la base des stratégies plus avancées) avec  $N = 1000$  particules.

### 5.1 Définition des mesures

La figure 4 présente le RMSE de l’estimé pour nos deux mesures  $peau$  et  $peau\_dist$ , très proches dans leur définition. Les autres métriques ne sont pas présentées ici, mais affichent des caractéristiques similaires : l’indice  $peau$  conduit à des résultats catastrophiques ( $\text{RMSE} \sim 1m$ ) comparativement à l’indice  $peau\_dist$ , quel que soit le paramètre  $\sigma_i$  choisi. Ceci est dû aux modes très prononcés de la distance de similarité  $D_{peau}$ . En effet, les régions de

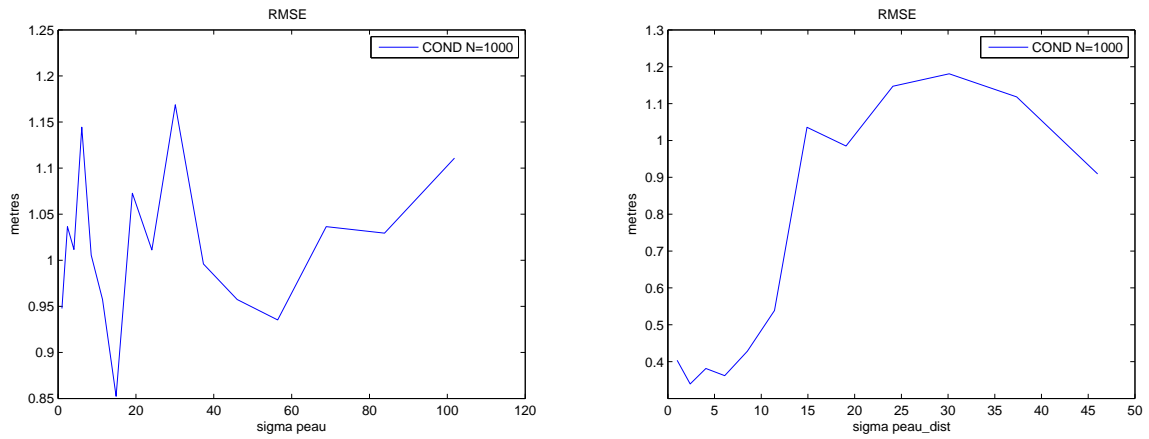


FIG. 4 – Influence du paramètre  $\sigma$  sur le RMSE pour 30 réalisations de l’algorithme CONDENSATION lorsque l’indice *peau* est utilisé seul (gauche) et lorsque l’indice *peau\_dist* est utilisé seul (droite). L’indice *peau* est intrinsèquement mal défini et n’apporte aucune information.

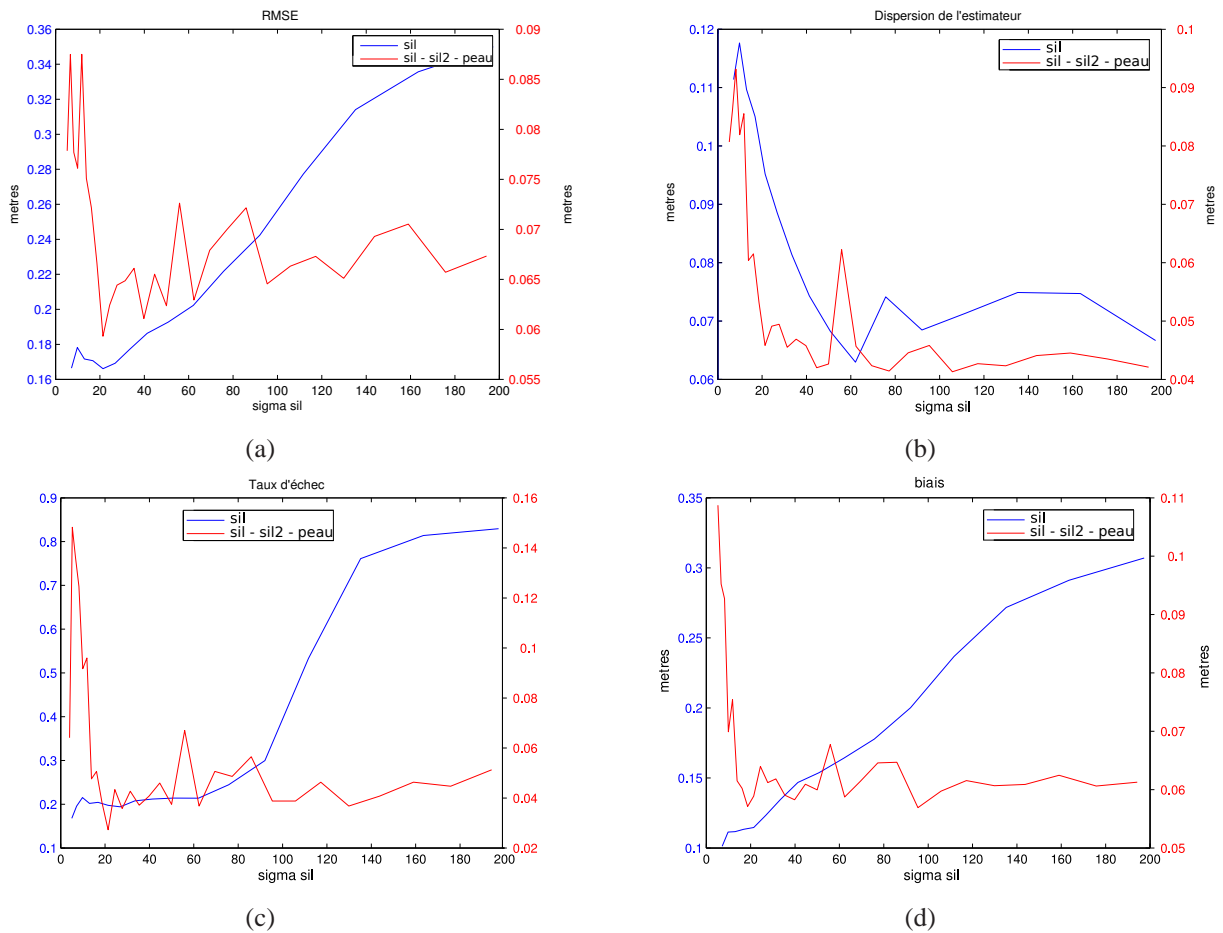


FIG. 6 – Influence de  $\sigma_{sil}$  pour 30 réalisations de l’algorithme CONDENSATION selon que l’indice *sil* est utilisé seul (bleu) ou conjointement avec les indices *sil2* et *peau\_dist* (rouge) : (a) RMSE, (b) dispersion de l’estimateur, (c) taux d’échec, (d) biais.

haute probabilité de couleur peau sont généralement de petite taille, et la tête et les mains de chaque particule ont peu de chance d’être situées exactement dans ces zones (distance de similarité faible), comme on peut le voir sur la figure 5 - gauche. Ainsi, leur poids est systématiquement écrasé. Parallèlement, l’indice *peau\_dist* présente un profil plus doux de par le calcul de la distance aux blocs de couleur peau, ce qui favorise une distribution plus régulière des distances de similarité (cf. Figure 5 - droite). Ceci montre à quel point la définition des mesures impliquées dans la fonction de vraisemblance est capitale. En outre, il convient également de bien les paramétrer *via* le choix du paramètre  $\sigma_i$  qui leur est associé, ce que nous présentons dans la partie suivante.

## 5.2 Paramétrisation des mesures

L’indice *sil* étant le plus utilisé dans la littérature, nous nous intéressons en premier lieu au comportement du système lorsque seul cet indice est exploité dans la fonction de vraisemblance globale. La figure 6 (courbe bleue) présente l’influence du paramètre  $\sigma_{sil}$  sur une séquence simple avec  $N = 1000$  particules. Tout d’abord, nous remarquons que la valeur optimale de  $\sigma_{sil}$  dépend du critère que l’on souhaite minimiser. Le RMSE indique  $\sigma_{sil} = 22$ , la dispersion est minimale pour  $\sigma_{sil} = 200$ , et le biais optimal est atteint pour  $\sigma_{sil} = 1$ . En deçà de cette valeur, le filtre diverge à cause de problèmes numériques (calculs trop proches de 0). De manière intuitive, une grande valeur de  $\sigma_{sil}$  résulte en un estimé moins dispersé d’une réalisation sur l’autre car l’indice considéré revêt une importance moindre au regard de la dynamique. Visuellement, cela se traduit par une évolution plus douce d’un instant à l’autre. Le RMSE et le biais sont plus faibles pour une valeur de  $\sigma_{sil}$  plus faible également, toutefois, le biais et le RMSE optimaux sont respectivement de  $0.1 m$  et de  $0.17 m$  par articulation, ce qui révèle que l’indice *sil* seul n’est pas suffisant pour permettre un suivi satisfaisant avec 1000 particules. Manifestement, un compromis doit être fait entre précision et dispersion.

Afin d’obtenir un certain équilibre, nous choisissons  $\sigma_{sil} = 30$ . Nous procédons de même en insérant dans la fonction de vraisemblance unifiée les indices *sil2* et *peau\_dist*, ce qui mène à  $\sigma_{sil2} = 0.07$  et enfin  $\sigma_{peau\_dist} = 5$ . Ce dernier indice permet un suivi satisfaisant et abaisse le RMSE à  $0.06 m$ . Cependant, nous pouvons nous demander si les indices *sil* et *sil2* ne sont pas redondants. La figure 6 (courbe rouge) présente les effets de  $\sigma_{sil}$  alors que les indices *sil*, *sil2* et *peau\_dist* sont exploités conjointement. Le comportement du filtre pour les grandes valeurs de  $\sigma_{sil}$  nous informe sur la pertinence de l’indice *sil*, puisque le système se comporte alors comme si cette mesure était absente. Ainsi, nous remarquons que l’indice *sil* améliore sensiblement le RMSE, mais aucun des autres critères. Ceci amène deux conclusions : tout d’abord, certains indices semblent être superflus lorsqu’ils sont combinés avec d’autres, et ne font qu’ajouter une charge calculatoire inutile sans améliorer les résultats globaux. Ensuite, le comportement observé ici témoigne d’une corrélation entre les mesures. Ceci semble suggérer que la condition d’indépendance des mesures conditionnellement

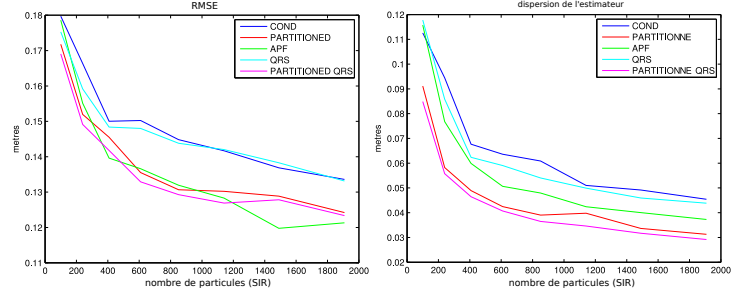


FIG. 7 – [gauche] : RMSE ; [droite] dispersion de l’estimateur, pour différents filtres.

Nom	RMSE	Dispersion	Taux d’échec	Biais
CONDENSATION	5	5	5	5
QRS	4	4	4	4
PARTITIONNE	3	2	3	3
<b>PARTITIONNE QRS</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>2</b>
APF	2	3	2	1

TAB. 2 – Classement des différents FP selon chaque critère.

à l’état est partiellement inappropriée. Enfin, on peut noter que l’ordre de grandeur identique entre le RMSE et le biais ( $\sim 0.06 m$ ) traduit le fait que notre système ne pourra pas arriver à un meilleur résultat en l’état. Nous atteignons ici la limite de nos modèles (simples) d’observation. Pour de meilleurs résultats il conviendrait de mettre en place des mesures plus fines.

## 5.3 Stratégies de filtrage

Dans cette partie, nous présentons une comparaison entre notre filtre PARTITIONNE QRS et les stratégies CONDENSATION [9], QRS [7], APF [4] et PARTITIONNE [13], qui constituent les principales méthodes de l’état de l’art. Toutes les stratégies sont normalisées par rapport au nombre d’évaluations de la fonction de vraisemblance, qui est –de loin– la plus coûteuse en temps de calcul. L’« Annealed Particle Filter » (APF) [4] est utilisé avec 3 couches pour être suffisamment efficace dans notre contexte et est paramétré selon les recommandations de [4]. Les filtres PARTITIONNE exploitent 2 partitions : une pour la localisation et l’orientation du torse, l’autre pour la configuration des membres. Les filtres exploitent les mesures *sil* et *peau\_dist* avec les paramètres  $\sigma_{sil} = 30, \sigma_{peau\_dist} = 5$  précédemment justifiés. Un condensé des résultats est présenté en table 2.

**RMSE** - La figure 7 (a) présente le RMSE. Tous les filtres sont globalement satisfaisants. Les erreurs sont raisonnables compte tenu de nos modèles frustes de l’homme et des mesures. Pour ce qui est des performances relatives, nous remarquons que les stratégies avancées fournissent un meilleur résultat que la CONDENSATION classique. Toutefois l’APF semble moins efficace que le PARTITIONNE QRS pour un faible nombre de particules. Selon [4], l’APF doit disposer d’un nombre minimum de particules pour être opérationnel. De manière globale, les techniques QMC proposent un meilleur estimé que leurs alternatives MC. Pour une erreur donnée, elles peuvent nécessiter 20 % de particules en moins dans le meilleur des cas, ce qui consti-

tue un gain important en temps de calcul.

**Dispersion de l'estimateur**- La figure 7 (b) présente la dispersion de l'estimateur. Les filtres PARTITIONNE affichent une dispersion plus faible. Les versions QMC semblent proposer un estimé plus « stable » que les techniques MC de par leur échantillonnage à faible discrétance, même pour nos versions simplifiées afin de garantir une complexité en  $\mathcal{O}(N)$ . Ce critère peu exploité dans la littérature révèle que la stratégie PARTITIONNE QRS propose la meilleure dispersion, y compris pour un grand nombre de particules. Elle peut améliorer la dispersion de 3 cm par articulation, ce qui constitue une progression significative.

**Taux d'échecs et biais** - Ces deux métriques ne sont pas présentées par manque de place. Le taux d'échec est cohérent avec le RMSE et la dispersion de l'estimé au regard des performances relatives identiques. Le biais est meilleur pour l'APF pour  $N > 1000$  dans notre contexte. En deçà de ce seuil, les filtres PARTITIONNE, PARTITIONNE QRS et APF fournissent des résultats similaires.

Malgré les résultats satisfaisants de notre filtre, nous pouvons remarquer que les différences par rapports aux techniques de la littérature restent relativement faible au regard des améliorations qu'ont apportées nos études sur le choix des mesures et leur paramétrisation.

## 6 Discussion et heuristiques générales

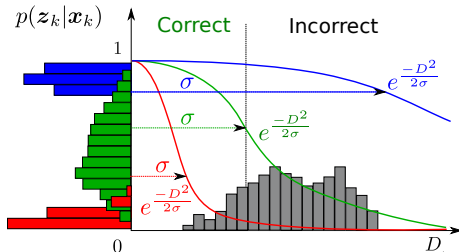


FIG. 8 – Histogrammes des poids des particules en fonction de l'histogramme des distances de similarité pour un nuage de particules distribué autour de la vérité terrain, et différents choix de  $\sigma$ .

De nos évaluations, nous pouvons dériver plusieurs observations concernant la paramétrisation des vraisemblances : (1) un compromis doit être établi entre précision et dispersion, (2) une faible valeur de  $\sigma_{(\cdot)}$  entraîne un biais et une erreur plus faibles donc une meilleure précision, (3) trop de mesures et/ou caméras impliquées ou de trop faibles valeurs de  $\sigma_{(\cdot)}$  entraînent une dispersion des estimés plus importante, et, dans le pire des cas, une divergence du filtre due à des problèmes numériques. Ainsi, nous proposons les heuristiques suivantes pour guider le concepteur de filtres particuliers à paramétrer au mieux ses fonctions de vraisemblance :

**Sélection d'une distance de similarité "douce"** - En premier lieu, il convient de privilégier une distance de similarité présentant une courbure douce autour des *optima* afin de permettre une distribution équilibrée de ses valeurs pour

un nuage de particules donné (histogramme gris sur la figure 8). Ceci est d'autant plus important que l'espace d'état est de grande dimension. Nous pouvons ici noter qu'une distance de similarité nulle (*i.e.* une correspondance parfaite entre une particule et les données visuelles) ne se produit jamais en pratique, de par l'inexactitude de nos modèles (de l'homme, de l'environnement, du lien état-mesures, des bruits ambiants, ...).

**Paramétrisation préliminaire des  $\sigma$**  - L'évolution de la distance de similarité par rapport au vecteur d'état peut constituer une bonne ligne de conduite pour la paramétrisation des  $\sigma$ . La figure 8 présente l'histogramme des distances de similarité (en gris) pour un nuage de particules distribué autour de la vérité terrain. La paramétrisation préliminaire des  $\sigma$  doit se situer à l'intérieur du support de cet histogramme. De manière simplifiée, les particules présentant une distance de similarité plus petite que la valeur de  $\sigma$  (en vert) seront considérées comme "bonnes" par rapport à l'indice concerné, tandis que les autres seront considérées comme "mauvaises" et seront affectées d'un poids plus faible. Ceci peut être justifié par le fait que le point de plus grande pente d'une gaussienne  $\mathcal{N}(x; \mu, \sigma^2)$  est atteint pour  $x = \sigma$ .

**Affinage des valeurs choisies** - Nous pouvons améliorer le choix initial par l'observation : si l'on augmente le paramètre  $\sigma$  (courbe bleue sur la figure 8), tous les poids sont proches de 1 et l'indice considéré apporte moins d'information. Ainsi, il est moins pris en compte dans la densité filtrée, ce qui réduit la dispersion de l'estimé. Si l'on diminue la valeur de  $\sigma$ , la fonction de vraisemblance présente un pic plus important, ce qui résulte en une sélection plus drastique de la particule la plus adaptée au détriment des autres (courbe rouge sur la figure 8). Dans un même temps, l'estimateur présentera une dispersion plus importante sur plusieurs réalisations. En outre, la valeur des poids calculés diminue, ce qui peut mener au problème suivant.

**Limites calculatoires** - La valeur de  $\sigma$  ne peut pas être réduite à volonté : en effet, une trop faible valeur peut conduire à une fonction de vraisemblance nulle ("underflow"). Par exemple,  $\exp(-\frac{1}{2} \frac{0.5^2}{0.01^2}) = 0$ , de telle sorte que choisir  $\sigma = 0.01$  résulte en un poids nul pour une distance de similarité de 0.5, et mène à une divergence du filtre. Ce phénomène est accentué lorsque de multiples indices ou caméras sont utilisés, car les fonctions de vraisemblance sont multipliées entre elles. Par conséquent, il convient d'être très prudent lorsque l'on souhaite diminuer les valeurs de  $\sigma$  et fusionner plusieurs mesures.

## 7 Conclusions et perspectives

Nous avons proposé une étude portant sur différentes stratégies de filtrage ainsi que sur la conception et la paramétrisation des fonctions de vraisemblance dans un contexte de suivi visuel de mouvement humain au regard de quatre métriques. Il apparaît que la fusion de mesures n'est pas triviale et que la paramétrisation des vraisemblances a une forte influence sur les performances du filtre. Des compromis doivent être faits entre la précision du suivi et la dispersion de l'estimé, ce qui –à notre connaissance– n'est pas mentionné dans la littérature. L'ajout de mesures peut

également impacter de manière significative sur la dispersion du filtre et peut ne pas améliorer son comportement tout en ajoutant une charge calculatoire inutile. À partir de ces observations, nous avons proposé quelques heuristiques simples afin d'estimer la valeur des paramètres libres (écarts-types  $\sigma$ ) et de guider ces choix, qui sont systématiquement inexplicables dans la littérature. Dans notre contexte multi-oculaire, et avec les mesures décrites ici, la stratégie PARTITIONNE QRS que nous proposons affiche de meilleurs résultats que les stratégies couramment trouvées dans la littérature. Toutefois, il semble que la mise en place d'une stratégie « optimale » soit conditionnée au choix et à la bonne exploitation des mesures.

Une perspective de ces travaux pourraient être de proposer une étude plus complète, faisant intervenir des mesures plus hétérogènes, en particulier basées sur des capteurs 3D de type "temps-de-vol" (Swissranger). En complément, il serait également intéressant de prendre en compte les détecteurs visuels et les méthodes par échantillonnage d'importance qui permettent l'initialisation automatique.

## Remerciements

Ces travaux ont été partiellement menés au sein du projet ANR AMORCES.

## Références

- [1] A. Balan, M. J. Black, H. Haussecker, and Sigal L. Shining a light on human pose : on shadows, shading and the estimation of pose and shape. In *International Conference on Computer Vision (ICCV'07)*, Rio de Janeiro, BRAZIL, October 2007.
- [2] A. Balan, L. Sigal, and M. Black. A quantitative evaluation of video-based 3D person tracking. In *International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS'05)*, pages 349–356, Washington, USA, October 2005.
- [3] F. Daum and J. Huang. Mysterious computational complexity of particle filters. In *Signal and Data Processing of Small Targets*, volume 4728 of *Proceedings of SPIE*, Bellingham, MA, USA, August 2003.
- [4] J. Deutscher and I. Reid. Articulated body motion capture by stochastic search. *International Journal of Computer Vision (IJCV'05)*, 21(3) :185–205, 2005.
- [5] A. Doucet, N. De Freitas, and N. J. Gordon. *Sequential Monte Carlo Methods in Practice*. Series Statistics For Engineering and Information Science. Springer-Verlag, New York, 2001.
- [6] K.-T. Fang, Y. Wang, and P. M. Bentler. Some applications of number-theoretic methods in statistics. *Statistical Science*, 9(3) :416–428, 1994.
- [7] D. Guo and X. Wang. Quasi-Monte Carlo filtering in nonlinear dynamic systems. *IEEE transactions on signal processing*, 54(6) :2087–2098, June 2006.
- [8] A. Gupta, A. Mittal, and L.S. Davis. Constraint integration for efficient multiview pose estimation of humans with self-occlusions. *Transactions on Pattern Analysis Machine Intelligence (PAMI'08)*, 30(3) :493–506, March 2008.
- [9] M. Isard and A. Blake. CONDENSATION – Conditional density propagation for visual tracking. *International Journal on Computer Vision (IJCV'98)*, 29(1) :5–28, 1998.
- [10] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1) :81–96, 2002.
- [11] O. Lanz and R. Brunelli. An appearance-based particle filter for visual tracking in smart rooms. In *International Evaluation Workshop CLEAR'07*, pages 57–69, May 2007.
- [12] J. Lichtenauer, Reinders M. J. T., and Hendriks E. A. Influence of the observation likelihood function on particle filtering performance in tracking applications. In *Automatic Face and Gesture Recognition (FGR'04)*, pages 767–772, Seoul, KOREA, May 2004.
- [13] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *European Conference on Computer Vision (ECCV'00)*, pages 3–19, Dublin, Ireland, 2000.
- [14] T. Moeslund, A. Hilton, and V. Krüger. A survey of advanced vision-based human motion capture and analysis. *Computer Vision and Image Understanding (CVIU'06)*, 104 :174–192, December 2006.
- [15] H. Moon and R. Chellappa. 3D shape-encoded particle filter for object tracking and its application to human body tracking. *EURASIP Journal on Image and Video Processing*, 2008, 2008.
- [16] D. Ormoneit, C. Lemieux, and D.J. Fleet. Lattice particle filters. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence (UAI'01)*, pages 395–402, San Francisco, CA, USA, 2001.
- [17] V. Philomin, R. Duraiswami, and L. S. Davis. Quasi-random sampling for CONDENSATION. In *European Conference on Computer Vision (ECCV'00)*, pages 134–149, Dublin, Ireland, 2000.
- [18] L. Sigal and M. J. Black. Humaneva : Synchronized video and motion capture dataset for evaluation of articulated human motion. Technical report, Department of Computer Science Brown University, 2006.
- [19] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling. *International Journal on Robotic Research (IJRR'03)*, 6(22) :371–393, May 2003.
- [20] P. Wang and J. Rehg. A modular approach to the analysis and evaluation of particle filters for figure tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pages 790–797, New York, USA, 2006.
- [21] <http://vision.cs.brown.edu/humaneva/> — humaneva.
- [22] <http://www.motionanalysis.com> — Motion Analysis Corporation.