Suivi 3D Monoculaire du Haut du Corps par une Propagation des Croyances sous Contraintes Articulaires

Multicues 3D Monocular Upper Body Tracking Using Constrained Belief Propagation

Philippe Noriega Olivier Bernier

France Télécom Recherche et Développement
2, Av Pierre Marzin, 22300 Lannion, France

{philippe.noriega, olivier.bernier}@orange-ftgroup.com

Résumé

Cet article décrit une méthode destinée au suivi du haut du corps en 3D pour des scènes filmées avec une caméra monoculaire. La structure articulaire est représentée par un modèle graphique probabiliste, à savoir un graphe de facteurs dans lequel la propagation des croyances permet d'évaluer la probabilité marginale de chacun des membres. Le modèle du corps est un modèle à membres indépendants incluant des facteurs d'attraction entre les membres adjacents et de répulsion pour empêcher les collisions. Pour résoudre les ambiguïtés résultant de la vision monoculaire, un ensemble d'indices robustes basés sur les contours et la couleur sont extraits de l'image et des contraintes articulaires sont intégrées au modèle. Les résultats tant qualitatifs que quantitatifs viennent confirmer l'efficacité de l'algorithme proposé.

Mots Clef

Containtes articulaires, filtre à particules, propagation des croyances, suivi du haut du corps en monoculaire.

Abstract

This paper describes a method for articulated 3D upper body tracking in monocular scenes using a graphical model to represent an articulated body structure. Belief propagation on factor graphs is used to compute the marginal probabilities of limbs. The body model is a loose-limbed model including attraction factors between adjacent limbs and constraints to reject poses resulting in collisions. To solve ambiguities resulting from monocular view, robust contour and colour based cues are extracted from the images. Moreover, a set of constraints on the model articulations is implemented according to human pose capabilities. Quantitative and qualitative results illustrate the efficiency of the proposed algorithm.

Keywords

Articulated constraints, particle filter, belief propagation, monocular upper body tracking.

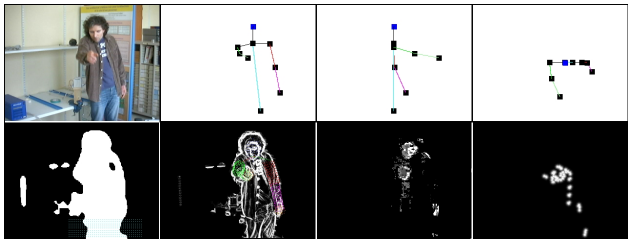

FIG. 1 – Suivi du haut du corps en monoculaire. Première ligne : image originale et les vues de face, de droite et de haut des poses estimées. Deuxième ligne : soustraction du fond, contours, carte de la couleur du visage et carte de l'énergie du mouvement.

1 Introduction

Les algorithmes pour le suivi du corps doivent s'accommoder d'espaces de grande dimension où la probabilité jointe est hautement multimodale et bruitée. Dans ce contexte, certaines méthodes déterministes peuvent effectuer un suivi en temps réel grâce à une caméra stéréo [7], mais elles peuvent êtres mises en échec en monoculaire du fait des nombreux optimums locaux et des ambiguïtés engendrées par la vision monoculaire [16].
Les contraintes articulaires limitent l'espace réel des poses à un sous-espace bien plus petit que l'espace théorique qui le contient. De ce fait, les méthodes de suivi basées sur l'apprentissage peuvent êtres efficaces à condition qu'elles parviennent à couvrir suffisamment ce sous-espace. Parmi elles, certaines utilisent une régression pour déduire la pose à partir des observation extraites de l'image dans le but de suivre des personnes marchant dans un environnement

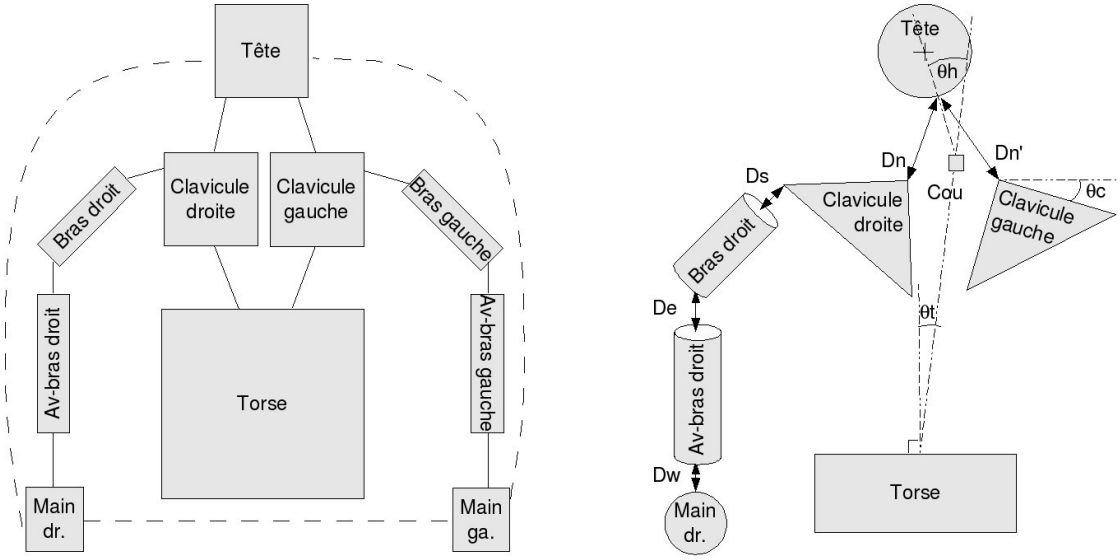

FIG. 2 – Interactions entre les membres (figure de gauche) : les noeuds correspondent aux membres, les contraintes articulaires sont représentées par les traits pleins et les pointillés représentent des contraintes de non-collision entre la tête et les mains. Modèle du haut du corps (figure de droite) : les bras et les avant-bras sont modélisés par des cylindres et la tête par une sphère. Les autres membres (mains, torse et clavicules) sont représentés par des éléments 2D. Les interactions entre les membres sont calculées à partir des distances (Dn, Ds, De et Dw) qui les séparent. Les autres contraintes articulaires sont déduites des angles θh, θc et θt. La base du cou se trouve à égale distance des deux clavicules.

contraint [2]. La technique de factorisation des matrices non-négatives [1] peut améliorer les performances de telles méthodes en rejetant les données non-pertinentes. D'autres méthodes comme les GPDM [18] introduisent des probabilités dans le calcul d'un espace latent pour adoucir les transitions entre les poses apprises mais les scènes de test sont restreintes aux mouvements cycliques. Les méthodes qui procèdent à une comparaison entre l'image et une base apprise exigent des bases de très grande dimension, même lorsque le résultat est déduit d'une régression localement pondérée sur les plus proches voisins [3]. L'augmentation de la taille de la base d'apprentissage a pour effet de ralentir drastiquement le processus de comparaison et, pour accélérer la sélection d'un ensemble de plus proches voisins, il est possible d'utiliser une méthode de hachage associée à une mesure des distances de Hamming [13][17]. La vraisemblance d'une pose est calculée avec cette dernière méthode en utilisant le formalisme Bayésien mais certaines poses sont mal estimées lorsqu'elles sont trop éloignées des exemples appris. En général, la très grande taille de l'espace des poses et la variabilité des paramètres externes tels que l'habillement ou la coiffure compliquent l'implémentation des méthodes à base d'apprentissage.

En vision monoculaire, les méthodes stochastiques qui utilisent un algorithme multi-hypothèses sont souvent efficaces pour lever les ambiguïtés dues aux inférences de la pose 3D à partir d'une image 2D. Le filtre à particules fait partie de ces méthodes mais, dans son implémentation classique, le nombre de particules très élevé nécessaire au suivi d'une personne entraîne des temps de calcul prohibitifs. Une solution à ce problème consiste à choisir un modèle de corps à membres indépendants [14] où la vraisemblance de chaque membre est calculée de manière indépendante. Utilisé avec le filtrage particulaire, un tel modèle permet d'associer un filtre à particules à chacun des membres et restreindre la dimension de l'espace d'exploration au nombre de *ddl* du membre considéré [4]. Les liaisons entre les membres sont prises en compte en propageant les croyances associées aux membres à travers un graphe de facteurs en utilisant la propagation des croyances [10]. Contrairement à [14] qui a recours à un échantillonneur de Gibbs pour propager les croyances dans un espace continu, la technique proposée dans [4] permet de les propager récursivement et plus simplement dans l'espace discret des échantillons.

D'autres approches montantes existent dans le cadre du suivi en monoculaire. L'une d'entre-elles utilise aussi la propagation des croyances mais exploite seulement un indice d'énergie de mouvement pour la détection des bras [9]. Une seconde mêle l'algorithme du champ moyen aux techniques de Monte-Carlo et prend en compte les contours et les niveaux de gris mais le modèle 2D adopté contraint les mouvements à rester dans le plan de l'image [19]. L'utilisation de tableaux noirs hiérarchisés permet également d'assurer la cohérence des solutions pour effectuer de l'estimation de pose 2D en exploitant un ensemble judicieux d'indices portant sur les contours, les couleurs et une soustraction de fond [11].

Dans cet article, l'approche montante adoptée en stéréo par [4] qui consiste à mettre en œuvre un ensemble de filtres à

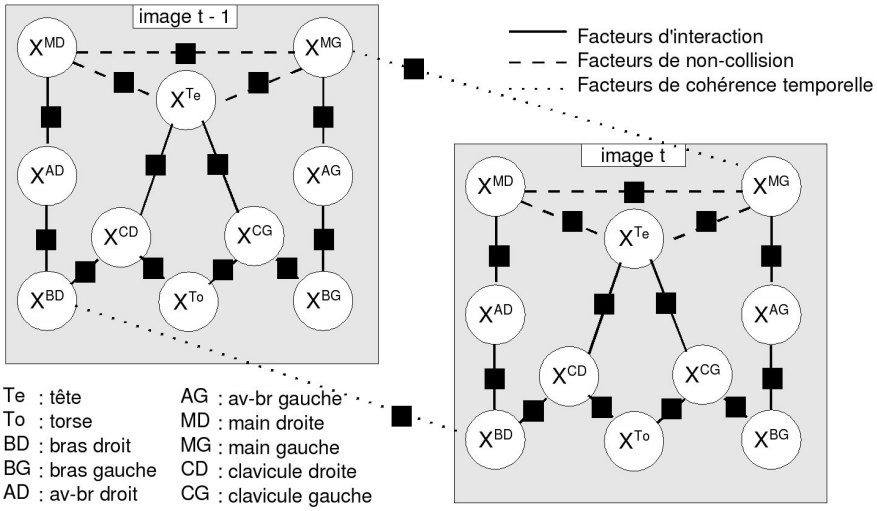

FIG. 3 – Graphe de facteurs. Les cercles correspondent aux noeuds qui contiennent les variables définissant l'état des membres et les carrés noirs aux facteurs entres ces différents noeuds (T^μ pour les facteurs de cohérence temporelle et $\psi^{\mu\nu}$ pour les facteurs d'interactions ou de non-collision). Dans un soucis de clarté, seulement deux images consécutives sont représentées avec deux facteurs de liens temporels. Les facteurs ainsi que les noeuds qui correspondent aux observations Y^μ sont omis.

particules en intéraction avec la propagation des croyances est réinvestie. Pour compenser le manque d'informations dû au passage en monoculaire et lever les ambiguïtés, un ensemble complémentaire d'indices pertinents est utilisé conjointement à des règles déduites des contraintes articulaires humaine. Ces règles articulaires sont directement intégrées au processus de propagation des croyances et permettent de traiter une plus large variété de poses qu'une contrainte modélisée par une mixture de Gaussiennes apprises sur des séquences de marche [14]. L'algorithme proposé est capable d'estimer la pose en 3D à partir d'une simple webcam à une cadence de 6 im/s.

2 Suivi Bayésien récursif

Le haut du corps est modélisé sous la forme d'un graphe incluant M membres représentés pas les noeuds. Les liens représentent les articulations ou les contraintes de non-collision entre les membres (figure 2). Un réseau de Markov suffirait à représenter une telle structure mais les contraintes de non-collision entre la tête et les mains génèrent des cliques à trois noeuds. Un graphe de facteurs est préféré en imposant un facteur entre chaque paires de noeuds [4]. La probabilité conditionnelle jointe peut être décomposée comme le produit de ces facteurs. Le graphe complet inclut les états précédents pour tenir compte de la cohérence temporelle (figure 3).

Soit μ un membre, X_t^μ son état à l'instant t et Y_t^μ les observations tirées de l'image pour ce membre. Les paramètres du modèle sont :

– les facteurs de compatibilité par rapport aux observations, $\phi^\mu(X^\mu, Y^\mu)$,
– les facteurs d'interaction temporels $T^\mu(X_t^\mu, X_{t-1}^\mu)$,
– le facteur d'interaction entre les membres μ et ν, $\psi^{\mu\nu}(X^\mu, X^\nu)$.

En adoptant ces notations, la probabilité jointe connaissant toutes les observations entre les instants 0 et T est :

$$P(X_{0:T}|Y_{0:T}) = \prod_{t=0}^{T} \Phi(X_t, Y_t)\Psi(X_t) \prod_{t=1}^{T} T(X_t, X_{t-1}) \ , \tag{1}$$

avec :

$$\Phi(X_t, Y_t) = \prod_{\mu=1}^{M} \phi^\mu(X_t^\mu, Y_t^\mu), \tag{2}$$

$$\Psi(X_t) = \prod_{(\mu,\nu)\in\Gamma} \psi^{\mu\nu}(X_t^\mu, X_t^\nu), \tag{3}$$

$$T(X_t, X_{t-1}) = \prod_{\mu=1}^{M} T^\mu(X_t^\mu, X_{t-1}^\mu), \tag{4}$$

où Γ est l'ensemble des liens entre les membres.

La probabilité marginale des membres est obtenue en appliquant l'algorithme de propagation des croyances à l'intérieur du graphe de facteurs utilisé pour modéliser le haut du corps [4]. Comme ce graphe comprends des boucles, la probabilité marginale obtenue n'est qu'une approximation de la probabilité réelle. De plus, la valeur de cette approximation dépend de l'ordre choisi pour mettre à jour les messages. Pour simplifier l'algorithme, les messages sont d'abord propagés dans une même tranche temporelle du graphe qui représente le modèle du corps à l'image $t-1$ avec un nombre fixé d'itérations (10 dans notre cas). Puis,

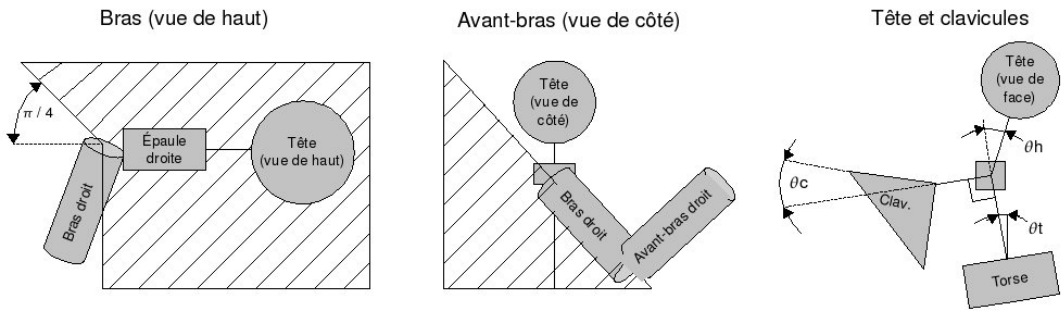

FIG. 4 – Contraintes articulaires. Bras et avant bras : les parties hachurées montrent les zones interdites. Les contraintes angulaires sont : $|\theta_c| \leq 15°$ pour les clavicules et $|\theta_h| \leq 25°$ pour la tête. L'inclinaison du torse θ_t n'est pas limitée.

ils sont propagés unidirectionellement vers l'image suivante à l'instant t en une seule itération (figure 3). Cette façon de propager les messages implique que l'estimation d'une probabilité marginale à l'instant t ne dépend pas des observations futures à cet instant. Les marginales peuvent donc être calculées récursivement.

Les messages sont représentés par un ensemble d'échantillons pondérés. Ils sont calculés d'une image à l'autre en utilisant le schéma classique du filtre à particules consistant en une phase de reéchantillonnage suivie d'une phase de prédiction basée sur les facteurs de cohérence temporels [5]. L'espace d'état de chaque membre étant réduit à ses propres échantillons, la propagation des croyance à l'image t correspond à une propagation des croyances discrète à l'intérieur d'un graphe bouclé. De plus, une estimation de la probabilité marginale d'un membre est fournie par la somme pondérée de ses échantillons. De cette manière, une estimation pleinement récursive est obtenue et l'algorithme est analogue à un ensemble de filtres à particules en interaction où le poids des échantillons est réévalué à chaque image grâce à la propagation des croyance pour tenir compte les liens entre les membres.

L'algorithme est relativement rapide car, contrairement à [14], les facteurs de compatibilité par rapport aux observations $\phi^{\mu}(X_t^{\mu}, Y_t^{\mu})$ doivent êtres évalués une seule fois pour chaque échantillon et le facteur d'interaction entre deux membres connectés μ et ν, $\psi^{\mu\nu}(X_t^{\mu}, X_t^{\nu})$, une seule fois pour chaque paire d'échantillons.

3 Application au suivi du haut du corps en monoculaire

Ce modèle est appliqué au suivi du haut du corps à partir d'images couleur issues d'une simple webcam. L'information couleur sert à suivre la tête et les mains tandis que les niveaux de gris sont utilisés pour extraire l'énergie de mouvement, la carte des contours et procéder à une soustraction de fond (figure 1).

3.1 Initialisation

Le visage est d'abord localisé dans l'image couleur grâce à un détecteur de visages robuste [8]. La pose de départ suppose que les bras se trouvent le long du corps et le torse en position verticale faisant face à la caméra. Le suivi peut aisément raccrocher la pose réelle tant que celle-ci n'est pas trop éloignée de cette hypothèse. Le résultat de la détection de visage sert aussi à initialiser un histogramme de la couleur du visage.

3.2 Modèle du corps et contraintes articulaires

La figure 2 montre le modèle de corps utilisé. Certains membres sont représentés en 3D en utilisant une sphère pour la tête et des cylindres pour les bras et les avant-bras. En revanche, les mains, les clavicules et le torse sont représenté respectivement par des cercles, des triangles et un rectangle faisant face à la caméra. Les membres sont discrétisés dans l'espace à l'aide de points régulièrement distribués autour d'eux, leur taille est choisie d'après des données anthropométriques moyennes.

Un ensemble de règles sur les contraintes articulaires agissent directement sur le calcul des facteurs d'interaction. Un facteur d'attirance entre les membres adjacents prend la forme d'une Gaussienne de la distance entre ces membres (voir la figure 2 pour les distances Dn, Ds, De et Dw). Ce facteur agit comme un ressort entre les membres adjacents procurant ainsi une certaine tolérance dans le choix de la taille des membres. La Gaussienne utilisée est centrée sur une valeur égale à la distance qui sépare les mains des avant-bras pour les poignets, et celle qui sépare la tête de la base du cou pour le cou. En revanche, elle est centrée en zéro pour les épaules et les coudes.

Pour respecter les limites articulaires, les facteurs d'interaction torse-tête ou torse-clavicules sont nuls si les angles θh ou θc dépassent un seuil fixé (figure 4). De même, si les bras ou les avant-bras entrent dans des zones définies comme interdites (figure 4) les facteurs résultants sont nuls. Pour compléter cet ensemble de règles destiné à fournir des poses cohérentes, des contraintes de non-collision (figure 2) constituent des liens additionnels qui renvoient une valeur de facteur nulle lorsque les mains et la tête entrent en collision.

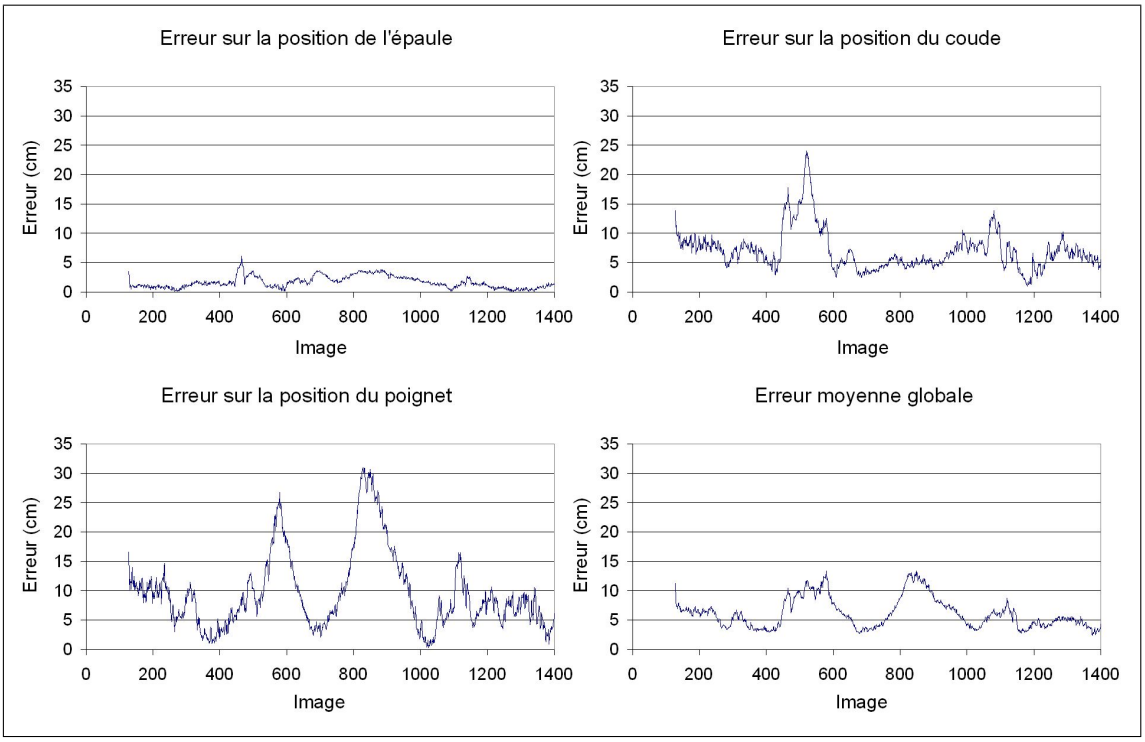

FIG. 5 – Résultats quantitatifs. Pour chaque articulation, l'erreur correspond à la différence entre la vérité de terrain et la position estimée. La moyenne des erreurs commises sur les trois articulations [15, 17] est donnée par la courbe "Erreur moyenne globale".

3.3 Facteur de cohérence temporelle

Les facteur de cohérence temporelle $T^{\mu}(X_t^{\mu}, X_{t-1}^{\mu})$ sont de simples Gaussiennes centrées sur les paramètres de la pose estimée à l'image précédente. Pour les mains qui peuvent bouger plus rapidement et dont la vitesse peut varier brusquement, le facteur de cohérence temporelle est un mélange de deux Gaussiennes, l'une centrée sur les paramètres de la pose de l'image précédente et l'autre sur la prédiction qui utilise la vitesse de la main à l'image précédente pour estimer la pose à l'image actuelle. L'écart type des Gaussiennes est de $10cm$ pour les mains et de $5cm$ pour les autres membres. Pour les angles, il est égal à $\pi/8$.

FIG. 6 – Position du torse. La grille de points noirs sur le bord inférieur de l'image modélise la position du bassin. Elle se déplace horizontalement pour maximiser la correspondance entre les points de la grille et les pixels détectés positifs par la soustraction de fond (pixels blancs). L'énergie est maximum lorsque la grille est centrée sur la zone inférieure de l'image marquée positivement par la soustraction de fond. Le haut du torse est situé entre les deux clavicules.

4 Exploitation des indices extraits de l'image

Les facteurs de compatibilité par rapport aux observations extraites de l'image $\phi^{\mu}(X_t^{\mu}, Y_t^{\mu})$ sont calculés à partir des scores S_f^{μ} représentant la compatibilité entre l'hypothèse d'un membre μ et l'indice f extrait de l'image. Contrairement aux images stéréo [4], la vision monoculaire exige plus d'indices pour atteindre un niveau de robustesse satisfaisant. Une compatibilité basée sur une fusion multi-indices procure un score global : $S^{\mu} = \prod_f S_f^{\mu}$. Pour éviter les effets néfastes dûs aux distracteurs issus du fond de l'image, une soustraction de fond robuste aux variation lumineuses et aux mouvements des petits objets est utilisée [12].

4.1 Suivi des mains et du visage

Un modèle de couleur est crée en calculant un histogramme normalisé des couleurs du visage à partir de la position de celui-ci détecté pendant la phase d'initialisation (§ 3.1). Les pixels p qui correspondent à la projection $proj(\mu)$ pour la pose candidate, des points appartenant au membre μ sont comparés à ce modèle afin de déterminer le score de couleur :

$$S_c^{\mu} = \sum_{p \in proj(\mu)} H(p) \qquad (5)$$

La fonction $H(p)$ renvoie la valeur de la case d'histogramme qui correspond à la couleur du pixel p.

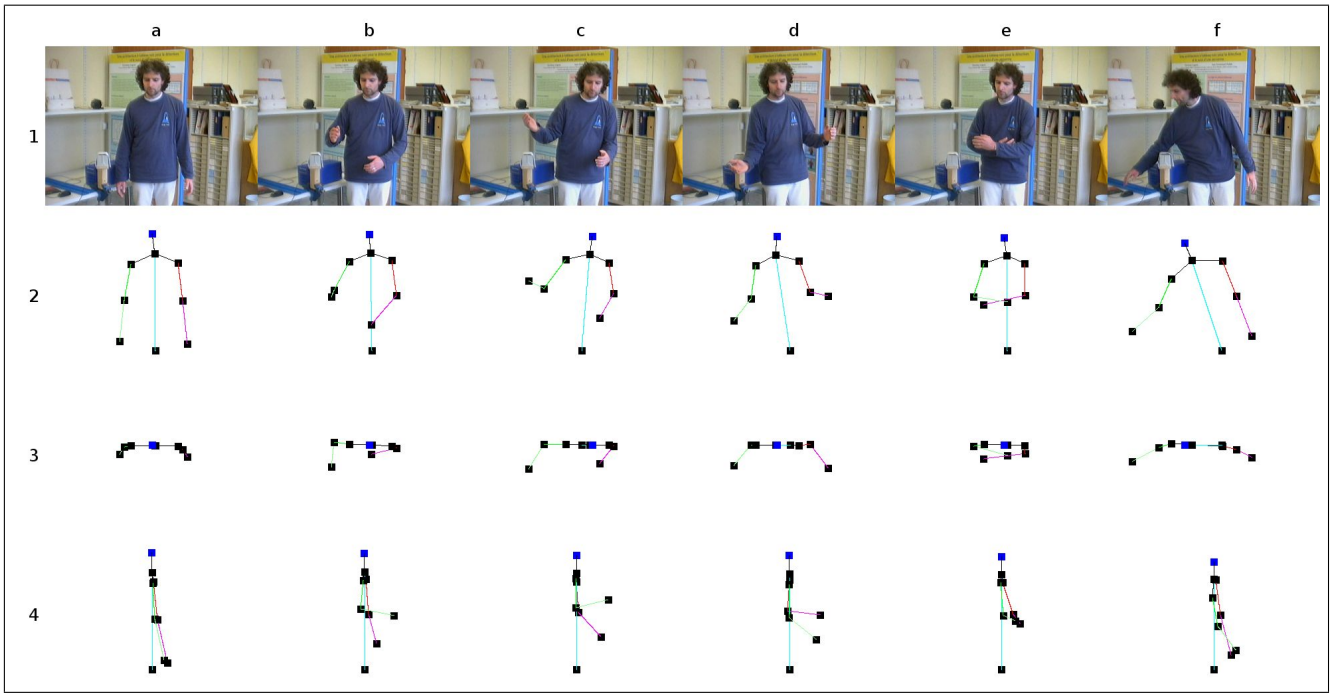

FIG. 7 – Résultats sur la séquence de test à partir de laquelle la vérité de terrain a été mesurée. Colonne (a) à (f) : images 246 (initialisation), 409, 709, 813, 1184 et 1437. Les colonnes 2, 3 et 4 montrent les vues de face, de haut et de droite des poses estimées.

4.2 Suivi du torse

Du fait de la déformation des vêtements ou des occultations qui se produisent lorsqu'une personne est en mouvement, le torse est une partie du corps difficile à localiser avec précision. Cependant, la position du bassin peut être estimée de manière précise si on suppose qu'il se trouve dans la zone inférieur de l'image. Il suffit pour cela de faire interagir une grille de points p pondérés et le résultat de la soustraction de fond (figure 6). Le score attribué au torse est :

$$S^t = \sum_{p \in t} W(p) Bg(p), \tag{6}$$

où $W(p)$ est le poids de p qui correspond à la Gaussienne de la distance entre p et le centre de la grille. $Bg(p)$ retourne la probabilité que le pixel p appartienne au premier plan en considérant le résultat de la soustraction de fond [12]. La grille est libre de se mouvoir horizontalement en restant collée sur le bord inférieur de l'image. Le haut du torse est situé au niveau de la nuque à demi-distance des deux clavicules (base du coup dans la figure 2).

4.3 Suivi des bras, des avant-bras, et des clavicules

Les bras ont tendance à bouger rapidement et sont fréquemment sujet aux occultations. Dans ces conditions, une fusion d'indices basée sur les contours et l'énergie de mouvement permet d'obtenir un meilleur degré de robustesse.
Un score de contours est estimé en prenant non seulement en compte l'intensité des contours mais aussi leur orientation. Soit $M(\|\overrightarrow{p}\|)$, une fonction qui pénalise les contours de faible et de forte amplitude $\|\overrightarrow{p}\|$:

$$M(\|\overrightarrow{p}\|) = \frac{1}{\lambda}\|\overrightarrow{p}\| tanh(\frac{\lambda}{\|\overrightarrow{p}\|}), \tag{7}$$

λ étant un paramètre de réglage. Un score S^{μ}_{or} est calculé pour l'hypothèse d'un membre supérieur μ en considérant la Gaussienne G_θ de la différence entre l'orientation θ_{limb} du membre et l'orientation du contour θ_p obtenue en chaque pixel p qui correspond à la projection dans l'image $proj(\mu)$ des points appartenant au membre hypothèse :

$$S^{\mu}_{or} = \sum_{p \in proj(\mu)} M(\|\overrightarrow{p}\|) G_\theta[\theta_{limb} - \theta_p] \tag{8}$$

Le score de l'énergie de mouvement avantage les hypothèses des membres situées dans des zones contenant du mouvement. Il est calculé en considérant la Gaussienne G_m de la distance $d(p)$ entre le pixel p qui correspond à la projection dans l'image d'un point appartenant au membre concerné et le plus proche pixel où un mouvement a été détecté :

$$S^{\mu}_{m} = 1 + \sum_{(p \in \mu)} G_m(d(p)). \tag{9}$$

Le calcul de $d(p)$ fait appel à la distance de chanfrein [6]. La constante égale à 1 additionnée au score permet de neutraliser celui-ci quand il n'y a pas de mouvement dans l'image. La détection de mouvement est assurée par la soustraction des images consécutives deux à deux. Pour les clavicules, seul le score de contours est activé car elles sont suffisamment contraintes par la position de la tête.

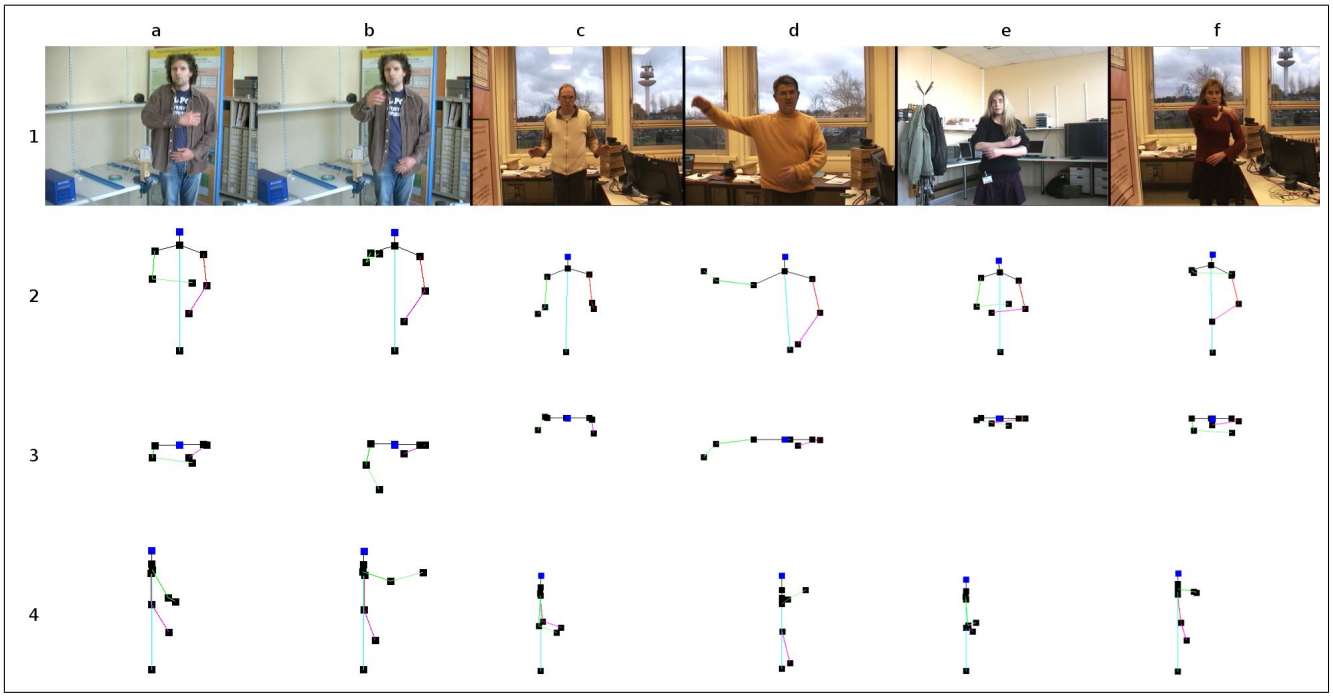

FIG. 8 – Suivi 3D en monoculaire. Quelques poses présentant des difficultés majeures telles que des occultations, des fonds complexes et des environnements non contraints (conditions lumineuses et habits variés).

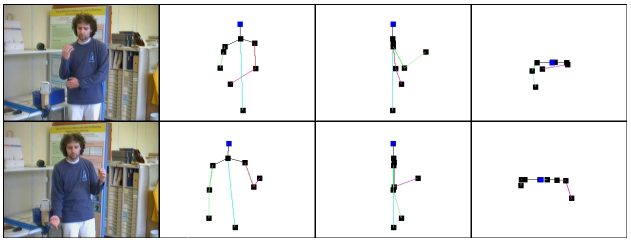

FIG. 9 – Exemple d'erreurs d'appréciation de la profondeur pour les images 581 (première ligne) et 850 (seconde ligne). Dans les deux cas, l'avant-bras droit estimé n'est pas suffisamment tendu entraînant une erreur de plus de $25cm$ sur la position du poignet.

5 Résultats expérimentaux

L'algorithme a été testé en utilisant une séquence acquise à partir d'une simple webcam cadencée à 30 im/s. Les résultats quantitatifs ont été obtenus en comparant les poses estimées à la vérité de terrain fournie par un capteur magnétique de mouvement. Il est doté d'une précision supérieure à $5mm$ et permet de mesurer la position des articulations du membre supérieur droit (épaule, coude et poignet). La transformation géométrique allant du repère capteur du magnétique à celui de la caméra est estimée grâce au calibrage de la caméra pour laquelle un modèle pinhole est adopté. La séquence de test comprend des mouvements dans les trois dimensions occasionnant des occultations avec un fond complexe (figure 7). Les gestes sont exécutés de manière naturelle sans ralentir la vitesse des mouvements lors de la prise de vue (tableau 1). En plus de l'erreur moyenne globale [15, 17], les résultats fournissent l'erreur d'estimation pour chaque articulation (figure 5). Des résultats qualitatifs sont également donnés à la figure 8 où des séquences présentant différents utilisateurs arborant des tenues vestimentaires variées devant différents fonds et sous différents éclairages sont testés avec succès.

En suivi monoculaire du corps, l'estimation de la profondeur entraîne les erreurs les plus importantes. Dans le cas de notre séquence de test, une erreur de profondeur sur l'estimation du coude autour de l'image 500 contraint exagérément le poignet vers l'avant (figure 9). Le même problème survient de nouveau autour de l'image 850 où l'avant-bras se plie dans la direction perpendiculaire à l'image entraînant une estimation faussée de la position du poignet par l'algorithme qui positionne l'avant-bras le long du corps (figure 9). Néanmoins, l'erreur estimée tout au long de la scène test ne dépasse pas 31 cm et reste en dessous de 15 cm si on considère le protocole de mesures globales utilisé dans [15, 17] (tableau 1). Le traitement des occultations entre membres ou entre un objet et un membre est gérée de manière satisfaisante (figure 10). En l'absence de scènes test standards, la comparaison des algorithmes de suivi entre-eux reste une tâche difficile du fait des dissemblances évidentes entre les scènes de test utilisées pour estimer les performances des différentes approches. Cependant, en s'en tenant aux résultats quantitatifs, la précision obtenue est aussi bonne voire meilleure que celle atteinte par des approches au sommet de l'état de l'art [15, 17]. De plus, l'algorithme décrit ici présente l'avantage de fonctionner en quasi temps réel (6 im/s) avec un matériel

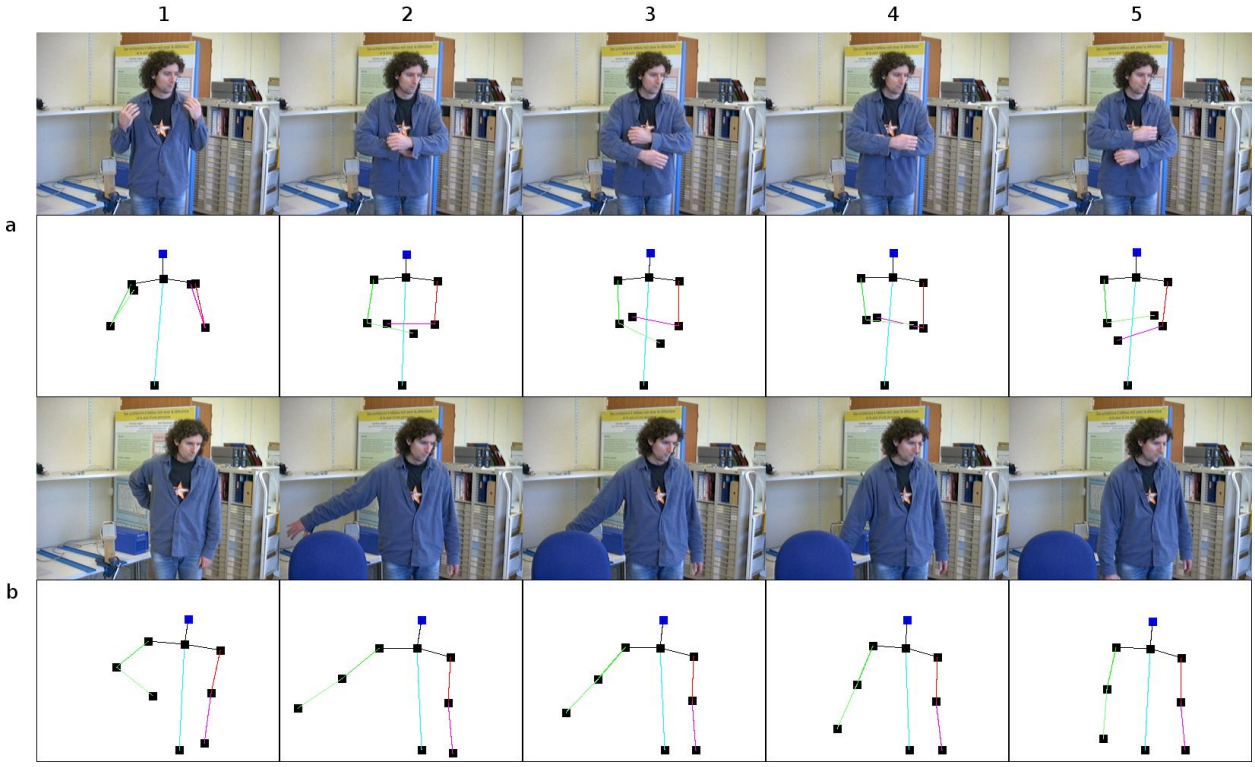

FIG. 10 – *Cas d'auto-occultations des bras (a1) et de la main par l'avant-bras (a2, a4), une main derrière le dos (b1) et occultation de la main par le dossier d'un fauteuil (b2 à b5).*

presque grand public (PC bi-processeur Xeon $3,4\ GHz$ équipé d'une webcam).

6 Conclusion

Nous avons présenté un algorithme destiné à accomplir un suivi du haut corps par caméra monoculaire capable de traiter des images issues d'une simple webcam à une cadence de $6\ im/s$. Cet algorithme fonctionne avec succès dans des environnements non contraints vis-à-vis des vêtements, du fond ou des conditions lumineuses. Une fusion d'indices complémentaires extraits de l'image permet de compenser le manque d'informations sur la profondeur. La propagation des croyances permet de réduire la dimension de l'espace dans lequel les hypothèses sont formulées et rend l'utilisation du filtre à particules pertinente. Les contraintes articulaires sont aisément implémentées dans le calcul des facteurs pour fournir des solutions de poses cohérentes. Les travaux futurs incluront un terme de compatibilité basé sur l'apprentissage pour gérer plus efficacement les auto-occultations et améliorer la précision avec laquelle la profondeur des membres est estimée.

Références

[1] Ankur Agarwal and Bill Triggs. A local basis representation for estimating human pose from cluttered images. In *ACCV (1)*, pages 50–59, 2006.

[2] Ankur Agarwal and Bill Triggs. Recovering 3d human pose from monocular images. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 28(1), jan 2006.

[3] Christopher G. Atkeson, Andrew W. Moore, and Stefan Schaal. Locally weighted learning. *Artif. Intell.*

Erreur (cm)	Épaule	Coude	Poignet	Erreur moyenne globale
Moyenne	1.7	7.1	9.7	6.1
Maximum	6.1	24.1	31.0	13.4
Écart-type	1.0	3.5	6.6	2.6
Vitesse moyenne ($cm.s^{-1}$)	2.83	4.28	8.5	

TAB. 1 – Moyenne, maximum et écart-type de l'erreur commise sur l'épaule, le coude et le poignet. L'erreur moyenne globale est la moyenne des erreurs commises en estimant la position de ces trois articulations. La vitesse moyenne est calculée sur toute la séquence pour chaque articulation.

Rev., 11(1-5) :11–73, 1997.

[4] Olivier Bernier and Pascal Cheung-Mon-Chang. Real-time 3d articulated pose tracking using particle filtering and belief propagation on factor graphs. In *British Machine Vision Conference*, volume 01, pages 005–008, 2006.

[5] Andrew Blake and Michael Isard. The condensation algorithm - conditional density propagation and applications to visual tracking. In *NIPS*, pages 361–367, 1996.

[6] Gunilla Borgefors. Distance transformations in digital images. *Comput. Vision Graph. Image Process.*, 34(3) :344–371, 1986.

[7] David Demirdjian, T. Ko, and Trevor Darrell. Constraining human body tracking. In *ICCV '03 : Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1071. IEEE Computer Society, 2003.

[8] Raphaël Féraud, Olivier Bernier, Jean Emmanuel Viallet, and Michel Collobert. A fast and accurate face detector based on neural networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 23(1) :42–53, 2001.

[9] Jiang Gao and Jianbo Shi. Multiple frame motion inference using belief propagation. In *FGR*, pages 875–882, 2004.

[10] Kschischang, Frey, and Loeliger. Factor graphs and the sum-product algorithm. *IEEETIT : IEEE Transactions on Information Theory*, 47, 2001.

[11] Christine Leignel and Jean Emmanuel Viallet. Une architecture à tableau noir pour la détection et le suivi d'une personne. In *RFIA*, 2004.

[12] Philippe Noriega and Olivier Bernier. Real time illumination invariant background subtraction using local kernel histograms. In *Proceedings of the British Machine Vision Conference*, pages 979–988, 2006.

[13] Gregory Shakhnarovich, Paul Viola, and Trevor Darrell. Fast pose estimation with parameter-sensitive hashing. In *ICCV '03 : Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 750. IEEE Computer Society, 2003.

[14] Leonid Sigal, Sidharth Bhatia, Stefan Roth, Michael J. Black, and Michael Isard. Tracking loose-limbed people. In *CVPR (1)*, pages 421–428, 2004.

[15] Leonid Sigal and Michael J. Black. Measure locally, reason globally : Occlusion-sensitive articulated pose estimation. In *CVPR '06 : Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2041–2048, Washington, DC, USA, 2006. IEEE Computer Society.

[16] Cristian Sminchisescu and Alexandru Telea. Human pose estimation from silhouettes - a consistent approach using distance level sets. In *WSCG*, pages 413–420, 2002.

[17] Leonid Taycher, David Demirdjian, Trevor Darrell, and Gregory Shakhnarovich. Conditional random people : Tracking humans with crfs and grid filters. In *CVPR (1)*, pages 222–229, 2006.

[18] Raquel Urtasun, David J. Fleet, and Pascal Fua. 3d people tracking with gaussian process dynamical models. In *CVPR '06 : Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 238–245, Washington, DC, USA, 2006. IEEE Computer Society.

[19] Y. Wu, G. Hua, and T. Yu. Tracking articulated body by dynamic markov network. In *ICCV*, pages 1094–1101, 2003.