

Extraction de silhouette et suivi des parties du corps dans un environnement non-contrôlé

Rada DEEB, Elodie DESSEREE, Saida BOUAKAZ
Université de Lyon, CNRS
Université Lyon1, LIRIS, UMR5205, F-69622, France
prénom.nom@liris.cnrs.fr

Résumé

Dans cet article nous nous intéressons à la détection de premier plan ainsi qu'à l'extraction de silhouette et le suivi des parties du corps dans une séquence vidéo monoculaire. La personne se trouve dans son domicile, dans un environnement domestique non contrôlé. Les problèmes rencontrés dans ce cas sont les changements d'illumination, les ombres, le camouflage, les occultations, etc. Dans cet article, nous proposons une approche en deux-niveaux pour extraire la silhouette en temps réel. Au premier niveau, nous extrayons le premier plan en modélisant le fond par un mélange de gaussiennes amélioré dans une représentation multi-niveaux (pixel, région, image) en utilisant des informations de chrominance. Au second niveau, pour améliorer les résultats de la segmentation dans les zones d'intérêt, nous utilisons une modélisation statistique du premier plan. Le modèle du premier plan est mis à jour par un algorithme de suivi des parties du corps. En s'appuyant sur une modélisation de type blobs et en ayant recours à la minimisation d'une distance dans l'espace des descripteurs, nous réalisons le suivi des différentes parties du corps identifiées dans l'image. La prise en compte de ces différents aspects (MOG amélioré, modélisation et suivi du premier plan) permet à notre méthode d'être robuste aux ombres, aux changements de luminance ainsi qu'au phénomène de camouflage. Les résultats obtenus sur des cas réel montrent la capacité de notre méthode de supprimer efficacement le fond, et d'extraire la silhouette.

Mots clefs

Extraction de silhouette, suivi des parties du corps, modélisation de premier plan, temps réel.

1 Introduction

L'analyse du mouvement humain est actuellement l'un des domaines de recherche les plus actifs dans la vision par ordinateur. Dépassant le cadre académique, ce domaine suscite l'intérêt de différents secteurs économiques tant les domaines d'application sont nombreux. On peut citer à titre d'exemple la surveillance « sécuritaire », la domotique, la robotique, l'interaction homme-machine, l'analyse des performances athlétiques,

etc. Dans ce contexte, les systèmes d'analyse du mouvement tentent de localiser la (ou les) personne(s) présente(s) dans une séquence vidéo, d'identifier et suivre les différentes parties du corps, et enfin d'interpréter les comportements et les situations selon l'application visée. Dans un cadre général d'un système d'analyse du mouvement humain, l'extraction de la silhouette est l'une des étapes initiales. La qualité du résultat a une forte incidence sur l'efficacité de l'ensemble du système, d'où l'importance de disposer d'approches fiables pour l'extraction de premier plan. Malgré les travaux réalisés, ce domaine est encore confronté à de nombreux défis induits par le non contrôle de l'environnement tels que les changements de lumière, la présence d'ombres et le camouflage, qui rendent l'extraction de la silhouette assez ardue. Par ailleurs, l'information contenue dans la silhouette extraite est utilisée pour détecter et suivre les mouvements de la personne dans la séquence.

Dans cet article nous présentons une méthode pour l'extraction de la silhouette et le suivi des parties du corps dans une séquence monoculaire prise dans un environnement non contrôlé. Dans notre démarche, nous avons appliqué une méthode de mélange de gaussiennes où le premier plan est extrait en utilisant deux types d'information : (i) information utilisant la composante de chrominance de l'espace couleur, (ii) information extraite des composantes de chrominance et de luminance. Pour obtenir les différentes composantes du premier plan, ce processus est complété par une segmentation de type croissance de région. L'algorithme de suivi s'appuie sur des modèles de type blobs que nous construisons pour chaque partie de corps. La minimisation d'une distance dans l'espace des descripteurs permet de réaliser le suivi des différentes parties du corps. Enfin, on introduit une fonction d'énergie, basée sur les champs markoviens tenant compte des modèles statistiques des parties du corps et de fond. La minimisation de cette fonction permet de corriger et compléter la silhouette extraite dans l'image en cours.

La présentation de notre méthode se déroule selon le plan suivant : dans la section 2 nous présentons les méthodes de l'état de l'art. Dans la section 3, nous développons la méthode d'extraction de premier plan par un mélange de gaussiennes amélioré. Dans la section 4, nous décrivons le

modèle de parties du corps ainsi que la méthode de suivi et de mise en correspondance. Dans la section 5, nous présentons la segmentation entre fond/parties du corps par une minimisation d'énergie. Section 6 montre les résultats obtenus. Enfin, la section 7 conclue notre travail et évoque les perspectives associées.

2 Etat de l'art

Les stratégies adoptées dans la littérature pour l'extraction de premier plan peuvent être divisées en deux grandes catégories : celles basées sur la modélisation de fond et celle qui s'attachent à la modélisation simultanée fond / premier plan. De façon schématique, le premier ensemble de méthodes utilise un modèle d'arrière-plan, suivi d'une soustraction du fond pour obtenir l'image de premier plan. Concernant la modélisation du fond, les premières méthodes se sont orientées vers une modélisation par une gaussienne unique par pixel. Pour tenir compte de l'aspect multimodal, Stauffer et Grimson [1] sont parmi les premiers à proposer un modèle de mélange de gaussiennes. Kim *et al.* [2] proposent une méthode dictionnaire « codebook », qui peut gérer des fonds mouvants et bruyants, mais cette méthode nécessite une étape d'apprentissage qui peut être longue. En règle générale, les techniques de modélisation de fond sont robustes au bruit et elles peuvent s'adapter à de nouveaux objets d'arrière-plan, cependant, elles demeurent vulnérables à beaucoup de phénomènes induits par l'environnement tels que les ombres, les changements d'illumination globale et le camouflage. Beaucoup de méthodes de détection d'ombre ont été utilisées pour réduire les effets des ombres [3]. La plupart de ces méthodes utilisent des informations locales au niveau pixel. Le principal défaut d'une telle démarche, est qu'en plus de l'indisponibilité des modèles de premier plan, un traitement local des ombres (pixel) engendre de faux négatifs et faux positifs. Par ailleurs, l'absence d'un modèle pour représenter le premier plan induit un phénomène de camouflage, qui résulte souvent de la proximité des couleurs du fond et de régions du premier plan.

De nombreuses alternatives ont été proposées afin de pallier ces inconvénients. Ces approches se basent sur la modélisation à la fois du premier plan et du fond. L'extraction du premier plan est dans ce cas traitée comme un cas de segmentation. Yu *et al.* [4] proposent l'utilisation des modèles de mélange de gaussiennes spatio-couleur à la fois pour le fond et le premier plan. Ils réalisent une segmentation en minimisant une fonction d'énergie contenant ces modèles. Toutefois, pour manipuler les mouvements de grandes amplitudes, ils font appel à un algorithme Maximisation de l'espérance ou *Expectation-Maximization* (EM) pour mettre ces modèles à jour avant l'étape de segmentation. Gallego *et al.* [5] utilisent le même modèle pour le premier plan, mais proposent l'utilisation d'une seule gaussienne par pixel

pour modéliser l'arrière-plan. Ces approches sont basées sur l'idée que l'utilisation de modèles de premier plan contribue à donner une meilleure classification et à surmonter l'un des problèmes les plus difficiles dans l'extraction de premier plan : le camouflage. Cependant, ces techniques ne sont pas bien adaptées aux changements de fond ou de premier plan. En outre, l'utilisation d'EM et de coupe de graphe (graph cut) sur tous les pixels de l'image limite l'application de ces méthodes dans le cas des systèmes nécessitant le temps réel.

En proposant une représentation par blobs du premier plan, Wren *et al.* [6] réalise une segmentation par une minimisation d'énergie et un suivi par un filtre de Kalman. En s'appuyant sur la forme à suivre, les auteurs améliorent les résultats. Néanmoins, le suivi avec le filtre de Kalman n'est pas efficace en cas d'occultation. Dans le but d'avoir un système de suivi qui peut traiter les occultations, des filtres à particules basée sur l'algorithme de condensation [7] ont été utilisés [7–9]. Ces approches dépendent notamment du nombre de particules utilisées, et peuvent s'avérer gourmandes en temps de calcul. Notons que le suivi de la personne reste vulnérable aux occultations et au camouflage, et que le suivi d'une composante partielle du corps ne suffit pas pour prédire la posture générale de la personne. Pour gérer les occultations Javed *et al.* [10] mettent en correspondance les régions à suivre par un calcul de distance de caractéristiques extraites de ces régions.

3 Extraction du premier plan

Dans cette partie, nous présentons la méthode que nous utilisons pour l'extraction du premier plan. Pour être opérationnelle dans un environnement quelconque, cette méthode doit être robuste aux ombres et aux changements rapides et/ou globaux de lumière dans la scène.

3.1. Modèle de fond

Pour extraire les pixels du premier plan, pour chaque pixel du fond, nous adoptons la modélisation par un mélange des gaussiennes (MOG)¹ proposée par Stauffer et Grimson [1]. Dans cette approche, l'histoire récente des observations à un pixel donné est considérée comme un processus statistique indépendant de celle de tous les autres pixels. Elle est modélisée par un mélange de L gaussiennes. Si on note par $X_{i,t}$ l'observation d'un pixel i au temps t , les observations sont prises dans l'espace couleur $L * a * b^*$ on peut poser :

$$X_{i,t} = [l_{i,t}, a_{i,t}, b_{i,t}] \quad (1)$$

Notons que dans la suite, l'indice i pourra être ignoré.

¹ Mixture Of Gaussian

Dans un modèle MOG, la probabilité qu'une valeur du pixel appartenant à l'arrière-plan est donnée par :

$$P(X_t) = \sum_{k=1}^L \omega_{k,t} \cdot \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \quad (2)$$

Où L est le nombre de gaussiennes, η est la fonction gaussienne donnant la densité de probabilité, $\omega_{k,t}$ est une pondération de la gaussienne, $\mu_{k,t}$ est le vecteur qui contient les moyennes pour chaque canal, $\Sigma_{k,t}$ est la matrice de covariances. Pour faciliter le calcul, nous adoptons la même hypothèse que [1] : indépendance des canaux l^*a*b^* , d'où $\Sigma_{i,t} = \sigma_{i,t}^2 I$.

Le modèle est construit et mis à jour au cours du temps par un algorithme de maximisation de l'espérance (EM) en utilisant l'ensemble des équations de mise à jour proposé en [11]. L'adaptation des poids, des moyennes, et des variances de chaque gaussienne est faite par rapport aux correspondances trouvées entre chaque pixel et les gaussiennes correspondantes. L'adaptation se base aussi sur le nombre d'images traitées permettant au modèle d'apprendre la scène rapidement, puis de conserver les connaissances acquises.

3.2. Niveau pixel

Dans un premier temps, une distance de Mahalanobis est calculée entre la valeur récente du pixel p et toutes les gaussiennes relatives à ce pixel. On considère qu'on a une correspondance entre la valeur de pixel et la i ème gaussienne calculée sur tous les canaux de l'espace couleur si :

$$\sqrt{(X_{t+1} - \mu_{k,t})^T \cdot \Sigma_{k,t}^{-1} \cdot (X_{t+1} - \mu_{k,t})} < 2.5 * \sigma_{i,t} \quad (3)$$

Nous obtenons l'image de premier plan (Figure 1. a) en appliquant un seuil sur le poids de la gaussienne correspondante pour décider s'elle correspond au fond ou au premier plan.

Pour les pixels de premier plan, en utilisant uniquement les composantes de chrominance, nous calculons une seconde distance de Mahalanobis. De la même façon, en appliquant un seuil sur le poids de la gaussienne correspondante, nous obtenons des régions qui peuvent potentiellement correspondre aux ombres ou aux changements de luminance (Figure 1. b).

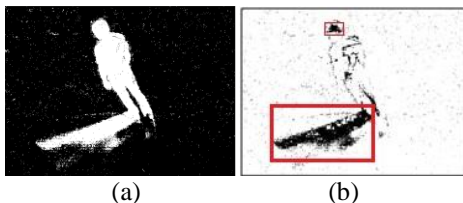


Figure 1 – Les résultats obtenus de niveau pixels.
(a) l'image résultat de l'extraction de fond obtenue en utilisant toutes les composantes de l'espace l^*a*b^* .

(b) les régions candidates d'être étiquetées ombres ou changements de luminance.

3.3. Niveau région

Le traitement au niveau région permet d'éliminer les régions du premier plan, susceptibles de correspondre aux ombres ou aux changements rapides de lumière. Notons que ces régions sont détectées sur la base des composantes de chrominance (Figure 1.b). Une telle région est considérée comme l'effet d'un changement de luminance si sa suppression n'affecte pas la connectivité du premier plan. Cette suppression est confirmée si elle n'implique pas de changements de la taille de la silhouette : cette vérification est faite par comparaison de boîtes englobantes estimées dans l'image courante et l'image précédente.

3.4. Niveau image

L'analyse au niveau image permet de détecter les changements globaux dans la scène tout en assurant la conservation de la silhouette jusqu'à la fin du processus d'adaptation. On comptabilise le nombre de pixels détectés comme premier plan : si ce nombre est assez important par rapport au nombre total de pixels dans l'image (soit plus de 75% de nombre totale de pixels), on considère qu'on a un changement global de luminance dans la scène. Dans ce cas, on supprime du premier plan toutes les régions susceptibles de correspondre aux ombres ou aux changements de lumière.

4 Suivi des parties du corps

4.1. Modèle des parties du corps

A l'instar de beaucoup de chercheurs, nous adoptons un modèle M simplifié pour représenter un personnage. Ce modèle est composé de 6 parties : une tête, deux membres supérieurs, un torse et deux membres inférieurs (ces 2 derniers peuvent être confondus pour gérer différents types d'habillement). Chaque partie du modèle est représentée par un ou plusieurs blocs. La création du modèle est réalisée en deux étapes : d'abord, les articulations sont détectées par une approche d'analyse de contour de la silhouette du personnage détecté dans l'étape précédent [12]. Une partie de corps est créée pour chaque région entre deux articulations successives.

Une segmentation par croissance de régions de chacune des parties du corps détectée permet de construire les blocs correspondants à cette partie. Un bloc correspond à une région homogène au sens de la couleur. A chaque bloc k , on associe un vecteur de caractéristiques contenant : les coordonnées du centre P_k^M , la couleur moyenne C_k^M et la taille S_k^M de la région associée ; et un vecteur d'état portant un vecteur de vitesse V_k^M , une

prédiction de changement de sa couleur ∇C_k^M et de sa taille ∇S_k^M .

La création du modèle se fait à l'initialisation sur la base de la détection des régions du premier plan.

4.2. Suivi des blobs :

Pour réaliser le suivi, l'image au temps t de la séquence est segmentée par un procédé de croissance de régions. Nous obtenons un ensemble de B composantes connexes. Un vecteur de caractéristique est associé à chaque composante j , de cet ensemble contenant les coordonnées de son centre P_j , sa couleur moyenne C_j , et sa taille S_j .

Ayant l'ensemble de B composantes connexes extraites de l'image courante et l'ensemble de T blobs modèles déjà construits, le suivi est effectué en cherchant des correspondances entre les composantes connexes de l'image courante et les blobs du modèle. Pour réaliser ce suivi, nous avons adapté à notre modèle l'approche proposée dans [10]. Un critère de minimisation d'une fonction de similarité permet une mise en correspondances entre l'ensemble des composantes connexes extraites et l'ensemble des blobs. Pour une composante connexe j , et un blob k , nous définissons la fonction de coût comme suit :

$$Cost(j, k) = \alpha_1 (||P_j - P_k^M|| / ||V_k^M||) + \alpha_2 (||C_j - C_k^M|| / \nabla C_k^M) + \alpha_3 (|S_j - S_k^M| / \nabla S_k^M) \quad (4)$$

Où : α_1, α_2 et α_3 sont des coefficients de pondérations.

4.3. Mis à jour des modèles

Les résultats de l'étape précédente sont utilisés pour mettre à jour chaque blob selon les cas suivants :

- Si une correspondance est établie entre un blob et une composante connexe : le vecteur de caractéristiques du blob ainsi que son vecteur d'état sont mis à jour.
- Si pour un blob, aucune correspondance n'a été établie, mais des correspondances ont été trouvées pour d'autres blobs appartenant à la même partie du corps que le blob considéré : alors ces correspondances sont utilisées pour estimer une vitesse moyenne, et selon cette vitesse la position de ce blob est mise à jour.
- Si pour un blob, aucune correspondance n'a été trouvée ni pour ce blob ni pour un blob appartenant à la même partie du corps : alors, une vitesse moyenne est estimée sur la base de différentes correspondances établies. Cette vitesse sert à la mise à jour de la position du blob.

5 Segmentation fond/parties du corps

La dernière étape de notre approche est de segmenter les pixels entre le fond et les parties du corps. Nous proposons d'affiner la segmentation de pixels entre le fond et les parties du corps en utilisant la minimisation d'énergie afin de réduire les faux positifs, et d'ajouter des pixels camouflés aux modèles. Au lieu d'effectuer cette étape sur tous les pixels de l'image, une partie de l'image est choisie ce qu'elle contient la boîte englobante de l'ensemble de blobs modèle dont les positions sont mises à jour via l'étape de suivi.

5.1. Les modèles probabilistes

L'appartenance d'un pixel à un blob k est modélisé par un modèle probabiliste représenté par une gaussienne spatio-couleur qui est mise à jour par l'algorithme de suivi présenté dans la section 4. Nous prenons un vecteur de caractéristiques à cinq dimensions pour décrire chaque pixel, $z_i = (x, y, l, a, b)$. La probabilité d'un pixel d'appartenir au blob k est donné par :

$$P(Z_i | Blob_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(Z_i - \mu_k)\Sigma^{-1}(Z_i - \mu_k)} \quad (5)$$

Où d est la dimension de modèle, dans notre cas $d = 5$. Si on suppose l'indépendance entre les composantes spatiales et celles colorimétriques, la matrice de covariance sera diagonale.

Pour la modélisation du fond, nous utilisons le mélange de gaussiennes présenté dans la section 3.1.

5.2. Segmentation par minimisation d'énergie

À tout instant t du temps, nous considérons les valeurs des pixels actuels, $z_{i,t}$, $i = 1, \dots, N$, où N est le nombre de pixels dans la zone de l'image où nous voulons appliquer la segmentation. Notons par $f_{i,t}$ l'étiquette du pixel i , $i = 1, \dots, N$; $f_{i,t}$ appartient à l'ensemble $\{0, \dots, T+1\}$, où T est le nombre de blobs, et le $T+1$ est le composant fond. Pour alléger, il sera fait abstraction de l'indice t .

Une fonction objective basée sur l'énergie peut être formulé sur les variables inconnues d'étiquetage de chaque pixel de la forme d'une fonction aléatoire du champ Markovien de premier ordre (MRF) :

$$\begin{aligned} E(f) &= E_{Data}(f) + \lambda E_{Smooth}(f) \\ &= \sum_{p \in P} D_p(f_p) + \lambda \sum_{(p,q) \in Ne} V_{p,q}(f_p, f_q) \quad (6) \end{aligned}$$

où Ne désigne le système de voisinage, un voisinage de 4-connectivité est utilisé dans notre cas.

Nous modélisons l'énergie mutuelle de lissage $E_{Smooth}(f)$ comme suit:

$$E_{smooth}(f) = \sum_{(p,q) \in Ne} \frac{1}{d(p,q)} e^{-\frac{(\|X_p - X_q\|)^2}{2\sigma^2}} \quad (7)$$

où X_p représente le vecteur de couleur du pixel p , σ est la moyenne des différences de couleur entre les pixels voisins dans l'image, et $d(p, q)$ est la distance entre deux pixels p et q .

Le terme d'énergie de données $E_{Data}(f)$ évalue la probabilité de chaque pixel de prendre une étiquette. Nous utilisons les modèles probabilistes expliqués dans la section précédente pour construire cette fonction : pour chaque pixel représenté par un vecteur saptio-couleur $z_i = (x, y, l, a, b)$, la probabilité de ce pixel d'appartenir au fond ou à un blob de parties du corps est :

$$p(z|l) = p(z|bg) + \sum_{k=1}^T p(z|Blob_k) \quad (8)$$

où l appartient à $\{0, \dots, T+1\}$.

Alors le terme d'énergie de données $E_{Data}(f)$ est défini comme suit :

$$E_{Data}(f) = \sum_{p \in P} p(z_p | f_p) \quad (9)$$

Enfin, la fonction d'énergie obtenue est ensuite minimisée par un algorithme de coupe de graphe par une méthode de développement alpha [13–15]. Et le modèle de fond est mis à jour pour les pixels d'arrière-plan avec un algorithme EM [11].

6 Résultats Expérimentaux

Afin d'évaluer notre approche, nous avons effectué plusieurs tests en utilisant un ensemble de séquences vidéo prises à l'intérieur par une seule caméra. Ces vidéos contiennent des différents scénarios actés par une seule personne. Elles ont été prises dans des différentes conditions de lumière et avec la présence d'ombres, de cas de camouflage, d'occultations et de changements d'éclairage.

La figure 2, montrent la performance de notre méthode sur plusieurs images d'une séquence vidéo. La première ligne montre la segmentation que nous avons obtenue lorsque la personne a commencé à marcher après l'étape de l'initialisation. La deuxième ligne illustre l'effet camouflage (Figure 2-ligne 2.b) en raison de similitude de couleurs entre le fond et le premier plan, on perd une partie de ce dernier. Notre méthode permet de corriger ce défaut : on arrive à extraire la partie camouflée dans les résultats de la segmentation finale (Figure 2-ligne 2.c). Avec la seule modélisation du fond, ce problème n'aurait pas été résolu. La troisième ligne montre nos résultats pour des postures compliquées, notre approche permet de suivre et extraire les parties du corps même lorsque la personne est par terre (Figure 2-ligne 3.c). En outre, la

silhouette extraite n'a pas été affectée par les ombres (Figure 2-ligne 1.b).

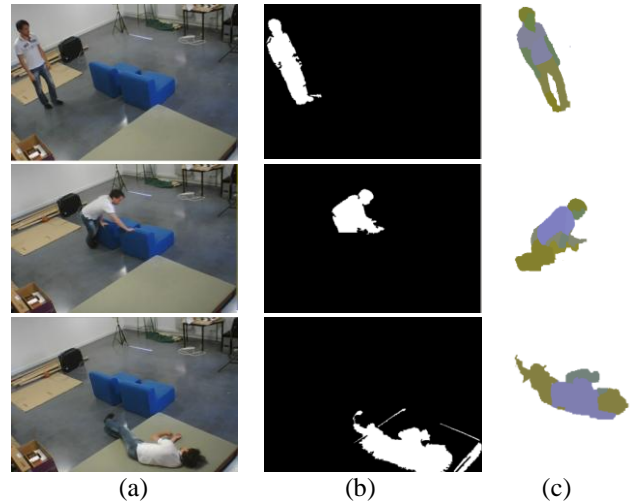


Figure 2 – Les résultats obtenus dans le cas de la marche ; (a) l'image de séquence vidéo. (b) Le premier plan résultat de notre approche MOG amélioré. (c) résultat final montrant la silhouette extraite représentée par les parties du corps.

En ce qui concerne le coût de calcul, notre système permet une vitesse moyenne de 15 images par seconde sans recours au calcul parallèle, pour la segmentation d'une séquence vidéo de 640x480 pixels en utilisant un processeur Intel Core i3-550. Le transfert de certaines parties du calcul sur GPU, permettrait d'accélérer le processus.

Le pourcentage de réussite de notre approche est donné dans le tableau (Table 1), nos résultats ont été comparés à des vérités terrain, et les pourcentages de réussite de la détection de silhouette en présence des différents problèmes posés par l'environnement non contrôlé ont été estimés.

	Ombre	Changement Rapide	Changement Globale	Camouflage
Réussite	82%	77%	80%	85%

Table 1 – Le pourcentage de réussite de notre approche.

7 Conclusion

Dans ce papier, nous avons présenté une approche pour l'extraction de silhouette dans une séquence vidéo et le suivi des différentes parties du corps, dans un environnement non-contrôlé. Notre approche s'est avérée efficace dans le cas des postures complexes. Tout d'abord une modélisation du fond par une MOG est utilisée pour extraire le premier plan, où une décision multi-niveaux en utilisant les informations de chrominance est utilisée pour

éliminer les effets de changement global et de présence d'ombres. Dans une étape d'initialisation, des modèles des parties du corps sont construits grâce à une analyse de contour et à une segmentation de région. Après l'initialisation, une approche basée sur la minimisation de coût est utilisée pour la mise en correspondance des modèles / composants connexes et le suivi. La dernière étape est la segmentation par la minimisation de l'énergie d'un champ de Markov contenant les modèles des parties du corps et le modèle du fond. La minimisation de l'énergie est appliquée sur une zone de l'image contenant les parties du corps.

Les résultats ont montré que notre approche est capable d'extraire la silhouette et de suivre les parties du corps pour des différentes postures et en présence d'ombres, de changements de lumière, de camouflage et d'occultations. Bien que notre méthode donne des bons résultats une implémentation en GPU, en cours de réalisation, permettra de prétendre au temps réel.

Pour l'amélioration de la qualité de résultats, un modèle du corps peut être utilisé pour améliorer la prédiction de l'emplacement des pièces manquantes. Cela peut conduire à de meilleurs résultats en cas de postures compliquées ou en présence d'occultations. En outre, un suivi de la silhouette entière peut être utilisé pour que le système soit applicable dans des scènes contenant plusieurs personnes, et pour relaxer le choix de seuil de suivi.

8 Références

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pp. 246-252.
- [2] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172-185, Jun. 2005.
- [3] A. Prati, I. Mikic, and R. Cucchiara, "Comparative evaluation of moving shadow detection algorithms," *on Empirical Evaluation*, no. 858, pp. 1-18, 2001.
- [4] T. Yu, C. Zhang, M. Cohen, Y. Rui, and Y. Wu, "Monocular Video Foreground/Background Segmentation by Tracking Spatial-Color Gaussian Mixture Models," *2007 IEEE Workshop on Motion and Video Computing (WMVC'07)*, pp. 5-5, Feb. 2007.
- [5] J. Gallego, "Enhanced bayesian foreground segmentation using brightness and color distortion region-based model for shadow removal," *Image Processing (ICIP), 2010 17th*, pp. 3449-3452, 2010.
- [6] C. R. Wren, a. Azarbayejani, T. Darrell, and a. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, Jul. 1997.
- [7] M. Isard and A. Blake, "CONDENSATION — Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5-28, 1998.
- [8] H. Nait-Charif, "Activity summarisation and fall detection in a supportive home environment," , *2004. ICPR 2004. Proceedings of the*, pp. 323-326 Vol.4, 2004.
- [9] C. Rougier and J. Meunier, "3D head trajectory using a single camera," *Image and Signal Processing*, vol. 3, no. 4, pp. 505-512, 2010.
- [10] O. Javed, "Tracking and object classification for automated surveillance," *History*, pp. 439-443, 2006.
- [11] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Proc. European Workshop Advanced Video Based Surveillance Systems*, 2001, vol. 1, no. 3, pp. 1-5.
- [12] N. Thome and S. Miguet, "A robust appearance model for tracking human motions," *IEEE DICTA-2005*, pp. 528-533, 2005.
- [13] Y. Boykov, O. Veksler, and R. Zabih, "Minimization via Graph Cuts," *Analysis*, vol. 23, no. 11, pp. 1-18, 2001.
- [14] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 2, pp. 147-59, Feb. 2004.
- [15] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 9, pp. 1124-37, Sep. 2004.