



HAL
open science

Méthode robuste pour la détection de chute dans un environnement non-contrôlé

Rada Deeb, Félix Lédée, Elodie Desserée, Saida Bouakaz

► **To cite this version:**

Rada Deeb, Félix Lédée, Elodie Desserée, Saida Bouakaz. Méthode robuste pour la détection de chute dans un environnement non-contrôlé. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00660974

HAL Id: hal-00660974

<https://hal.science/hal-00660974v1>

Submitted on 19 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Méthode robuste pour la détection de chute dans un environnement non-contrôlé

R. Deeb^{1,2} F.Lédée¹ E.Desserée^{1,2} S.Bouakaz^{1,2}

¹ Université de Lyon, CNRS

² Université Lyon1, LIRIS, UMR5205, F-69622, France

rada.deeb@liris.cnrs.fr

Résumé

L'allongement de la durée de vie a pour conséquence l'augmentation du nombre de séniors dans la population. La mutation sociétale ainsi que l'insuffisance des structures d'accueil poussent inévitablement vers le choix du maintien à domicile. Cette tendance ouvre la voie à l'émergence de nouveaux concepts, et on assiste donc à un développement des systèmes d'assistance de services à la personne (SAP). Ces dispositifs permettent à la personne âgée de vivre chez elle tout en assurant sa sécurité. Outre les systèmes d'alarmes bien connus (pendentif, montre, télécommande avec bouton d'alerte) de nouveaux systèmes plus sophistiqués se développent. Ils s'intéressent à la détection d'événements à risque dont la chute est l'un des plus dangereux. La recherche s'est récemment dirigée vers l'utilisation de cameras au lieu de capteurs portables ; ils sont en effet plus confortables et plus fiables. Dans cet article nous présentons une approche de détection automatique de chute, basée sur une méthode de suivi 3D de la tête. Cette méthode s'appuie sur un filtre à particules pour estimer la localisation et la vitesse de la tête afin de détecter la chute. Ce suivi est réalisé en utilisant la silhouette extraite en temps réel avec une approche robuste d'extraction de fond. Sur la base d'une modélisation du fond par un mélange de gaussienne, nous proposons de classifier les pixels entre « fond » et « premier plan » ou « plan d'intérêt ». L'originalité de notre méthode est l'utilisation de trois niveaux de résolution : local (pixel), semi-global (région), global (image). La décision est prise en utilisant deux images de fond, une calculée sur tout l'espace couleur et l'autre sur les composants de chrominance. Les expériences réalisées ont montré une bonne robustesse de notre méthode face aux changements brusques de luminance et à l'effet des ombres, ce qui permet son application dans un environnement non-contrôlé. De plus, notre approche de détection de chute semble indépendant du point de vue de la camera.

Mots Clefs

Détection de chute, système monoculaire, mélange de gaussiennes, extraction de fond, environnement non-contrôlé.

Abstract

Home care systems are becoming more and more necessary due to the growth of elderly population. Such systems are ought to ensure the independency as well as the safety of an elderly person. In addition to well-known alert techniques: (clocks, remote controls with alert button, etc.), more sophisticated systems have been developed to detect dangerous activities in general, and falls in particular known to be one of the most risky events that may happen to an old person living alone. Researches have been currently focused on using cameras which are more comfortable and more reliable than wearable and mobile sensors. In this article, we present an approach for automatic fall detection by a 3D head tracking. We use a particle filter in order to estimate head's location and speed both used to detect falls. The tracking method utilizes the silhouette extracted in real-time with a robust approach of Mixture of Gaussian background modeling. Our contribution is the classification of pixels between background and foreground via three levels of resolution: local (pixel), semi-global (region) and global (frame). In addition, decision is taken using two background images; the first is computed using all the components of color space, and the second is obtained using only the chromatic components. Our experiences have shown that the proposed approach is robust to sudden light changes and shadows. Therefore, it is suitable to be applied in a non-controlled environment. Moreover, our falls detection approach seems to be independent of camera's point of view.

Keywords

Falls detection, monocular system, mixture of Gaussian, background extraction, non-controlled environment.

1 Introduction

La mutation de la société, l'allongement de la durée de vie et l'augmentation de la population âgée, créent un besoin nouveau en termes de système d'assistance et d'aide au maintien à domicile des personnes âgées. Cela nécessite le développement de solutions innovantes pour aider ces personnes à vivre mieux, plus longtemps et en sécurité dans leur environnement. Disposer d'un système pour la détection automatique d'événements à risque qui peuvent arriver à une personne âgée vivant seule, devient un enjeu technologique et économique important. Parmi les risques évoqués ci-dessus, la chute est l'un des plus dangereux et les plus fréquents. Différents dispositifs existent dans le commerce comme des appareils de détection de mouvements rapides et soudain (exemple accéléromètres) ou des générateurs d'alarmes. Si ce genre de dispositifs s'avère pratique, l'utilisation de tels capteurs peut être dérangement voire invasif, et il y a un risque non négligeable pour que la personne oublie de le porter ou de l'activer. Pour pallier cet inconvénient, la recherche s'est récemment orientée vers l'utilisation de systèmes intégrés au lieu de vie des utilisateurs comme des caméras et des microphones. L'un des défis est de développer un système fiable et qui soit le mieux accepté et le moins intrusif possible tout en assurant une détection fiable des événements dangereux, en particulier la chute ou l'immobilité, et ceci en limitant l'intervention d'agent extérieur.

Dans ce contexte nous nous intéressons à la détection d'événements dangereux, dont le plus courant est la détection de chute, à partir de séquences vidéo (monoculaire). Généralement, les systèmes de détections d'événements à partir de vidéo se basent sur la détection de la silhouette. La silhouette extraite est utilisée pour obtenir des caractéristiques de mouvements utiles à la détection d'événements « à risque ». Dans cet article, nous présentons une approche robuste pour extraire la silhouette de la personne. Le suivi de cette silhouette nous permet d'identifier des cas de chute. L'extraction de la silhouette nécessite de la séparer du fond. Pour cela, nous proposons une adaptation de la méthode dite : Mélange de Gaussiennes « MOG » (Mixture Of Gaussian) pour tenir compte des changements brusques de luminosité et/ou présence de zone camouflage. Un suivi 3D de la trajectoire de la tête permet d'identifier les cas de chute et de distinguer des fausses alertes..

2 Etat de l'art

L'extraction de silhouette est généralement l'étape initiale dans les systèmes de détection des événements basés sur la vidéo. Les techniques de modélisation de fond sont utilisées pour pouvoir extraire les pixels d'intérêt dans un environnement particulier. Beaucoup de méthodes d'extraction de fond ont été développées afin de répondre aux nombreuses difficultés qui affectent l'efficacité de ces méthodes. Le bruit, les changements graduels ou rapides de lumière, les ombres et le camouflage sont des exemples de ces difficultés. Le modèle de fond le plus simple suppose que les valeurs d'intensité relatives à chaque pixel sont modélisées par une distribution uni-modale. Stauffer et Grimson [1] ont proposé de modéliser les valeurs du pixel avec un mélange de gaussiennes « MOG » (Mixture Of Gaussians) pour pouvoir l'appliquer dans un environnement complexe. Sur le même sujet, on peut également citer les travaux de Elgammal *et al.* [2] qui proposent l'utilisation d'une estimation par noyau pour modéliser le fond, ou encore les travaux de Kim *et al.* [3] qui préconisent l'utilisation d'un codage « codebook » comme un modèle de fond. Notons que cette dernière méthode nécessite une longue étape d'apprentissage.

Le mélange de gaussiennes « MOG » est l'une des méthodes les plus utilisées pour la modélisation de fond [4]. Cette méthode est en effet un bon compromis entre précision des résultats et temps de calcul. Elle permet également de s'affranchir de la longue étape d'apprentissage. MOG est généralement connue pour être robuste aux bruits et aux changements graduels de lumière. Malgré ces avantages, des problèmes subsistent ; ils sont en général liés aux changements globaux et rapides d'intensité lumineuse dans la scène, aux ombres et au camouflage. La littérature propose différentes études et solutions pour corriger ces problèmes. Dans le cas de systèmes multi-caméras, l'information de profondeur est utilisée conjointement aux couleurs pour avoir un système plus robuste aux changements d'intensité lumineuse [5]. Dans les systèmes monoculaires, des post-traitement sont généralement utilisés pour éliminer les effets dus aux ombres [6]. En revanche, le problème de changement rapide ou global de luminance dans la scène est généralement plus difficile à surmonter. Dans le travail de Harville [7], ces changements sont détectés en utilisant le nombre de pixels de premier plan ; ensuite, deux modèles MOG sont utilisés: le premier fait appel à des gaussiennes pré-calculées, et le second, temporel, prend en compte les changements globaux et rapides. L'avantage de cette approche est de conserver les informations déjà apprises pour adapter rapidement les modèles par l'estimation de nouvelles gaussiennes

construites durant la période de changement de lumière. Cependant, la silhouette est momentanément perdue. Le fait de perdre l'objet d'intérêt pendant un moment peut causer des échecs dans la détection d'événements. Pour pallier ce problème nous proposons un modèle de fond construit avec un MOG qui emprunte le principe proposé par Stauffer *et al.* [1]. Notre proposition consiste à classifier les pixels (fond et premier plan) selon trois niveaux en s'appuyant sur deux images de fond, la première calculée sur tout l'espace couleur et la seconde estimée sur les composants de chrominance.

Concernant l'identification de situations dangereuses, la littérature propose différentes méthodes dont la plupart n'utilise qu'une seule caméra. La méthode la plus simple utilise un calcul du ratio entre la largeur et la hauteur de la boîte englobant la silhouette de la personne dans l'image. Outre le problème bien connu des occultations, de telles méthodes tombent en défauts dans le cas de mouvements frontaux (par rapport à la caméra). Une réponse aux occultations est proposée dans [8] en positionnant la caméra au plafond. Si une telle solution répond aux problèmes des occultations, il est cependant connu qu'une telle projection fait perdre l'information de la vitesse verticale de la personne « et écrase la silhouette » ce qui restreint la détection.

Dans une étude biomédicale Wu [9] a montré qu'il était possible de différencier une chute d'une activité normale en regardant les vitesses verticales et horizontales du corps. Pour avoir un champ de vision plus large, la caméra est de préférence installée dans un coin de la pièce et en hauteur. On peut alors détecter les chutes à partir de la vitesse du sujet calculée dans l'image. Il faut souligner à ce propos que, plus le sujet est proche de la caméra plus cette vitesse est grande. Un tel phénomène rend difficile le choix d'un seuil de détection pour faire la différence entre une chute et une personne qui s'assoit brutalement. Pour différencier les différentes situations, en utilisant une seule caméra calibrée, Rougier *et al.* [10] ont montré qu'il est possible de calculer la trajectoire 3D de la tête avec une seule caméra calibrée. Cela permet d'obtenir les vitesses verticales et horizontales de la tête. Le fait de s'intéresser à la tête, partie du corps qui tombe le plus vite, évite des problèmes d'occultations. Cependant, l'approche de [10] dépend du point de vue de camera : les chutes qui arrivent contre ou vers la caméra sont plus difficiles à détecter. Pour apporter une réponse à ce problème, nous proposons de prendre en compte un second point d'appui pour déterminer la trajectoire de la tête. Ainsi, nous nous

sommes intéressés à la position des pieds du sujet. Cette idée découle du constat que cette position ne varie pas lors d'un dérapage de la personne mais elle change lors d'un déplacement. Ceci suppose que les pieds de la personne touchent le sol.

Concernant l'étape de suivi les filtres à particules basés sur l'algorithme Condensation [11] ont été utilisés pour suivre des personnes [8], et aussi pour suivre des parties du corps comme la main et la tête [11] [12]. Cette approche donne des bons résultats à condition d'avoir suffisamment de particules pour avoir un bon maximum de vraisemblance pour l'état estimé. Deutscher *et al.* [13] ont proposé de gérer les particules sous la forme de plusieurs couches. Cette méthode de recuit à base de filtre à particules permet d'obtenir un résultat plus précis avec le même nombre de particules.

3 Contexte et extraction de la silhouette

Le travail décrit dans cet article est développés dans le cadre d'un projet ANR-Tecsan : CIRDO. Le but est de développer un système d'assistance pour la détection automatique de situations à risque –en particulier les chutes- et ceci en s'appuyant sur des séquences vidéo, acquise dans un environnement de vie quotidienne : environnement non contrôlé. Pour des raisons de coût et de respect de l'intimité, le système n'utilise qu'une seule caméra par pièce. Cette caméra est installée en hauteur dans un coin de la pièce

3.1 Le modèle de mélange de gaussiennes

Pour un pixel p de l'image, on note par X_t^p le vecteur contenant la valeur de ce pixel au moment t . Cette valeur est prise dans l'espace couleur $l^*a^*b^*$, comme les composants de couleurs sont bien séparés de celui d'intensité lumineuse.

Nous modélisons les valeurs de chaque pixel p de la scène par k gaussiennes (généralement k compris entre 3 et 5) [1]. Donc, on peut écrire la probabilité d'observer la valeur du pixel p au temps t par l'équation :

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

Où : k désigne le nombre de gaussiennes ; $\omega_{i,t}^p$ le poids associé à la gaussienne, et η est la fonction gaussienne donnant la densité de probabilité :

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X_t - \mu)\Sigma^{-1}(X_t - \mu)} \quad (2)$$

Avec : $\mu_{i,t}^p$ est un vecteur qui contient les moyennes pour chaque canal et $\Sigma_{i,t}^p$ est la matrice de covariance. Pour faciliter le calcul, nous adoptons la même hypothèse que [1] : indépendance des canaux $1*a*b*$: $\Sigma_{i,t} = \sigma_{i,t}^2 I$.

3.2 Initialisation et mise à jour

MOG contient un historique des observations de la scène. Il est construit et mis à jour au cours du temps par un algorithme *Expectation-Maximisation* (EM) : chaque nouvelle valeur du pixel est comparée aux gaussiennes, puis l'adaptation des poids, des moyennes, et des variances de chaque gaussienne est faite par rapport aux correspondances trouvées. Dans notre travail nous suivons l'approche proposée dans [14] où les deux étapes (initialisation et mise à jour), sont séparées, et l'adaptation est faite en prenant en compte le nombre d'images traitées. En effet, pour éviter la disparition –même momentanée– de l'objet d'intérêt, nous suivons une approche sélective ; tous les pixels ne participent pas à l'adaptation des paramètres du modèle : les pixels appartenant a priori à l'objet d'intérêt sont exclus de l'étape d'adaptation.

3.3 Détection du fond

Pour séparer les pixels de fond des pixels de premier plan nous proposons une approche à trois niveaux : pixel, région, et image.

- Niveau pixel:

Dans un premier temps, la distance de Mahalanobis est calculée entre la valeur récente (temps $t+1$) du pixel p et toutes les gaussiennes relatives à ce pixel. Dans l'approche générale de MOG la distance est calculée sur tous les canaux de l'espace couleur. Dans notre approche, deux distances de Mahalanobis sont calculées, la première sur tous les canaux, et la deuxième en utilisant seulement les composantes de chrominance. 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 :

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

On note par \tilde{X}_{t+1}^p le sous vecteur de X_{t+1}^p limité aux deux composantes de chrominance. Et on considère qu'on a une correspondance sur les composants chrominance si:

$$\sqrt{(\tilde{X}_{t+1}^p - \tilde{\mu}_{i,t}^p)^T \cdot \tilde{\Sigma}_{i,t}^{p-1} \cdot (\tilde{X}_{t+1}^p - \tilde{\mu}_{i,t}^p)} < 2.5 * \sigma_{i,t} \quad (4)$$

Ici, le vecteur de moyenne et la matrice de covariance correspondent aux composantes de chrominance.

Après avoir trouvé les correspondances, une première classification des pixels entre fond et premier plan est effectuée. On considère qu'un pixel appartient au fond si la gaussienne correspondante est parmi les gaussiennes représentant le fond. Ces gaussiennes sont les B premières gaussiennes après la mise en ordre selon le ratio poids/variance:

$$B = \operatorname{argmin}_b (\Sigma_{i=1}^b \omega_i > T) \quad (5)$$

Où T est un seuil prédéfini.

Par conséquent, on obtient une image I_1 qui représente la classification des pixels de l'image précédente entre fond (en noir) et premier plan (en blanc). Cette première classification est faite en utilisant tous les canaux de l'espace couleur. De la même façon, et pour tous les pixels du premier plan qu'on a obtenus, on procède à une deuxième classification en utilisant seulement les composantes de chrominance, et on met les résultats dans l'image I_2 . Cette image I_2 va contenir seulement les régions de fond trouvées en utilisant les composants de chrominance et qui correspondent aux régions du premier plan dans I_1 .

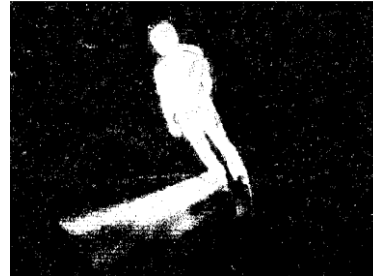


Figure 1 - I_1 l'image résultat d'extraction de fond obtenue en utilisant toutes les composantes de l'espace $1*a*b$. Le premier plan est en blanc et le fond est en noir.

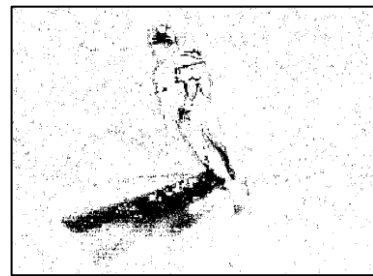


Figure 2 - I_2 l'image obtenue après une classification appliquée sur le premier plan de I_1 en n'utilisant que

les composantes de chrominance. Le premier plan est en blanc et le fond est en noir.

- Niveau région :

La décision au niveau région a pour but d'éliminer les régions du premier plan en I_1 liées aux ombres ou aux changements rapides de lumière. Cette élimination se base sur les régions de fond détectées avec les composants de chrominance (image I_2). Pour chaque région (composante connexe) en I_2 , on pose des critères sur la taille de cette région. En plus, on regarde si la suppression de cette région perturbe la connectivité du premier plan. Si une zone apparaît dans la région étiquetée fond, et si sa suppression n'affecte pas la connectivité du premier plan, on la considère comme une région liée aux changements de luminance. Ainsi on peut la considérer comme « candidate à la suppression » de I_1 . Cette suppression est confirmée ou infirmée si elle n'implique pas de modification de la taille de la silhouette. Cette vérification est faite par comparaison de boîte englobante estimée dans l'image précédente du flux vidéo. Avec cette approche on élimine les effets des ombres et de changements de lumière tout en conservant les parties du corps dont la couleur est proche des couleurs de fond.

- Niveau image :

La décision au troisième niveau : niveau image, sert à 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-là, on ne prend pas de décision au niveau région, mais on supprime de I_1 tous les pixels de fond en I_2 . Cela signifie qu'on ajoute les pixels de fond qu'on a trouvés en utilisant les composants de chrominance aux pixels de fond obtenue avec tous les composants de l'espace couleurs dans le but de réduire les effets de changements de lumière sur le premier plan.

4 La détection de chute

Un système monoculaire donne par défaut des informations 2D. Pour obtenir une information 3D, le système a besoin d'être calibré par rapport à des éléments de référence. Le calibrage nous permet de

calculer paramètres intrinsèques et extrinsèques de la caméra.

Pour la détection de chute on suit l'approche proposée par Rougier et *al.* [10]: Nous modélisons la tête par un ellipsoïde. Ce dernier est ensuite projeté dans le plan image en utilisant les paramètres intrinsèques et extrinsèques de la caméra. Le suivi du modèle dans le repère monde est ensuite réalisé à l'aide d'un filtre à particules. Dans le but d'avoir un suivi indépendant de point de vue de la caméra, nous proposons l'utilisation de la position des pieds dans le repère monde pour guider le suivi de la tête.

Un filtre à particules est particulièrement robuste aux brusques changements de trajectoire et donc adapté au suivi du personnage. Le but est de déterminer le vecteur correspondant au déplacement du modèle suivi. Chaque particule est représentée par un vecteur contenant les coordonnées (X, Y, Z) du centre de l'ellipsoïde dans le repère monde et son inclinaison par rapport aux axes X et Y :

$$V = \{X, Y, Z, \theta_x, \theta_y\}$$

Chaque échantillon est décrit par N vecteurs avec leur poids p respectifs à un temps t .

$$E = \{S_t^{(n)}, p_t^{(n)}, 1 \leq n \leq N\} \quad (6)$$

Le filtrage se déroule en quatre étapes :

- Sélection : N ellipsoïdes sont sélectionnés aléatoirement parmi les précédents.
- Prédiction : La nouvelle position des ellipsoïdes est ensuite déterminée à l'aide du modèle dynamique stochastique suivant :

$$S_t^{(n)} = AS_t^{(n)} + BSw_t^{(n)} \quad (7)$$

Avec A et B les composantes déterministe et stochastique du modèle et w un vecteur de variables aléatoires.

- Mesures : Un poids $p_t^{(n)}$ est assigné à chaque ellipsoïde.
- Estimation : Le vecteur est alors estimé à partir des ellipsoïdes pondérés :

$$E[S_t] = \sum_{n=1}^N p_t^{(n)} S_t^{(n)} \quad (8)$$

Pour améliorer les performances, on utilise un filtre à particules basé sur un schéma hiérarchique [13]. A chaque temps t , on estime la position quatre fois de suite en réduisant à chaque passage la composante stochastique du modèle. Lors du premier passage $B =$

$\{X, Y, Z, \theta_x, \theta_y\}$. A chaque passage cette composant est diminuée de la manière suivante : $B_l = B_{l-1}/2$.

Le calcul du poids de chaque particule se base sur 3 coefficients distincts.

- Coefficient de premier plan :

Le premier coefficient se calcule à partir de la projection de l'ellipsoïde dans le repère image. Il correspond à la distance moyenne entre les points de l'ellipse et les points du contour de la silhouette correspondants. Pour N_e points de l'ellipse, on trace la normale à l'ellipse passant par le point pour trouver l'intersection avec la silhouette du sujet.

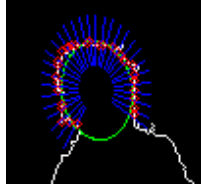


Figure 3 - Distance entre l'ellipse et la silhouette

On calcule ensuite la distance d entre le point de l'ellipse et le point détecté correspondant. Chaque distance calculée est normalisée.

$$C_f = \frac{1}{N_e} \sum_{n=1}^{N_e} \frac{D - d(n)}{D} \quad , C_f \in [0 \dots 1] \quad (9)$$

- Coefficient de couleurs :

Le second coefficient est basé sur le calcul de l'intersection entre l'histogramme de couleur correspondant à la zone rectangulaire contenant l'ellipse et un histogramme de couleur référence de la tête du sujet. L'histogramme de référence correspond à l'image de la tête lors de l'initialisation. Il est ensuite mis à jour lors du déplacement du sujet si le coefficient est suffisamment élevé. Les histogrammes sont calculés dans l'espace de couleur *RGB* avec une dimension $Nb = 8*8*8$.

$$C_c = \sum_{i=1}^{Nb} \min(H(i), H_{ref}(i)) \quad , C_c \in [0 \dots 1] \quad (10)$$

- Coefficient de corps :

Le troisième coefficient consiste à évaluer la distance entre la position des pieds détectés dans l'image puis projetée dans le repère monde et la position des pieds dans le repère monde estimée en fonction de l'axe principal de l'ellipsoïde. Le but de ce coefficient est de ne pas confondre un déplacement de la personne avec une position du corps incliné dans une certaine direction.

Premièrement, pour repérer la position des pieds du sujet dans l'image, on utilise une ellipse englobant la silhouette du sujet. On sélectionne ensuite le point de l'ellipse, le plus éloigné de la dernière position connue de la tête. Le système peut ainsi gérer le fait que les pieds ne correspondent pas forcément au point le plus bas de la silhouette, par exemple lors d'une chute en direction de la caméra.

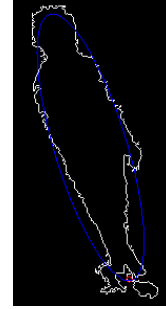


Figure 4 - Détection de la position des pieds

Dans un second temps, nous devons estimer l'emplacement des pieds dans le repère monde. Pour trouver cette position, on calcule l'intersection entre l'axe principal de l'ellipsoïde et le sol.

Pour finir on calcule la distance d entre ces deux positions. Chaque distance est normalisée en fonction de la moitié de l'axe principal de l'ellipse englobante.

$$C_d = \frac{D - d}{D} \quad C_d \in [0 \dots 1] \quad (11)$$

Le coefficient final, correspondant au poids de la particule, est calculé à partir de la formule suivante :

$$C_{final} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{C_f C_c C_d}{2\sigma^2}\right) \quad (12)$$

avec $\sigma = 0.25$

Lors de l'initialisation du suivi, la taille et les proportions de la tête de la personne sont connues et le sujet est supposé debout dans la pièce. Pour situer la position initiale de la tête dans le repère monde on projette la position des pieds du sujet du repère image dans le repère monde pour obtenir la position en (X, Y) . On utilise ensuite la taille de la personne pour trouver la position en Z . On obtient alors la trajectoire 3D de la tête.

La détection de chute se fait ensuite à partir du calcul de la vitesse verticale appuyée par la position en Z de la tête. Si la vitesse moyenne V_v durant un temps T descend au dessous d'un certain seuil et que la position du modèle est proche du sol, on considère que la personne est en danger.

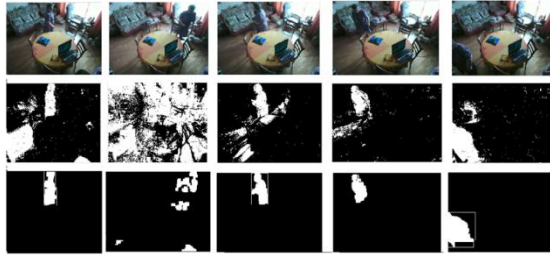


Figure 5 - Les résultats de notre approche d'extraction de silhouette appliquée sur une séquence de vidéo prise chez une personne âgée dans un environnement non-contrôlé.

5 Résultats Expérimentaux

5.1 Extraction de silhouette

Nous avons testé notre approche de modélisation de fond, construit en utilisant 3 gaussiennes, sur plusieurs vidéos prises dans un environnement non-contrôlé avec différentes conditions d'éclairages, en présence d'ombres et de changements rapides de lumière. Notre approche a pu extraire la silhouette en temps réel tout en étant robuste face aux changements de luminosité. En figure.5 on montre les résultats sur plusieurs images d'une vidéo pris chez une personne âgée, la deuxième ligne c'est les résultats obtenus avec une méthode MOG sans faire appel aux composantes de chrominance sur plusieurs niveaux. La troisième ligne montre les résultats obtenus en appliquant notre approche. Nos tests montrent que nous avons pu, grâce aux améliorations proposées réduire les ombres de l'image de premier plan (Figure.6-b), alors que dans l'approche originale les ombres sont considérées comme appartenant au premier plan et généralement attachées à la silhouette (Figure.6-a). De plus, on a pu éliminer les effets de changement rapides de lumière dans la scène. La figure.2 montre les images de premier plan obtenues lorsqu'une personne passe devant une fenêtre ce qui change le niveau d'éclairage dans la scène. Notre approche montre une bonne robustesse à ces changements et la silhouette est conservée dans l'image de premier plan (Figure.7-b). Dans l'approche MOG originale, plusieurs régions sont ajoutées au premier plan induisant la perte de la silhouette (Figure.7-a).



Figure 6 – les résultats d'extraction de fond en présence de l'ombre. (a) premier plan obtenu avec MOG originale. (b) premier plan obtenue avec notre approche.

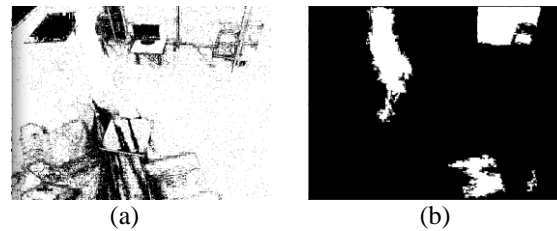


Figure 7 – les résultats d'extraction de fond en présence de changement rapide de lumière. (a) premier plan obtenu avec MOG originale. (b) premier plan obtenue avec notre approche.

5.2 Détection de chute

Nous avons testé l'approche de détection de chute sur une base de données de vidéos qui contiennent différents cas de chute comme : glisser, tomber en avant ou en arrière par rapport à la caméra. Les vidéos ont été prises dans un environnement complexe en utilisant une seule caméra fixée au mur en hauteur. En utilisant notre approche il était possible de détecter les cas de chute quelle que soit la position de la caméra. La figure.8 montre la détection d'un cas de chute en utilisant la position en Z de la tête ainsi que sa vitesse verticale. Figure.9 montre l'amélioration apportée à l'estimation de position 3D de la tête par l'utilisation de position 3D de pieds.

6 Conclusion

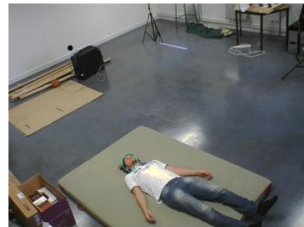
Dans cet article nous avons présenté une approche de détection de chute, et une méthode d'extraction de fond en temps réel dans un environnement non-contrôlé. Concernant la détection de fond, nous avons utilisé un mélange de gaussiennes pour modéliser les pixels de la scène. Nous avons classifié les pixels de l'image entre fond et premier plan selon trois niveaux (pixel, région, et image) en s'appuyant sur les parties de fond trouvées sur les composantes de chrominance sans prendre l'intensité lumineuse en compte. Cette approche a prouvé sa robustesse face aux changements rapides et globaux de lumière. De

plus, aucun post-traitement pour enlever les régions d'ombre ne s'est avéré nécessaire.

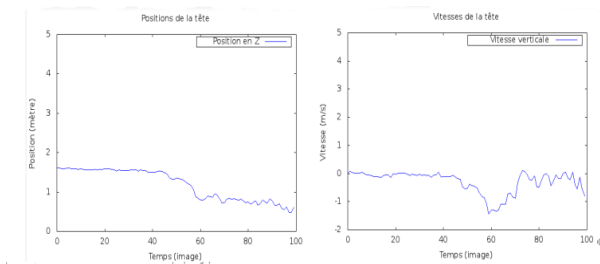
Concernant la détection de chute, nous avons présenté une approche indépendante de la position de la caméra dans la scène. Nous avons utilisé la silhouette extraite dans un suivi 3D de la trajectoire de la tête basé sur un filtre à particules. Le filtre à particules est guidé par la localisation des pieds. Ce suivi nous permet de détecter les cas de chute en s'appuyant sur la position et la vitesse verticale de la tête. Notre approche a montré son efficacité pour la détection des chutes et en particulier pour les cas ambigus des chutes frontales.

Plusieurs pistes d'amélioration de la méthode proposées sont d'ores et déjà envisagées :

- éliminer les parties de silhouette perdues à cause du phénomène de camoufflage en utilisant des informations de haut niveau.
 - Améliorer la classification de pixels entre fond et silhouette en utilisant un modèle de l'apparence générale de la personne ou des parties du corps.
- S'affranchir de la contrainte imposant que les pieds soient toujours visibles et attachés au sol en identifiant et en suivant à la fois le mouvement de la tête et de pieds.



(a)



(b)

(c)

Figure 8 – Exemple de la détection de chute. (a) l'image de séquence vidéo où on a un cas de chute. (b) position de la tête en Z. (c) vitesse verticale de tête.

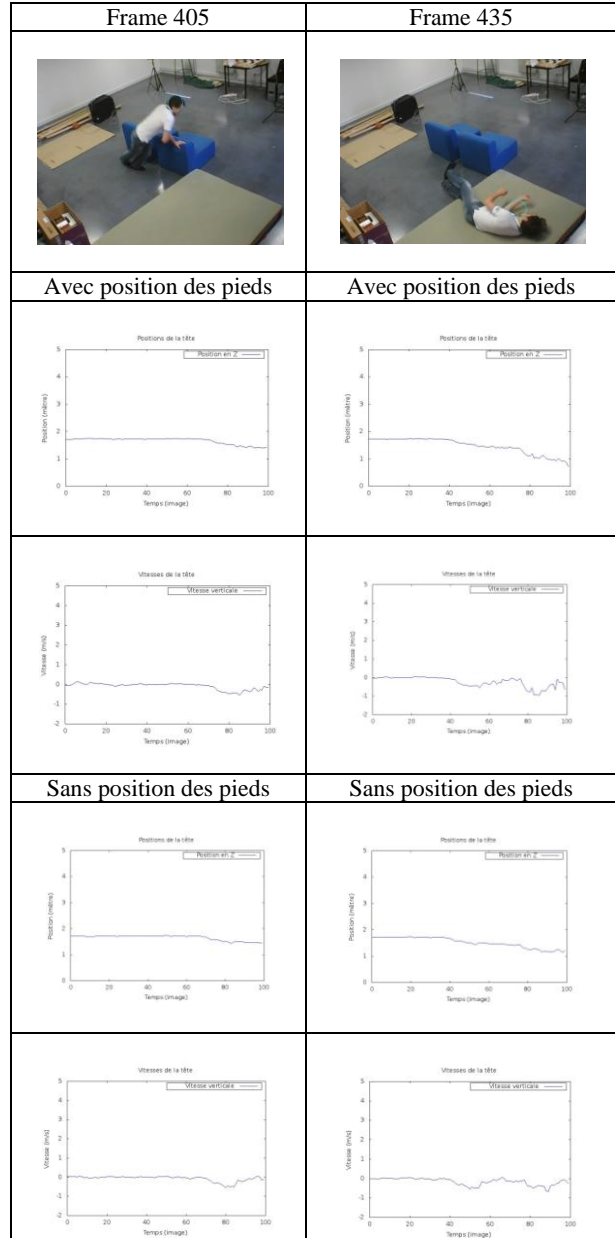


Figure 9 – Exemple de la détection de chute. On voit la position de la tête en Z et sa vitesse verticale estimées avec et sans la position 3D de pieds. On note sur la dernière image l'apport de l'utilisation de pieds dans l'estimation de la position de la tête.

Bibliographies

- [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] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," *Computer*, pp. 751-767, 2000.
- [3] 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.
- [4] T. Bouwmans, F. El Baf, and B. Vachon, "Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219-237, Jan. 2010.
- [5] M. Harville, G. Gordon, and J. Woodfill, "Foreground segmentation using adaptive mixture models in color and depth," *Proceedings IEEE Workshop on Detection and Recognition of Events in Video, 2001*, pp. 3-11.
- [6] A. Prati, I. Mikic, and R. Cucchiara, "Comparative evaluation of moving shadow detection algorithms," *on Empirical Evaluation*, no. 858, pp. 1-18, 2001.
- [7] M. Harville, "A Framework for High-Level Feedback to Adaptive, Per-Pixel, Mixture-of-Gaussian Background Models," *Framework*, 2002.
- [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] G. Wu, "Distinguishing fall activities from normal activities by velocity characteristics.," *Journal of biomechanics*, vol. 33, no. 11, pp. 1497-500, Nov. 2000.
- [10] C. Rougier and J. Meunier, "3D head trajectory using a single camera," *Image and Signal Processing*, vol. 3, no. 4, pp. 505-512, 2010.
- [11] 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.
- [12] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99-110, Jan. 2003.
- [13] J. Deutscher, a Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 2, pp. 126-133.
- [14] 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.