

La détection et le suivi de régions d'intérêt pour la vidéosurveillance de la prise de médicaments

HUNG HUYNH-HUU¹, JEAN MEUNIER², JEAN SEQUEIRA¹, MARC DANIEL¹

¹ Laboratoire des Sciences de l'Information et des Systèmes (LSIS)
Campus de Luminy, case postale 925, 13288 Marseille cedex 9, France

² Laboratoire Traitement d'Images, DIRO, Université de Montréal,
CP 6128, Succursale Centre-ville, 2920 Chemin de la tour, Montréal, Québec, Canada, H3C 3J7

¹ Hung.Huynh-Huu@univmed.fr, Marc.Daniel@univmed.fr, Jean.Sequeira@univmed.fr
² Meunier@iro.umontreal.ca

Résumé. Dans le contexte de l'accroissement du nombre de seniors dans les pays développés et des efforts fournis dans les services de soin à la maison, nous avons développé un algorithme nous permettant de détecter et de suivre en temps réel des objets mobiles, dans une scène correspondant à une personne prenant des médicaments, sous la vue d'une webcam fixe. Les expériences sont présentées et les améliorations dans la détection de bouche, la détection de régions des doigts, le suivi de mains et de visage, le traitement d'occultation, de notre algorithme sont discutées

Abstract. In the context of the growing proportion of seniors in the developed countries and the efforts provided in home care services, we have developed an algorithm that allows us to detect and track real-time mobile objects in the scene of medication intake. Experiments are proposed and improvements in the mouth detection, the detection of regions of the fingers, the hands tracking, the occultation handling, of our algorithm are discussed.

1 Introduction

Récemment, des systèmes ont été développés pour la surveillance de la prise de médicaments [2,14]. Cet article propose une amélioration du travail présenté dans [14]. Il fait partie d'un plus grand projet, la vidéosurveillance intelligente des personnes âgées pour leur soutien à domicile ou au centre de soin

L'algorithme général pour la surveillance de la prise de médicaments est décrit par les étapes suivantes: (a) construire l'arrière plan (background), (b) détecter les objets mobiles, (c) segmenter la couleur de peau, (d) détecter le visage et autres régions d'intérêt dans les blobs (régions) de couleur de peau, (e) suivre les objets, (f) représenter et reconnaître l'activité.

L'algorithme se compose de deux parties principales : (1) les traitements de bas niveau (a-e) et (2) la reconnaissance de l'activité de prise de médicaments (f), qui est un traitement de haut niveau.

Dans le cadre de cet article, nous présentons les traitements de bas niveau : segmentation de couleur de peau, détection des régions d'intérêt dans les blobs de couleur de peau, détection des flacons, suivi des objets mobiles et traitement d'occultation. La représentation et la reconnaissance d'activité fera l'objet d'une recherche ultérieure.

2 Détection

2.1 Détection des objets en mouvement

La détection des objets mobiles est faite par la soustraction de l'arrière-plan dans l'espace de couleur YCrCb. Nous calculons au préalable (avant qu'une

personne soit présente dans la scène) l'image moyenne des images pour obtenir l'arrière-plan ainsi que l'écart-type pour chaque pixel et pour chaque canal de couleur. Pour chaque nouvelle image, nous calculons la différence absolue entre les pixels de l'image moyenne (arrière-plan) et ceux de la nouvelle image. Le pixel est marqué comme pixel de l'avant-plan si la différence dépasse un certain nombre (par exemple 2.0) d'écart-type. Autrement, le pixel est marqué comme pixel de l'arrière-plan.

Après avoir détecté des objets mobiles, il est nécessaire d'éliminer les ombres. Les pixels qui font partie de l'ombre ont la même couleur que ceux de l'arrière-plan mais une intensité différente. Cette caractéristique est utilisée pour éliminer les ombres. Ensuite, des opérations morphologiques sont appliquées pour enlever le bruit.

2.2 Segmentation de couleur peau

La détection de couleur de peau humaine est largement utilisée et s'est avérée être une caractéristique efficace dans plusieurs applications, parmi lesquelles, la détection du visage et le suivi des mains. Bien que la couleur de la peau puisse largement varier, les récentes recherches montrent que la différence principale est plutôt dans l'intensité que dans la chrominance [10]. Plusieurs espaces de couleur sont utilisés pour étiqueter les pixels comme pixels de couleur de peau : RGB, RGB normalisé, HSV (ou HSI), YCrCb, YIQ,... De plus, des méthodes sont proposées pour construire un modèle de couleur de peau : méthode utilisant la tonalité de pixel, méthode basée sur l'histogramme [8], méthode utilisant une fonction gaussienne [3], ...

Dans l'application de surveillance de prise de médicaments, il existe souvent trois blobs de couleur de peau. Donc, une méthode simple, efficace, robuste et temps réel pour segmenter les blobs de couleur de peau est le bon choix pour notre travail. Parmi les approches utilisées pour la segmentation de couleur de peau, l'approche présentée dans [13], utilise la combinaison de deux espaces de couleur HSV et YCrCb. Celle-ci convient bien pour notre travail, avec un haut taux de précision [10].

2.3 Détection des régions d'intérêt du visage

Après avoir détecté les régions de peau dans le cadre de vidéo, il est aussi nécessaire de distinguer le visage et les mains, détecter la bouche et de localiser les régions de mains.

Plusieurs recherches se font sur la détection du visage, utilisant l'information de couleur de peau et en supposant qu'il existe seulement un blob de couleur de peau dans l'image. Avec cette supposition, le blob de couleur de peau est converti en niveaux de gris, et en binaire après. Les yeux et la bouche sont ensuite détectés en projetant l'image binaire sur l'axe vertical, sur l'axe horizontal en cherchant les minima locaux [4]. Il est aussi possible de réaliser cette détection à l'aide de formes paramétrées [4,9]. Dans le cas où il existe plusieurs blobs de couleur de peau, une contrainte morphologique, telle la proportion entre la largeur et la hauteur de blob, est utilisée pour identifier le blob représentant le visage. Les approches par apprentissage [11,12] sont utilisées aussi pour la détection du visage : réseaux neuronaux, AdaBoost, machine à vecteurs de support (SVM),... Parmi ces approches par apprentissage, l'AdaBoost est évaluée comme une approche rapide, robuste et efficace [16].

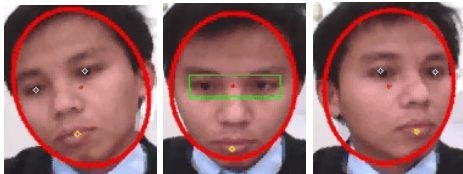


Figure 1. Détection des régions d'intérêt : les yeux et la bouche

La détection des régions d'intérêt dans notre travail inclut la détection des yeux et de la bouche (figure 1). La détection des yeux permet de différencier le visage et les mains (régions de même couleur de peau) : la détection de la bouche sert dans la phase de reconnaissance d'activité de prise de médicaments. Il n'est pas raisonnable d'appliquer directement les approches existantes pour notre système, pour les raisons principales suivantes :

- la supposition qu'il existe seulement un blob de couleur de peau (le visage) n'est pas applicable, car il existe souvent trois blobs de couleur de peau.
- la détection des régions d'intérêt par la projection sur l'axe horizontal, et l'axe vertical n'est pas possible, car il existe souvent des occultations.

- L'apprentissage pour détecter le visage entier avec fiabilité nécessiterait un nombre trop important d'échantillons.
- il est trop délicat de paramétrer les différentes caractéristiques de visage (approche utilisée pour les formes déformables), car l'image capturée par la caméra est petite et de basse qualité.
- la détection des régions d'intérêt par l'AdaBoost présente de nombreux cas d'échec décrits ci-après.

Il est donc difficile d'appliquer les approches existantes, et par conséquent, nous proposons une approche hybride pour notre travail.

La détection des yeux (deux yeux ensemble, œil gauche, ou œil droit) et de la bouche est faite par la méthode d'AdaBoost. C'est une méthode par apprentissage, présentée dans [15]. La détection des régions d'intérêt par l'AdaBoost est rapide mais elle commet des erreurs : c'est le cas lorsque la bouche est reconnue comme l'œil, et inversement l'œil est reconnu comme la bouche. On peut détecter aussi les yeux et la bouche dans l'espace de couleur YCrCb [7]. Cependant cette approche nécessite beaucoup de calculs pour mettre en évidence des régions d'intérêt du visage, surtout dans notre travail, car il existe souvent trois blobs de couleur de peau.

L'approche hybride que nous proposons pour surmonter ces désavantages se compose de deux étapes : (1) la détection des yeux est faite par l'AdaBoost, et (2) la détection de la bouche dans la région sous les yeux est faite par l'AdaBoost ou par l'approche présentée dans [7]. La distance entre les yeux et la distance entre la bouche et les yeux sont utilisées pour éliminer les fautes [13]. La taille des blobs de couleur de peau doit être normalisée dans le but de donner un bon résultat.

Dans notre travail, nous proposons le concept de '*candidature du visage*' pour décrire un blob de couleur de peau, dans laquelle les yeux (un ou deux) sont bien détectés, avec la bouche détectée ou non. Après avoir détecté les différents blobs de couleur de peau, chaque blob dans l'image actuelle est vérifié et étiqueté pour savoir s'il est une candidature du visage ou non, ce travail ayant pour but de faciliter le calcul de matrices de similarités entre les blobs. De plus, la distance entre la bouche et les yeux est utilisée pour éliminer le cou (aussi de couleur de peau) au cours de la détection du visage.

2.4 Détection des mains

La détection des mains peut-être faite facilement en supposant que les mains portent des gants colorés. D'autres approches détectent les mains sans les bras, ou sans visage dans l'image. Dans notre contexte, nous détectons la prise de médicaments par des personnes âgées, qui portent des vêtements à manches longues ou courtes. Donc ces approches sont inadaptées à notre travail. De plus, la détection des mains doit être robuste dans le cas où les personnes portent des vêtements à manches longues ou courtes, et la région des doigts doit être bien détectée, car elle sert à la reconnaissance de la prise de médicaments.

La détection des régions correspondant aux mains est faite en prenant l'esprit de la méthode présentée dans [2], avec des modifications simplificatrices et plus rapides qui améliorent le résultat. Dans [2], les auteurs utilisent une contrainte géométrique sur la position de la main dans les blobs de couleur de peau (autres que le visage) et une mesure du gradient moyen dans une région circulaire qu'ils cherchent à maximiser pour obtenir la position finale de la main. Le résultat de la détection n'est pas trop précis - les régions des doigts sont mal localisées.

Pour détecter les régions des mains, les blobs de couleur de peau sont convertis en niveaux de gris. On voit que dans la région de la main la variation des niveaux de gris est grande, donc cet élément sert à la segmentation de cette région. La variance des niveaux de gris est calculée pour chaque pixel, à l'intérieur d'un masque 3x3. Après avoir calculé la variance de gris, la région de la main est identifiée en déplaçant une fenêtre carrée, dont le côté a la largeur du bras, le long du bras pour trouver la fenêtre ayant la variance moyenne maximale (voir figure 2).

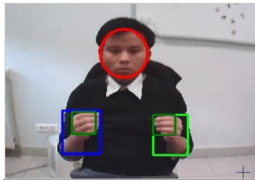


Figure 2. Résultat de la détection des régions des mains

2.5 Détection des flacons

La détection des flacons peut être faite en utilisant l'information de forme et la carte des bords de Canny [2], ou en reconnaissant les lettres assignées aux flacons. Pour simplifier le travail, dans notre système, les flacons sont colorés et la détection des flacons est faite dans l'espace de couleur YCrCb. De cette façon, les flacons sont détectés même s'ils sont tournés au moment de la détection.



Figure 3. Détection des flacons

2.6 Représentation des blobs de couleur peau, des flacons

Chaque blob de couleur de peau est représenté par une structure de données, avec les informations suivantes : le rectangle englobant, le centre du blob, la couleur, la surface, les paramètres du filtre de Kalman et d'autres informations. L'information est la même pour représenter les flacons à l'exception du filtre de Kalman.

3 Suivi

3.1 Algorithme de suivi des objets

Plusieurs recherches ont été entreprises sur le suivi d'objets. L'occultation n'y est pas forcément traitée. D'autre part, le suivi des objets rigides est plus facile que celui des objets non rigides. Dans la phase de suivi, les objets sont souvent identifiés avant et après l'occultation. Il nous est cependant nécessaire d'identifier aussi les objets pendant l'occultation, ce qui est beaucoup plus délicat, en particulier du fait que les objets ne sont pas rigides.

L'algorithme de suivi des régions d'intérêt présenté dans cette partie s'inspire du principe présenté dans [1,5,6]. Pour chaque image, on calcule la similarité entre les blobs courants (les blobs actuels) et les blobs dans l'image précédente (les blobs passés). La similarité est une fonction du déplacement, du rapport entre les surfaces et d'autres caractéristiques. La similarité entre les blobs est présentée dans une matrice. Si deux blobs ont la même similarité, on compare les autres caractéristiques pour mettre à jour la similarité. Ainsi, l'exemple présenté dans la table 1, on peut associer les blobs actuels 1, 2 et 3 aux blobs passés 2, 1 et 3 respectivement.

Tab 1. Exemple de matrice de similarité

	blob-pré-1	blob-pré-2	blob-pré-3
blob-act-1	x	max1	x
blob-act-2	max2	x	x
blob-act-3	x	x	max3

3.2 Traitement de l'occultation

Le traitement de l'occultation main-visage est un problème difficile, objet de plusieurs recherches, surtout dans les applications de reconnaissance des langages des signes. Le traitement est fait par soustraction du blob actuel avec des formes prédéfinies de mains en supposant que le visage ne bouge pas au moment de l'occultation. Dans notre travail, les occultations se passent souvent entre les blobs de couleur de peau : c'est l'occultation entre la main et le visage, ou entre les mains. L'occultation se passe aussi entre les flacons et la main. La détection et la reconnaissance des objets avant et après l'occultation ne suffisent pas pour notre travail. Il faut de plus bien détecter et reconnaître des objets au cours de l'occultation, surtout dans la phase de prise de médicaments. Dans notre travail, le traitement d'occultation est fait avec l'image en gris, les blobs de couleur de peau ayant été convertis en niveaux de gris.

Pour résoudre ce problème, nous proposons le concept de '*blob caché*' pour exprimer un blob disparu lorsqu'un blob chevauche un autre blob car il n'existe alors plus de blob correspondant. Dans ce cas, on utilise un filtre de Kalman pour prédire la position du blob caché. Pour cela, on déplace le blob caché, visible dans l'image précédente, autour de la position prédite dans l'image courante en cherchant la position où la distance

en gris est minimale avec le blob actuel. Cette approche nous permet un traitement en temps réel en supposant que la forme des blobs, surtout la forme de mains ne change pas beaucoup au moment de l'occultation.

Quant à l'occultation des flacons, lorsqu'un flacon est caché, il est supposé caché par le blob ayant la distance minimale avec lui, ce qui correspond en général aux blobs associés aux mains.

4 Résultats

Nous avons appliqué la démarche décrite dans les paragraphes précédents pour la détection de la couleur de peau et de bons résultats ont été obtenus. Des améliorations dans la détection des régions des mains sont présentées aussi dans cet article. La détection et le suivi présentés sont robustes, le traitement temps réel. La méthode est expérimentée avec une webcam utilisant une résolution de 320x240 pixels. Quatre séquences vidéo, d'une longueur totale de 5 minutes, similaires à celles de la figure 4, et mettant en scène deux personnes distinctes et des éclairages différents, ont été testées. Globalement l'algorithme a détecté et suivi correctement les régions d'intérêt 99,7% du temps.

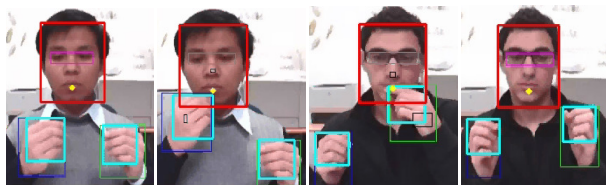


Figure 4. Résultat de détection et de suivi

5 Conclusions

L'algorithme proposé – combinaison de la méthode AdaBoost avec la méthode utilisant les espaces de couleurs YCrCb, HSV – présente différents avantages en comparaison avec d'autres approches existantes : rapidité, robustesse au changement d'éclairage, et résultats en temps réel. Cependant, il présente encore certaines limites : par exemple, au moment de l'occultation, l'algorithme ne fonctionne pas toujours bien si la forme de la main change beaucoup et rapidement. De plus l'algorithme demande beaucoup de puissance de machine au moment de l'occultation car il fait alors beaucoup de calculs pour trouver la position précise du blob. Dans le futur, nous proposerons un algorithme de reconnaissance de gestes (prise de médicaments) basé sur des observations de distances entre régions d'intérêt qui seront soumises à des modèles de Markov cachés (HMM).

6 Références

[1] Amer A. Voting-based Simultaneous Tracking of Multiple Video Objects. *IEEE Transactions on Circuits and Systems for Video Technology* 2005;15(11):1448-62.
 [2] Batz D, Batz M, Lobo NdV, Shah M. A Computer Vision System for Monitoring Medication Intake.

2nd Canadian Conference on Computer and Robot Vision 2005:362-9.
 [3] Bernd M, Mathias W. Segmentation and Tracking of Facial Regions in Color Image Sequences. *SPIE Visual Communications and Image Processing* 2000;4067:731-40.
 [4] Chai D, Ngan KN. Face Segmentation Using Skin Color Map in Videophone Applications. *IEEE Transactions on Circuits and Systems for Video Technology* 1999;9(4):551-64.
 [5] Dedeoglu Y. Moving Object Detection, Tracking and Classification for Smart Video Surveillance; 2004.
 [6] Fuentes LM, A.Velastin S. People Tracking in Surveillance Applications. *Image and Vision Computing, Image and Vision Computing* 2006;24(11):1165-71.
 [7] Hsu R-L, Abdel-Mottaleb M, K.Jain A. Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002;24(5):696 - 706.
 [8] J.Swain M, H.Ballard D. Color Indexing. *International Journal of Computer Vision* 1991;7(1):11-32.
 [9] Lai W-H, Li C-T. Detecting Faces in Color Images. Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK 2005.
 [10] Nallaperumal K, Ravi S, Babu CNK, Selvakumar RK, Fred AL, Christopher S, Vinsley SS. Skin Detection Using Color Pixel Classification with Application to Face Detection: A Comparative Study. *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications* 2007;3:436-41.
 [11] Osuna E, Freund R, Girosi F. Training Support Vector Machines: an Application to Face Detection. *IEEE Conference on Computer Vision and Pattern Recognition* 1997:130-6.
 [12] Rowley HA, Baluja S, Kanade T. Neural Network-Based Face Detection. *IEEE Transactions On Pattern Analysis and Machine intelligence* 1998;20:23-38.
 [13] Singh SK, Chauhan DS, Vatsa M, Singh R. A Robust Skin Color Based Face Detection Algorithm. *Tamkang Journal of Science and Engineering* 2003;6(4):227-34.
 [14] Valin M, Meunier J, St-Arnaud A, Rousseau J. Video Surveillance of Medication Intake. *International Conference: IEEE Engineering in Medicine and Biology Society* 2006.
 [15] Viola P, Jones M. Rapid Object Detection Using a Boosted Cascade of Simple Features. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2001;1:511-8.
 [16] Viola P, Jones M. Robust Real-Time Object Detection. *International Journal of Computer Vision* 2001.