

# Une méthode génétiquement optimisée d'extraction de personnes par coupe de graphe de superpixels pour l'environnement embarqué

Christophe CONIGLIO<sup>1</sup>, Cyril MEURIE<sup>1</sup>, Olivier LÉZORAY<sup>2</sup>, Marion BERBINEAU<sup>1</sup>

<sup>1</sup>Univ Lille Nord de France, F-59000 Lille, IFSTTAR, COSYS, LEOST, F- 59650 Villeneuve d'Ascq

<sup>2</sup> Normandie Univ., UNICAEN, ENSICAEN, GREYC UMR CNRS 6072, Caen  
christophe.coniglio@ifsttar.fr, cyril.meurie@ifsttar.fr  
olivier.lezoray@unicaen.fr, marion.berbineau@ifsttar.fr

**Résumé** – Dans cet article, nous présentons une nouvelle méthode d'extraction de personnes dans des environnements transport. De nombreuses méthodes de soustraction de fond existent dans la littérature, mais n'offrent pas toujours de résultats satisfaisants en conditions réelles et dans un contexte difficile contenant de nombreux verrous scientifiques tels que les changements rapides de luminosité, le bruit, les ombres, l'arrière-plan défilant à travers des parois transparentes. La méthode que nous présentons permet de pallier ces difficultés et est basée sur une segmentation de l'image en superpixels, associée à une coupe de graphe initialisée par une détection de premier plan, paramétrée à l'aide d'un algorithme génétique. La stratégie proposée est alors composée de quatre blocs majeurs: un bloc de pré-traitement qui utilise des filtres et des invariants colorimétriques afin de limiter la présence d'artefacts dans les images; Un bloc de détection de premier plan qui permet de localiser grossièrement la personne dans l'image; Un bloc de post-traitement qui élimine les divers artefacts dont les ombres; Un bloc d'extraction de personnes qui sur-segmente l'image en superpixels puis effectue une classification binaire par une coupe de graphe afin d'extraire précisément la personne de l'arrière-plan. Les tests sont réalisés sur une base de données issue du projet européen BOSS et sont optimisés à l'aide d'un algorithme génétique.

**Abstract** – In this paper, we present a new method for people extraction in complex transport environments. Many background subtraction methods exist in the literature but don't give satisfactory results on complex images acquired in moving transport systems that include several locks such as fast brightness changes, noise, shadow, scrolling background, etc. To tackle this problem, a new method for people extraction is proposed. It is based on an image superpixel segmentation coupled with graph cut binary clustering. The proposed strategy is composed of four majors blocks: A pre-processing block that uses filters and colorimetric invariants to limit the presence of artifacts in images; A foreground detection block that enables to locate moving people in images; A post-treatment block that removes shadow regions of no-interest; A people extraction block that segments the image into superpixels and performs a graph cut binary clustering to precisely extract people. Tests are realized on a real database of the BOSS European project and optimized with a genetic algorithm.

## 1 Introduction

De nos jours, les systèmes de vidéo surveillance sont très répandus dans des environnements de transport. Ils peuvent être utilisés pour détecter des situations potentiellement dangereuses, suivre des objets en mouvement ou des personnes afin d'améliorer leur confort et leur sécurité ainsi que celle de l'infrastructure. Il y a donc beaucoup d'attente sur le développement d'applications telles que la détection, le suivi, la réidentification de personnes et la reconnaissance de l'action [11]. La caractéristique commune de ces recherches est la détection d'objets en mouvement. C'est pourquoi de nombreux algorithmes de soustraction de fond (extraction d'objets par comparaison de l'image courante avec une image d'arrière-plan de référence) peuvent être trouvés dans la littérature (plus de détail dans [9]). La littérature est vaste, mais les méthodes existantes n'offrent généralement de bons résultats que dans le cas d'environnements contrôlés et non le cas d'environnements complexes (par exemple, en environnement embarqués). Dans le cadre de notre application globale d'optimisation de la ges-

tion des flux de passagers en contexte multimodal, l'extraction de la personne est nécessaire pour suivre les trajectoires et ré-identifier (anonymement) ces même personnes lors d'un passage devant une autre caméra. Cette étape doit donc être aussi précise que possible, même si la scène contient de nombreux verrous scientifiques (tels que les changements rapides de luminosité, les ombres, ou encore l'arrière plan défilant derrière des parois transparentes, etc.). En effet, ces difficultés sont généralement présentes lorsque les acquisitions d'images sont réalisées en conditions réelles et en embarqué. Dans ce contexte, nous proposons ici une méthode robuste d'extraction de personnes.

## 2 Méthode proposée

La méthode que nous proposons va du pré-traitement de l'image acquise à l'extraction précise de la personne. Elle est composée de plusieurs étapes décrites ci-dessous et d'une optimisation globale par une métaheuristique, comme illustré sur la Figure

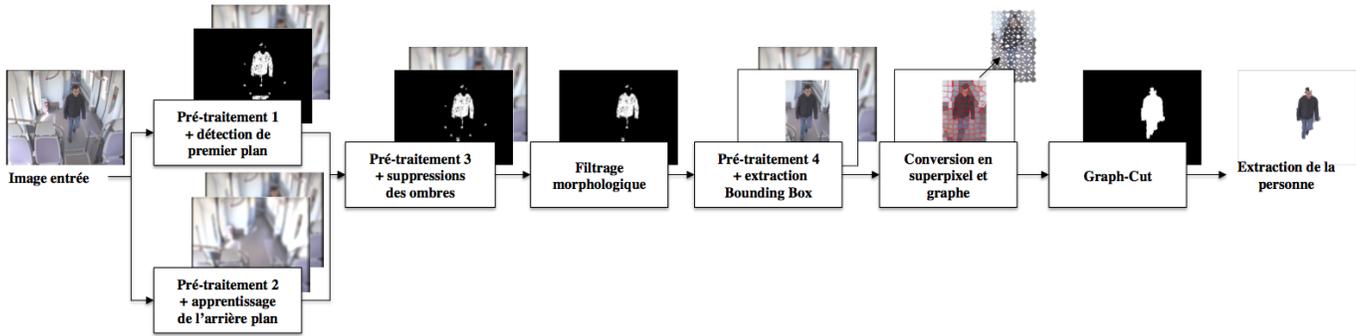


FIGURE 1 – Synopsis de la méthode proposée.

1.

- Optimisation par l’algorithme génétique : Devant le nombre important de méthodes disponibles et les différentes valeurs de paramètres possibles, nous avons fait le choix d’utiliser un algorithme génétique pour paramétrer notre méthode. Notre démarche n’a pas vocation à déterminer la meilleure méthode pour chaque étape mais de trouver la meilleure association possible en testant une multitude de possibilités. Pour ce faire, nous utilisons un modèle classique d’algorithme génétique avec une procédure en 3 phases (sélection, reproduction, mutation). Les chromosomes contiennent 30 gènes, chacun correspondant au paramètre d’une étape. La meilleure configuration obtenue après convergence de l’algorithme génétique sera donnée dans la section " Résultats expérimentaux".

- Pré-traitements : le but est de réduire les effets des changements rapides de luminosité et de bruits présents dans les images. Notons que le pré-traitement utilisé peut être différent selon qu’il est associé à la détection de premier plan, l’apprentissage de l’arrière plan, la suppression des ombres ou encore l’extraction de la Bounding Box. Le choix de celui-ci est effectué par l’algorithme génétique, nous avons implémenté quatre filtres et six invariants colorimétriques [5].

- Détection de premier plan : Dans le domaine de la détection de premier plan, l’état de l’art est vaste [9], mais les méthodes existantes voient leurs performances dégradées lorsqu’elles sont appliquées sur des bases de données complexes. Néanmoins, nous pouvons utiliser cette information comme une première approximation de la position des personnes dans les images. Nous avons donc implémenté les principales méthodes de l’état de l’art (Mélange de gaussienne, logique floue, réseau de neurones, analyse statistique couleur et texture).

- Suppression des ombres et filtrage morphologique : L’image issue de l’étape de détection de premier plan peut contenir de nombreux artefacts que nous pouvons corriger en supprimant les ombres et les légères erreurs de détection. Le problème des ombres est un problème bien connu de la littérature [8, 7] et spécialement dans le cas de la détection de premier plan. C’est pourquoi, nous avons choisi d’implémenter les principales méthodes basées sur la chromaticité, la géométrie, la texture et les caractéristiques physiques de l’image comme l’orientation des sources lumineuses. D’autre part, de nombreux auteurs semblent s’accorder sur le fait que les résultats sont améliorés lorsque

l’arrière-plan est connu. Nous réalisons donc un apprentissage de l’arrière-plan à partir des méthodes énoncées dans l’étape de détection de premier plan. Néanmoins nous dissocions l’apprentissage de l’arrière-plan de la détection du premier plan pour optimiser les données fournies en entrée de la méthode de suppression des ombres. Le deuxième post-traitement consiste à supprimer les petites régions de non-intérêt qui subsistent à l’aide d’opérations de morphologie mathématique.

- Une Bounding Box est ensuite extraite pour réduire la taille de la zone d’intérêt à traiter.

- Conversion en superpixel et graphe : L’image issue du post-traitement précédent donne une détection de la personne avec un score de précision élevé mais un score de rappel médiocre. Les deux étapes qui suivent ont donc pour but de trouver les pixels de l’arrière-plan mal labélisés. Nous commençons donc par segmenter l’image, en superpixels à l’aide de la méthode SLIC [2], puis créons le graphe d’adjacence de régions correspondant. Nous définissons ensuite, pour chaque noeud du graphe, une probabilité d’appartenance aux deux classes (premier plan  $c_{pp}$  et arrière plan  $c_{ap}$ ) que nous comparons à deux histogrammes couleurs décrivant  $c_{pp}$  et  $c_{ap}$ , initialisés à partir de l’information fournie en sortie de l’étape de post-processing.

- Coupe de graphe : Une fois ce graphe obtenu, nous le segmentons en deux classes ( $c_{pp}$  et  $c_{ap}$ ) à l’aide d’une coupe de graphe - Graph-Cut - [4]. Pour cela, nous avons besoin de deux informations exprimant la similarité, qui correspond au poids entre les noeuds du graphe (ici les superpixels), et la capacité qui correspond à l’appartenance de chaque noeud aux deux classes. La similarité  $S(s_i, s_j)$  est calculée à partir d’une distance couleur  $d(s_i, s_j)$  entre deux superpixels  $s_x$  (nous avons ici implémenté les normes  $L_1$ ,  $L_2$  et la distance de *Bhattacharyya*) :

$$S(s_i, s_j) = \exp\left(-\frac{d(s_i, s_j)}{2\theta^2}\right) \cdot \frac{1}{\text{dist}(s_i, s_j)}$$

La capacité  $W^{c_{pp}/c_{ap}}$  est calculée à partir des probabilités d’appartenance à chaque classe  $P^{c_{pp}/c_{ap}}$  des pixels, issues des deux histogrammes couleurs :

$$W^{c_{pp}/c_{ap}}(s_i) = -\log(P^{c_{pp}/c_{ap}}(s_i))$$

Le résultat illustré sur la Figure 4 donne une segmentation précise du premier plan correspondant à la silhouette de la personne.

### 3 Résultats expérimentaux

La méthode proposée étant décrite, nous allons maintenant montrer les résultats obtenus sur une séquence d'image extraite de la base de données du projet BOSS [1]. Les séquences issues de cette base de données ont été acquises à l'intérieur d'un train en mouvement pendant un après-midi très ensoleillé. Nous avons choisi une séquence regroupant un grand nombre de verrous scientifiques cumulés (changements rapides de luminosité, ombres portées, arrière-plan défilant à travers les vitres du train, etc). La séquence choisie contient 4258 images. La séquence comprend un passage complet de douze personnes de face. La qualité d'extraction des personnes a été effectuée avec la méthode F-mesure. L'apprentissage réalisé par l'algorithme génétique doit être effectué sur chaque séquence testée (même si les paramètres optimaux fournis pour un apprentissage donné offrent des résultats satisfaisants sur d'autres séquences de test). Pour cette base, l'apprentissage est réalisé sur les deux premières personnes de la séquence vidéo (ce qui correspond à 530 images). Une fois l'apprentissage réalisé, le temps de traitement d'une image utilisant l'optimum calculé est de 8 secondes avec un processeur cadencé à 1.8 Ghz. Ainsi, les méthodes de détection de premier plan basées sur la logique floue *Fuzzy Sugeno Integral* [10] et *Fuzzy Choquet Integral* [3] sont les meilleures pour la détection de premier plan et l'apprentissage de l'arrière-plan. La méthode basée sur les caractéristiques physiques de l'image [6] (l'orientation de la lumière) est considérée comme la plus efficace pour supprimer les ombres, les filtres *bilatéral* et *gaussien* et les invariants *normalisation affine* et  $l_1l_2l_3$  sont respectivement les meilleurs pour l'apprentissage de l'arrière plan et la suppression des ombres.

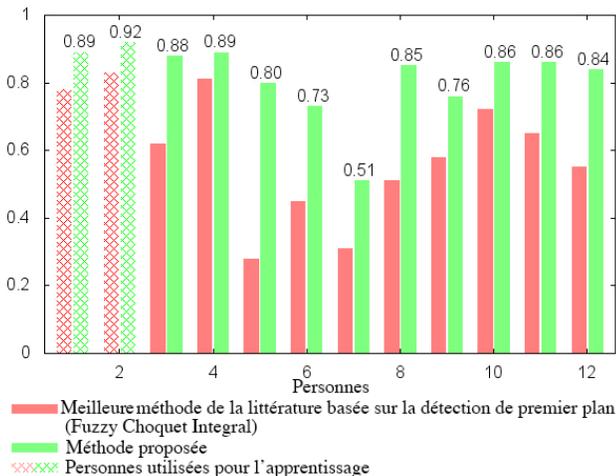


FIGURE 2 – Score F-mesure (F1) obtenu pour chaque passage de personnes devant la caméra, avec la meilleur méthode de détection de premier plan de la littérature et avec la méthode proposée

La figure 2 illustre sur la séquence choisie, les résultats d'extraction de personnes obtenus avec la méthode proposée (en vert) et avec la meilleure méthode de détection de premier plan

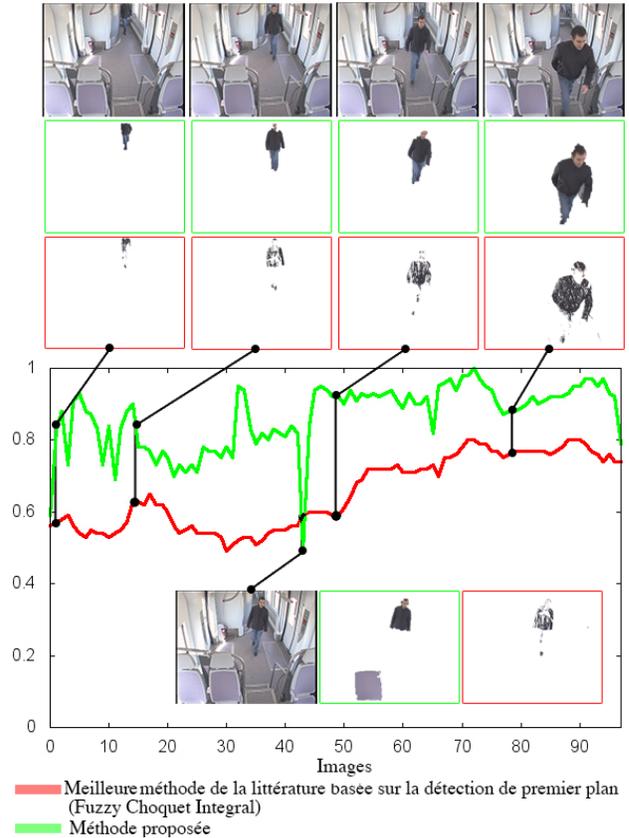


FIGURE 3 – Score F-mesure pour le passage complet (image par image) d'une personne devant la caméra

de la littérature définie par l'algorithme génétique (en rose). Il est important de noter que le F-mesure obtenu correspond à la moyenne des F-mesures calculés pour un passage d'une personne devant la caméra (environ 120 images/personne). Globalement notre méthode donne de meilleurs résultats (F-mesure = 0.82) que la meilleure détection de premier plan issue de la littérature (en F-mesure = 0.59). Pour une meilleure visualisation, nous illustrons sur la Figure 4, les résultats de l'extraction de personnes obtenus à partir de la méthode proposée. Notons également que la méthode proposée permet d'obtenir un F-mesure toujours supérieur à 0.73 excepté pour la 7ème personne. Nous pouvons expliquer ce mauvais résultat sur la 7ème personne (F-mesure à 0.51) par une détection des sièges qui n'aurait pas du l'être (cf Figure 4).

La Figure 3 quant à elle montre le score F-mesure obtenu image par image pour le cas d'un déplacement complet d'une personne devant la caméra. Nous pouvons constater que la personne est correctement extraite dans chaque image quelle que soit sa distance à la caméra. En effet, il y a peu de pixels d'arrière-plan détectés comme premier plan et l'extraction des contours de la personne, même si elle n'est pas parfaite, reste très satisfaisante. Ce dernier point s'explique par la difficulté qu'a la méthode de segmentation à segmenter les zones floues causées par un déplacement rapide de la personne (par rapport aux caractéristiques de la caméra). Il subsiste néanmoins, le problème

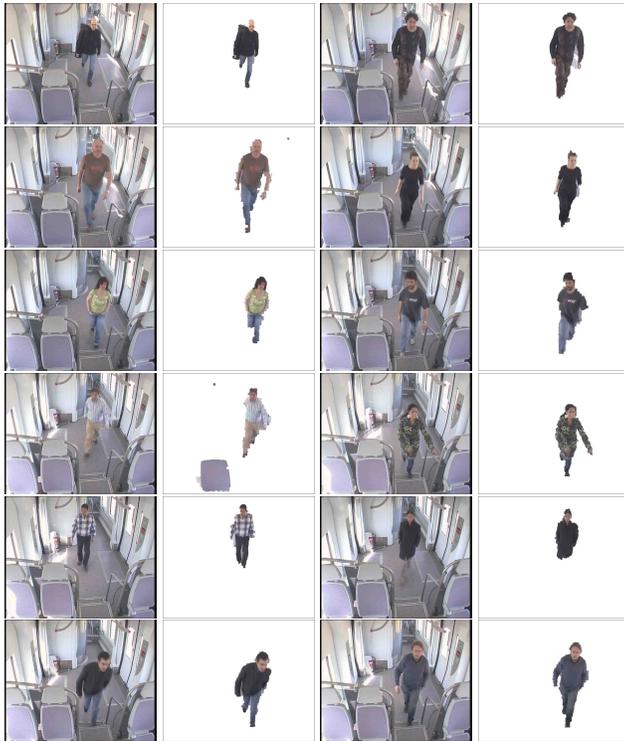


FIGURE 4 – Résultats d’extraction de personnes (colonnes 1-3 : images acquises ; colonnes 2-4 : résultats d’extraction).

de détection évoqué précédemment sur l’image 43 et illustré sur le graphique de la Figure 3. Ce problème mineur pourra être facilement corrigé par la suite en intégrant l’aspect temporel (avec une étape de suivi de la personne) dans la phase d’extraction de personnes.

## 4 Conclusion

Dans cet article, nous avons montré que les meilleures méthodes de détection de premier plan de la littérature ne sont pas toujours robustes et suffisamment précises pour obtenir une bonne extraction des personnes dans un environnement de transport complexe compte tenu de la combinaison des verrous scientifiques comme les changements brusques de luminosité, le bruit, les ombres, un arrière-plan défilant, des parois vitrées, etc. Néanmoins, l’utilisation d’une détection de premier plan semble constituer une première base pour déterminer la position grossière des personnes dans l’image. Sur ce constat, nous avons proposé une nouvelle méthode d’extraction de personnes basée sur la segmentation de l’image en superpixels couplée à une coupe de graphe initialisée par les informations de la détection de premier plan. Pour renforcer cette détection de premier plan, nous avons proposé une étape de pré-traitement robuste basée sur l’utilisation de filtres et d’invariants colorimétriques ainsi qu’une étape de post-traitement visant à supprimer les ombres et les artefacts. Le paramétrage de la méthode est réalisé par un algorithme génétique et évalué selon le critère

F-mesure. Les résultats d’extraction de personnes obtenus avec la méthode proposée sont meilleurs que ceux obtenus avec la meilleure méthode de détection de premier plan de la littérature (F-mesure de 0,82 contre 0,59). Même si la méthode proposée améliore largement les résultats d’extraction de personnes, les contours des personnes extraites ne sont pas parfaits et certaines zones de non-intérêts subsistent. Pour résoudre ce problème, nos futurs travaux consisteront à tenir compte de l’aspect temporel travers un suivi de la personne.

## Références

- [1] Boss european project (on bord wireless secured video surveillance), <http://www.multitel.be/image/research-development/research-projects/boss.php>.
- [2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11) :2274–2282, 2012.
- [3] F. El Baf, T. Bouwmans, and B. Vachon. Fuzzy integral for moving object detection. In *FUZZ-IEEE*, pages 1729–1736, 2008.
- [4] Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary region segmentation of objects in n-d images. In *IEEE International Conference on Computer Vision*, volume 1, pages 105–112, 2001.
- [5] S. D. Hordley, G. D. Finlayson, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38 :2005, 2005.
- [6] J. Huang and C. Chen. Moving cast shadow detection using physics-based features. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2310–2317, 2009.
- [7] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. Detecting moving shadows : Algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 :918–923, 2003.
- [8] A. Sanin, C. Sanderson, and B. C. Lovell. Shadow detection : A survey and comparative evaluation of recent methods. *Pattern Recognition*, 45(4) :1684 – 1695, 2012.
- [9] A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122 :4 – 21, 2014.
- [10] H. Zhang and D. Xu. Fusing color and texture features for background model. In *Fuzzy Systems and Knowledge Discovery*, volume 4223 of *Lecture Notes in Computer Science*, pages 887–893. Springer Berlin Heidelberg, 2006.
- [11] D-N. Truong Cong, L. Khoudour, C. Achard, C. Meurie, O. Lezoray, People re-identification by spectral classification of silhouettes. *Signal Processing*, 90(8) :2362 – 2374, 2010