Extraction des objets en mouvement : une approche mixte contours-régions

L. Biancardini[1] S. Beucher[2] L. Letellier[1]

[1] CEA LIST, Laboratoire Calculateurs Embarqués et Images
[2] Ecole des Mines de Paris, Centre de Morphologie Mathématique

CEA LIST Saclay, DRT/DTSI/SARC/LCEI, B.528
91191 Gif/Yvette Cedex, France

biancardini@cea.fr

Résumé

Dans cet article, nous proposons une nouvelle méthode pour extraire les objets en mouvement dans une séquence d'images sans aucune estimation de flot optique, ni aucun a priori sur la scène, comme une image de fond. L'objectif est d'obtenir une méthode d'extraction rapide et efficace, robuste au bruit et aux forts déplacements. Cette approche est basée sur des différences d'images successives combinées avec une segmentation hiérarchique de l'image au temps courant. Nous proposons tout d'abord un nouveau schéma de différenciation basé sur le gradient spatial d'images successives, permettant d'extraire les contours des objets en mouvement. Dans un second temps, les régions en mouvement sont déduites des contours précédemment définis, à l'aide de la segmentation hiérarchique. Les objets mobiles ainsi obtenus sont représentés par un graphe d'adjacence de régions. La méthode a été validée dans le contexte de la vidéosurveillance, sous l'hypothèse d'une caméra statique.

Mots Clef

Détection du mouvement, segmentation hiérarchique, vidéosurveillance.

Abstract

In this paper, we propose a new method to extract moving objects from a video stream without any motion estimation. The objective is to obtain a fast and efficient extraction method, robust to noise and large motions. Our approach consists in a frame differencing strategy combined with a hierarchical segmentation approach. First, we propose to extract moving edges with a new difference scheme, based on the spatial gradient of successive frames. In the second stage, the moving regions are extracted from previously detected moving edges by using the hierarchical segmentation. The obtained moving objects description is represented as a region adjacency graph. The method is validated on real sequences in the context of videosurveillance, assuming a static camera hypothesis.

Keywords

Motion detection, hierarchical segmentation, videosurveillance.

1 Introduction

La vidéosurveillance automatique est un domaine en plein essor dans la communauté de la vision. L'intérêt croissant porté à ce secteur est notamment lié à l'apparition de nouveaux capteurs vidéo et de puissance de calcul à bas coût. Aujourd'hui des applications aussi diverses que complexes sont envisagées : sécurité dans les lieux publics (détection de colis suspects, mouvements de foules, altercations, etc.), contrôle du trafic routier (détection des comportements à risque, des accidents), comptage de personnes en vue d'établir des statistiques, domotique, etc. Comme on peut le constater au travers de ces quelques exemples, l'enjeu d'un tel système réside aujourd'hui dans la capacité à reconnaître des attitudes et interpréter des scénarios [24]. Pour accomplir une tâche aussi complexe, une analyse de l'image au niveau pixellique n'est pas suffisante. Il est nécessaire de disposer d'une représentation de l'image plus proche de sa sémantique comme une segmentation en régions. Dans notre approche, cette représentation, associée à un graphe valué codant les relations d'adjacence, permet de mettre en oeuvre un algorithme de fusion des régions. On peut ainsi obtenir plusieurs segmentations de l'image correspondant à plusieurs niveaux de résolution.

Quel que soit le degré de complexité d'analyse recherché, la détection des objets d'intérêt dans l'image reste l'un des problèmes majeurs. Les objets en mouvements constituent un ensemble de régions d'intérêt à la fois générique et bien adapté au contexte de l'interprétation de scénarios.

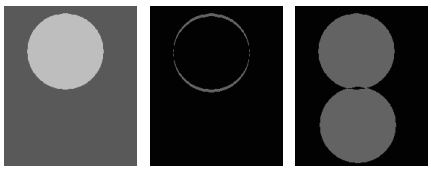
On peut classer les méthodes de détection de mouvement en trois catégories : les méthodes de soustraction de fond [7, 8, 9, 17, 21], celles qui s'appuient sur un calcul du flot optique [7, 10] et enfin celles qui se basent sur des différences d'images successives [7, 13, 18, 14, 19]. Dans le domaine de la vidéosurveillance, la soustraction de fond est la méthode la plus fréquemment employée. Malheureusement celle-ci nécessite la connaissance d'une image de référence, souvent difficile à obtenir et qui doit être remise à jour au cours de la séquence. Dans la seconde catégorie de méthodes, l'estimation du flot optique n'est généralement qu'une première étape à l'extraction des objets mobiles. Cette estimation est non seulement coûteuse en temps de calcul mais aussi très sensible aux mouvements de forte amplitude. De plus, les estimées sont souvent bruitées aux frontières des objets mobiles et délicates à obtenir à l'intérieur de larges régions homogènes.

La méthode proposée dans cet article appartient à la troisième catégorie. Ces techniques permettent de détecter les objets en mouvement avec un faible coût de calcul. Cependant, il est en général impossible d'extraire simultanément les objets rapides et le objets lents à l'aide de ce type de méthode. Dans ce cas, un compromis entre le nombre de cibles manquées et les fausses détections est difficile à trouver. Pour surmonter ces problèmes, nous proposons, dans un premier temps, un nouveau schéma différentiel basé sur les gradients de l'image, permettant d'extraire les contours des objets en mouvement. Ensuite une segmentation hiérarchique [5, 2, 3] de l'image courante est utilisée pour extraire les régions en mouvement à partir de ces contours.

Cet article est organisé comme suit : la section 2 présente brièvement l'état de l'art en insistant sur les problèmes rencontrés dans le cas de déplacements importants, la section 3 introduit la méthode d'extraction des contours en mouvement. Dans la section 4, c'est l'utilisation de la segmentation hiérarchique pour l'extraction des régions mobiles qui est expliquée. La section 5 présente des résultats expérimentaux obtenus en conditions réelles sur des scènes de vidéosurveillance. La dernière section conclut sur la méthode et présente le travail à venir.

2 Travaux antérieurs

De nombreuses approches visant à extraire les objets mobiles, en s'appuyant sur des différences d'images successives, ont déjà été proposées dans la littérature [4, 13, 15, 17, 20, 22]. Ces méthodes sont avantageuses dans la mesure où elles ne nécessitent pas de calcul de mouvement. Dans le principe, elles s'appuient sur les occultations : lorqu'un objet se déplace, il couvre et découvre d'autres parties de la scène (statiques ou elles-mêmes en mouvement). La plupart du temps, en traitement d'images monocaméra, les occultations sont détectées par la présence de fortes valeurs dans la différence absolue de deux images successives. Le résultat obtenu de cette façon, correspond simultanément aux parties couvertes et découvertes par l'objet mobile, et en l'absence d'information de mouvement, on ne peut pas déterminer quels sont les pixels occultés et les pixels occultants. D'autre part, les occultations ne sont jamais situées complètement à l'intérieur, ni à l'extérieur des objets en mouvement. En effet les pixels désoccultés entre t et $t+1$ se situent à l'intérieur de l'objet dans la première image tandis que les pixels occultés se situent à l'extérieur et vice versa dans la seconde image. De la même façon, les occultations ne correspondent jamais aux contours des objets mobiles ni dans la première image, ni dans la seconde image. Elles sont souvent désignées comme les frontières du mouvement [4] puiqu'elles indiquent le début et la fin de celui-ci. Le phénomène de délocalisation s'aggrave lorsque les objets sont rapides et/ou que la fréquence d'acquisition des images est faible. Dans le cas d'un objet homogène très lent l'intensité lumineuse change peu à l'intérieur de celui-ci entre deux images successives. La différence d'images donne dans ce cas un résultat proche de la silhouette de l'objet 1(b), mais pas d'information à l'intérieur des régions, et seules les régions texturées sont détectées dans leur ensemble. Par contre, si entre deux images, un objet s'est complètement déplacé par rapport à sa position initiale, les valeurs de différence d'images sont fortes aux positions occupées par l'objet à chacun des deux instants (figure 1(c)). Ce phénomène appelé fantôme [16], entraîne de nombreuses fausses détections. Dans [17, 18], les auteurs proposent une alternative afin d'extraire un ensemble de pixels appartenant à l'objet en mouvement. L'opérateur proposé effectue la différence de deux paires d'images successives au temps $(t-1,\ t)$ et $(t,\ t+1)$. Les deux images résultantes sont alors binarisées et un ensemble de points appartenant à l'objet mobile est obtenu en effectuant l'intersection des masques binaires correspondants. La méthode est utilisée avec succès afin d'extraire la silhouette des objets dont le mouvement est grand entre deux images ainsi que les régions mobiles suffisamment texturées. Par contre, dans le cas d'objets homogènes se déplaçant lentement, le nombre de points extraits (zone grisée sur la figure 2) devient rapidement insuffisant pour être exploité.

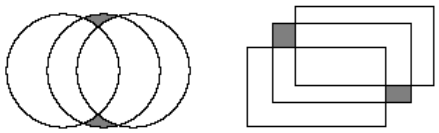


(a) (b) (c)

FIG. 1 – *Différence de deux images successives dans le cas d'un disque homogène se déplaçant de haut en bas : (a)image originale, (b)résultat pour un faible déplacement, (c) résultat pour un déplacement important.*

3 Extraction des contours en mouvement

Dans la suite, $I^t : \mathbb{Z}^2 \rightarrow \mathbb{N}$ désigne la luminance de l'image RVB correspondant au temps t. Cette image est choisie

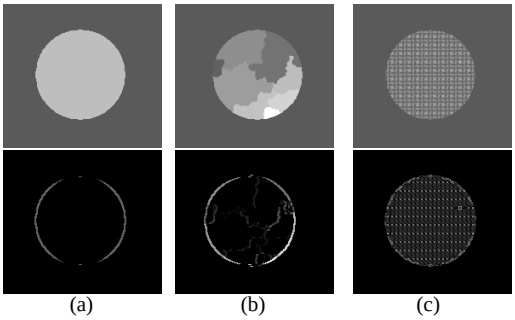


(a) (b)

FIG. 2 – *Résultats de l'opérateur de double différence (en gris) pour : (a) cercle se déplacant horizontalement, (b) un rectangle se déplacant obliquement.*

comme image de référence. L'objectif de ce paragraphe est de détecter dans $I(t)$, les contours des régions en mouvement entre les instants $t-1$, t et $t+1$. En se basant sur le fait que les contours de l'image sont calculés à partir du module du gradient, on peut raisonnablement faire l'hypothèse que les contours en mouvement se déduisent de l'évolution de la norme du gradient $\|\nabla I^t\|$. Cependant, la différence de deux gradients successifs présente les mêmes inconvénients que ceux précédemment exposés dans le cas de deux images consécutives. Pour cette raison, la méthode proposée prend en compte trois images successives I^{t-1}, I^t et I^{t+1}. On calcule le module du gradient de chaque image, $g^t = \|\nabla I^t\|$ (respectivement $g^{t\pm1}$). Ensuite, on applique symétriquement la différence absolue sur les deux couples successifs de module du gradient. La mesure des contours en mouvement (mcm) à un temps t donné est alors définie par le minimum de ces deux différences.

$$mcm^t = \inf\left(\left|g^{t+1} - g^t\right|, \left|g^t - g^{t-1}\right|\right) \qquad (1)$$



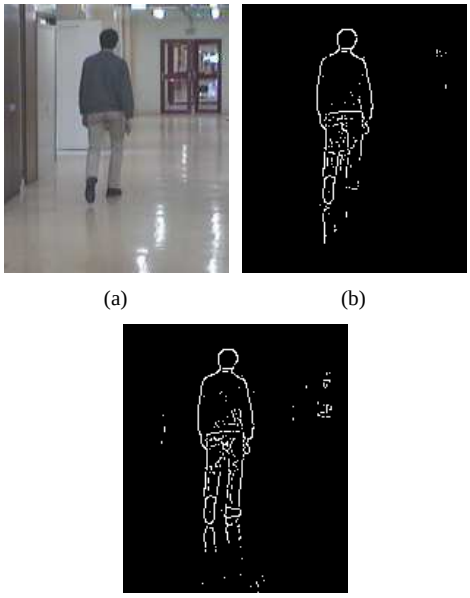
(a) (b) (c)

FIG. 3 – *Résultats du mcm obtenus pour : (a) une région homogène, (b) un groupe de régions homogènes, (c) une région texturée.*

Les propriétés de l'opérateur du minimum et l'analyse du gradient au fil de trois images entraînent les caractéristiques ci-dessous :

1. Une réponse correctement localisée à la frontière des objets en mouvement. Dans le cas (idéal) d'une région traversant une zone homogène, l'opérateur mcm est égal au gradient de l'image.
2. Une robustesse significative à l'amplitude du mouvement. Dans le cas d'un mouvement rapide, le résultat n'est pas délocalisé. La quasi-totalité des fantômes liés à la présence de forts déplacements est éliminée. Ainsi, on peut envisager l'utilisation de trois images $t - \Delta t$, t, $t + \Delta t$, plus espacées dans le temps (i.e. de sous-échantillonner la séquence) afin d'extraire des objets se déplaçant lentement sans générer de fausses alarmes sur ceux qui sont les plus rapides.
3. L'opérateur répond convenablement qu'il s'agisse d'une région texturée (figure 3(c)) ou homogène : on peut en effet constater sur la figure 3(a) que contrairement à l'opérateur discuté dans la section 2, la majorité des points du cercle est détectée même si le mouvement du disque est faible.
4. Une robustesse significative au bruit aléatoire puisque celui-ci ne se répète pas à l'identique d'une image à l'autre.

Différentes méthodes d'obtention du gradient ont été testées [5, 11, 12]. Il ressort de ces expérimentations que la méthode peut être employée quel que soit la méthode utilisée pour calculer le gradient. La seule limitation réside dans l'utilisation d'un gradient trop fortement régularisé. En effet, dans le cas d'un plateau de gradient (ou plus généralement d'un région floue dans l'image), la localisation de la mesure par rapport aux frontières n'est plus assurée, ce qui est prévisible puisque les contours de l'image eux-même ne peuvent pas être localisés correctement. L'utilisation d'un gradient régularisé est cependant envisageable, à condition que le lissage de l'image soit faible.
Les figures 4(c) et 4(b) illustrent les deux premières propositions et une application possible de la mesure décrite ci-dessus à la détection des contours en mouvement. Elles ont été obtenues de la façon suivante : une fois que le mcm est calculé, les contours de l'image sont détectés et pondérés par le mcm (pour chaque point de contour, une recherche locale du maximum de mcm dans la direction du gradient est effectuée). On réalise ensuite un seuillage par hystéresis du résultat précédent. La première des deux figures montre le résultat en considérant des images à $t-1$, t, $t+1$, la seconde en considérant $t-15$, t, $t+15$. Cependant, cette méthode reste sensible à certains facteurs comme le problème des contours en glissement sur eux-même (problème d'ouverture). Comme dans le cas d'une différence d'images successives, la réponse de l'opérateur est liée à l'amplitude du gradient spatial de l'image de référence et des régions en mouvement. Par conséquent, on peut s'attendre à ce que l'opérateur ne réponde pas uniformément tout au long du contour d'une région donnée, puisque celle-ci présentera des contrastes différents avec ses régions voisines. De plus un objet mobile faiblement constrasté avec ses régions voisines sera généralement difficile à extraire. Par conséquent, dans ces situations mais aussi dans le cas d'un mouvement de faible amplitude, l'ensemble des contours mis en évidence par le mcm sera probablement incomplet. Le problème de l'obtention de contours fermés (i.e. des régions



(a) (b)

(c)

FIG. 4 – *(a)image originale, (b), (c) contours en mouvement détectés en considérant les images aux temps (respectivement) ($t-1$, t, $t+1$) et ($t-15$, t, $t+15$).*

en mouvement sous-jacentes) est donc abordé dans la section suivante, dans laquelle une segmentation de la luminance de l'image est utilisée afin de pallier le manque de données concernant le mouvement.

4 Extraction des régions en mouvement

Même si dans de nombreux cas, l'obtention de contours en mouvement peut suffire, dans le cadre de la vidéosurveillance et de l'interprétation de séquences auquel appartient ce travail, il est indispensable d'obtenir des régions. Dans cette section, une méthode d'extraction des régions en mouvement, basée sur la mesure présentée au paragraphe précédent est proposée. Dans un certain nombre de situations, l'opérateur mcm ne fournit pas suffisamment d'information pour que tous les points d'un contour en mouvement puissent être détectés isolément. L'opérateur mcm ne permet donc pas d'obtenir (par exemple, par un seuillage), un contour fermé définissant correctement la région mobile sous-jacente. Du fait de la localisation du mcm par rapport aux frontières des objets en mouvement dans l'image courante, on peut supposer qu'il est réalisable d'établir une correspondance entre celui-ci et les contours obtenus par une segmentation de la luminance de l'image. On recherche alors dans cette partition les régions en mouvement comme celles qui présentent de fortes valeurs du mcm, réparties de façon suffisamment dense et importante au long de leurs frontières. En d'autre termes, les régions de la segmentation spatiale sont utilisées comme un ensemble de régions en mouvement potentielles et les régions détectées sont celles qui s'ajustent le mieux aux contours mis en évidence par le mcm.

Cependant, obtenir en une seule partition un découpage adapté à toute l'image est une tâche ardue. Généralement, on aboutit à une sur-segmentation de certaines parties alors que des détails ou des régions faiblement contrastées sont perdus, même s'ils sont sémantiquement importants (par exemple la tête ou les bras d'une personne, dans le contexte qui nous intéresse). En fait, une description fiable et complète de l'image requiert plusieurs niveaux de détails. Par conséquent dans notre approche les régions en mouvement sont recherchées parmi les niveaux d'une segmentation hiérarchique.

4.1 Segmentation hiérarchique et ensemble de candidats à la détection

Dans [3, 2, 20], les auteurs proposent d'obtenir une partition représentative de l'image en groupant les régions d'une sur-segmentation initiale. Cependant, on ne peut examiner toutes les possibilités d'assemblage afin de détecter celle qui correspondrait de façon optimale aux gradients en mouvement. Elles représentent en effet un nombre trop important de cas et le temps de calcul nécessaire serait pénalisant. La complexité de cet examen peut être réduite en présélectionnant des régions de la partition initiale, à l'aide de caractéristiques des cibles, comme par exemple leur couleur [20].

Quand on ne dispose pas d'un tel a priori sur les objets à extraire, une autre façon de réduire le nombre de régions à considérer est de construire une partition hiérarchique de l'image : dans un premier temps, une partition initiale est calculée et son graphe d'adjacence G est construit en créant un noeud pour chaque région et une arête pour chaque paire de régions adjacentes. Les arêtes de ce graphe sont alors valuées en accord avec un critère de dissimilarité (par exemple, la différence de luminance moyenne entre les deux régions). Une segmentation hiérarchique est obtenue en fusionnant progressivement les régions de la segmentation initiale par ordre de dissimilarité croissante. Le processus est ainsi itéré jusqu'à ce qu'il ne reste qu'une seule région égale à l'image entière.

En conservant la séquence des fusions effectuées lors de la création de la partition hiérarchique, on peut construire un ensemble de régions "candidates à être des régions en mouvement" : la liste est initialisée avec l'ensemble des régions de la partition la plus fine, puis chaque fois que deux régions sont fusionnées, la région résultante est ajoutée à l'ensemble des candidats. On peut remarquer d'une part que les régions candidates sont triées par rapport à leur niveau d'apparition dans la hiérarchie. Le nombre total de régions dans la liste est alors 2N-1, où N est le nombre de régions initialement présentes dans la partition initiale. La segmentation hiérarchique ne contient que les régions les plus significatives au sens du critère choisi. Comme le montre la figure 5, la liste de fusions peut être représentée par un arbre binaire, dans lequel chaque niveau correspond

à une région candidate générée par une fusion. Sur cette figure la numérotation des noeuds correspond à l'ordre d'apparition des régions dans la hiérarchie (sauf les noeuds du niveau le plus fin qui sont numérotés par ordre lexicographique d'apparition dans l'image). Un sous-arbre donné correspond alors à un sous-ensemble de régions candidates, liées les unes aux autres par une relation d'inclusion.

Dans la mesure ou l'on ne repondère pas les arêtes du graphe lors de la création de la hiérachie (cela donnerait une segmentation hiérachique plus robuste mais serait aussi beaucoup plus coûteux en temps de calcul), la donnée de la séquence de fusions (ou de l'arbre binaire des fusions) est équivalente à la donnée de l'arbre de poids minimum du graphe G (chaque arête de cet arbre correspond à une fusion [3]). Celui-ci peut être efficacement obtenu à l'aide de l'un des algorithmes classiques (et rapides) de construction de l'arbre de poids minimum. Dans notre approche la segmentation initiale est obtenue par la ligne de partage des eaux [5] du gradient de l'image courante. Cependant aucune limitation n'est imposée sur la façon d'obtenir celle-ci, si ce n'est qu'elle doit être suffisamment fine pour contenir les cibles à segmenter. Le critère de dissimilarité choisi est un critère de contraste robuste : pour une paire de régions adjacentes donnée, la valeur du critère est définie comme la valeur médiane du module du gradient le long de la ligne de partage des eaux séparant les deux régions.

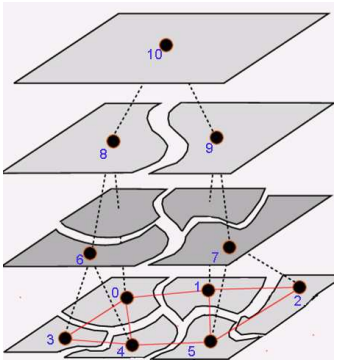


FIG. 5 – *Le graphe d'adjacence G du niveau le plus fin (en rouge) et l'arbre binaire de fusions superposés à la segmentation hiérarchique.*

4.2 Sélection des régions en mouvement

La sélection des régions en mouvement est effectuée en parcourant la hiérarchie du niveau le plus fin vers le niveau le plus grossier. Pour chaque noeud i de l'arbre de fusions, on évalue un score mesurant la présence de contours en mouvement (fortes valeurs du mcm) aux frontières de la région candidate C_i associée.

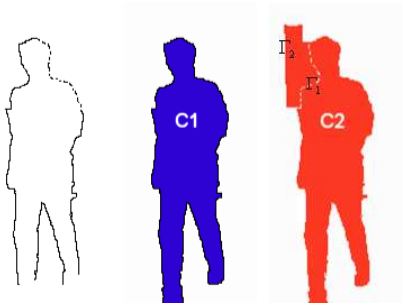
Par la suite, ∂C_i désignera la frontière de la région C_i définie comme l'ensemble des points de la ligne de partage des eaux entourant celle-ci. Pour chaque assemblage de régions autorisé par la hiérarchie, le score est calculé comme suit :

$$ms(C_i) = 1 - \frac{\sum_{t \in partial C_i} \exp(\frac{-mcm(t)}{\beta})}{card(\partial C_i)} \qquad (2)$$

où $card$ renvoie le nombre d'éléments de l'ensemble ∂C_i. En d'autres termes, on effectue une classification floue de chaque pixel de la frontière, en lui attribuant un score dans l'intervalle $[0,1]$. Un score nul indiquant l'absence de mouvement au point considéré et un score égal à un la présence certaine d'un tel contour. L'équation (2) équivaut alors à estimer le nombre de points en mouvement autour de la région candidate en pondérant la contribution de chaque point par son score.

Chaque candidat généré par la hiérarchie est successivement testé par ordre d'apparition. Un candidat est labellisé comme étant une région en mouvement si son score $ms(C_i)$ est supérieur à un seuil $T_{score} \in [0,1]$.

Contrairement à d'autres approches [12, 13], comme la hiérarchie est obtenue à l'aide d'un critère de segmentation spatial, une seule et unique région ne correspond pas *a priori* à un objet en mouvement donné (sauf si celui-ci répond au critère en question). Dans le cas général, les objets mobiles sont donc détectés sous la forme d'un agrégat de régions qui répondent au critère spatial. Les candidats étant emboîtés les uns dans les autres, lors du parcours de la hiérarchie plusieurs candidats se recouvrant partiellement peuvent être détectés à une même position. Certaines configurations comme celle décrite par la figure 6, peuvent alors entraîner de fausses détections (sous-segmentation des cibles). A titre d'exemple, dans la figure 6(c) la région C_2 est détectée parce qu'elle possède une importante portion de frontière en commun avec la région C_1 (figure 6(b)) qui présente elle même un score très élevé. Dans le cas présent, la région C_2 est obtenue en remplaçant la portion de contour Γ_1 par le contour Γ_2 (figure 6(c)), le reste des contours de C_1 étant maintenu. Or la contribution au score du contour Γ_1 est plus importante que celle de Γ_2. Par conséquent le score de la région C_1 est supérieur à celui de C_2. Malgré un score supérieur à T_{score}, la région C_2



(a) (b) (c)

FIG. 6 – *(a) valeurs du $mscm$ (seuillées et inversées pour les besoins de la figure), (b) la région à détecter, (c) sous-segmentation de la région à détecter.*

doit alors être rejetée. Il est à noter que que ce phénomène se produira plus fréquemment lorsqu'une région en mouvement est présente dans la hiérarchie. La région en mouvement "optimale" est alors localement définie dans la hiérarchie comme la plus grande région maximisant le score (2). Pour éliminer les situations comme celle dépeinte par la figure 6, lors de la classification d'une région donnée, il faut donc tenir compte des scores et des décisions précédemment prises sur les régions contenues dans C_i. Comme

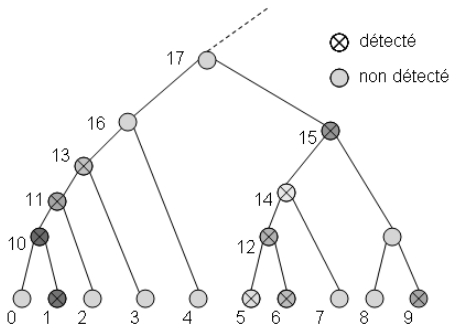


FIG. 7 – *Illustration du processus de sélection des régions en mouvement dans la hiérarchie.*

précédemment exposé dans le section 3.1, chaque candidat (exceptés ceux de la partition initiale) est le résultat d'une séquence de fusions (du point de vue de l'arbre de fusions, ces régions correspondent aux descendants de C_i). La région C_i résulte donc de la fusion de deux sous-arbres correspondant à deux sous-régions disjointes. Chacun d'entre eux peut contenir zéro, une, voire plusieurs sous-régions qui ont successivement été sélectionnées lors du processus. Dans le premier cas, en l'absence de décision antérieure, la détection d'une région candidate dépend uniquement du score obtenu par l'équation 2. Si au minimum une région a été détectée dans l'une des sous-branches (cette région ne correspond pas forcément à l'un des fils de la région courante), alors la dernière région détectée contient toutes les autres et présente la plus grande aire en commun avec la région C_i. D'autre part, le procédé de sélection étant appliqué récursivement à partir du niveau le plus fin, cette région a elle-même été choisie en accord avec ses propres prédécesseurs dans le processus et l'équation 2.
Si un seul des sous-arbres contient une détection antérieure (c'est par exemple le cas pour $i = 10{,}11{,}13$ dans la figure 7, c'est aussi le cas exposé par la figure 6), nous sommes en présence d'une région détectée qui s'agrandit en "absorbant" des régions non détectées. Dans ce cas, le score de la région C_i est directement comparé avec le score de la dernière région détectée. Si les deux sous-arbres contiennent des régions déjà détectées (c'est le cas des noeuds 15 et 17 dans la figure 7), il faut simultanément prendre en compte les deux décisions antérieures. Le score de la région courante est alors comparé avec la moyenne des scores obtenus sur les deux sous-régions, pondérée par la proportion occupée par chacune d'entre elle dans la région C_i. Il a pu expérimentalement être constaté que certaines régions de petite taille se trouvant près des frontières d'un objet en mouvement pouvaient présenter des scores importants. D'autre part, une large région en mouvement présente rarement de fortes valeurs de mcm tout au long de sa frontière. Afin que ces petites régions ne biaisent pas la détection, il suffit de ne pas comparer les scores de deux régions si leurs dimensions respectives sont trop différentes. Le ratio des aires est utilisé afin de déterminer si l'échelle des deux régions est compatible. Les résultats présentés dans la section 5, ont été obtenus en imposant un ratio des aires supérieur à 0.1.
Comme le montre la figure 7, chaque cible correspond initialement à un ensemble de régions distribuées irrégulièrement dans la hiérarchie et entremêlées de régions non détectées (statiques). Dans un premier temps, l'ensemble des noeuds détectés est complété en ajoutant tous les descendants de régions détectées (flèches descendantes, du noeud 12 vers les noeuds 10, 9, 4 et 3). Par re-projection, une cible correspond donc à un amas de régions connexes (CC sur la figure 8) du niveau le plus fin. On suppose qu'une cible peut éventuellement représenter un groupe. Chacune de ces composantes connexes est un sous-graphe du graphe d'adjacence de l'image. En appliquant l'algorithme proposé dans la section 4 à ces sous-graphes on obtient une représentation de chaque cible sous la forme d'une hiérarchie de graphes. Le critère de fusion utilisé étant toujours le même (section 4), l'arbre obtenu pour chaque objet en mouvement (valuation des arêtes comprises) sera pratiquement identique au sous-arbre correspondant à cet objet dans l'arbre global. Dans le cas de la figure 8, l'arbre de la cible composée des noeuds 0 à 5 (les régions 4 et 5 sont supposées adjacentes), peut être déduit de l'arbre initial, en remplaçant les branches constituées des noeuds 6, 7, 8, 11 classifiées comme statiques, par une arête reliant les noeuds 5 et 13 (en pointillés, sur la figure 8). Conformément, à l'algorithme de la section 4, cette arête est repondérée par la valeur minimum des arêtes du graphe d'adjacence encore actives (reliant deux régions détectées). Ainsi la représentation de chaque cible peut être déduite de l'arbre global.

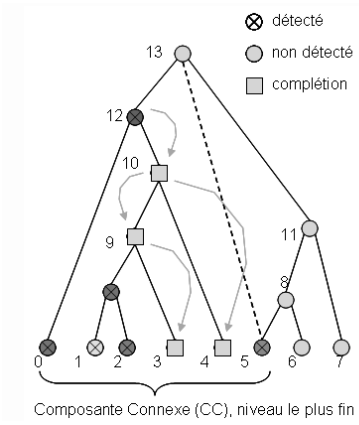


FIG. 8 – *Extraction finale des régions en mouvement et représentation des cibles.*

5 Résultats Expérimentaux

Dans cette section nous présentons une série de résultats obtenus sur des séquences prises dans des conditions réelles de fonctionnement d'un système de vidéosurveillance. Le gradient de Deriche [11] est utilisé afin d'obtenir la mesure des contours en mouvement (voir section 3) ainsi que la ligne de partage des eaux de l'image de référence. Le paramètre de régularisation σ est choisi supérieur à deux, de façon a préserver les structures fines et les régions peu contrastées.

Le nombre de régions nécessaires dans la partition initiale dépend des constrastes présents dans l'image et de l'échelle des cibles à détecter. Plus les cibles sont petites et/ou peu contrastées, plus le nombre de régions de la segmentation initiale doit être important. Les cibles de grande taille et plus proches de la caméra peuvent être représentées avec peu de régions. La qualité du résultat dépend du fait que l'on ait su préserver les régions en mouvement dans cette segmentation initiale. Cependant, le degré de sursegmentation influe fortement sur le coût de construction de la segmentation hiérarchique et la durée de la recherche des régions en mouvement. On est donc amené à ne considérer qu'un jeu de marqueurs réduit [5] dans le calcul de la ligne de partage des eaux. Dans cet article, nous proposons d'utiliser les marqueurs obtenus par l'opérateur h-minima [19] avec $h = 3$. Dans le type de scènes que nous avons traitées (plusieurs ordres de profondeur, plusieurs cibles), des valeurs inférieures à cinq n'entraînent généralement pas une perte d'information trop importante et accroissent considérablement la vitesse de traitement. Si l'on a besoin de maîtriser le coût de l'algorithme, on peut aussi fixer le nombre de régions du niveau le plus bas, en choisissant les k marqueurs les plus judicieux [1].

On peut aussi constater que lorsque l'on progresse dans la hiérarchie, au dessus d'un certain niveau, les régions obtenues correspondent aux régions homogènes les plus grandes de la scène (comme le sol) puis sont de moins en moins significatives à mesure qu'elles regroupent de plus en plus de sous-régions avec des caractéristiques différentes. Ces régions sont peu intéressantes du point de vue de l'analyse des objets en mouvement et peuvent être écartées de la recherche.

Le paramètre β permet de limiter l'influence du bruit et des mouvements considérés comme secondaires (les ombres des cibles, les reflets et les variations mineures de l'éclairage). Dans nos expérimentations, des valeurs de β de l'ordre de 5 ont été utilisées. Au cours du processus de détection, le seuil T_{score} détermine la proportion minimum de points de contour mobiles nécessaire pour étiqueter une région comme étant en mouvement (c.f. section 4.2). La valeur de ce paramètre est en partie déterminée par la valeur de β. Pour $\beta = 5$, $T_{score} = 0.65$ donne les résultants les plus satisfaisants. Les valeurs de ces paramètres sont stables pour une scène déterminée. De faibles variations des paramètres β et T_{score} autour des valeurs proposées entraînent de faibles modifications sur la solution obtenue.

Les résultats présentés ont été obtenus en appliquant la méthode proposée précédement suivie d'une étape de post-traitement. Cette étape permet, tout d'abord, la suppression des régions en mouvement de petites tailles (inférieure à 100 pixels). Dans un second temps, les régions restantes sont fusionnées en accord avec le critère de dissimilarité précédemment choisi.

Comme le montrent ces résultats la validité du système a été vérifiée dans différentes conditions. L'algorithme répond correctement, que les régions soient texturées ou homogènes. Les personnes isolées sont convenablement détectées et segmentées. Néanmoins, on peut constater que lorsque plusieurs régions proches sont en mouvement, l'espace entre celles-ci peut être partiellement détecté. Ces régions subissent elles aussi des changements très importants au niveau de leurs contours, puisqu'elles sont entourées de régions en mouvement. Comme le montrent les résultats, l'algorithme reste cependant valable lorsqu'il s'agit de détecter les groupes dans leur ensemble. La structure de graphe hiérarchique de la solution peut ensuite prendre le relais afin d'isoler les différents sous-éléments du groupe repéré. Il a pu être constaté que la méthode était relativement robuste aux reflets : par exemple, sur les figures 9, c, e et 9, a, b les reflets (respectivement) sur la vitre de l'escalator et la porte du fond ne sont pas détectés. Les problèmes liés aux reflets s'amplifient cependant lorsque l'on considère des images plus éloignées dans le temps dans le calcul du mcm. La méthode reste sensible aux ombres portées très marquées mais parvient néanmoins à gérer les ombres plus estompées ainsi que certaines variations d'éclairages (les lampes dans les figures 9, c, d et les fenêtres de la rame dans les figures 10, m, p).

6 Conclusions

Cet article se concentre sur l'extraction des objets en mouvement dans le contexte de la vidéosurveillance. L'objectif est de détecter toutes les zones d'intérêt potentielles et d'en créer une représentation adaptée à une analyse ultérieure. Dans un premier temps nous avons introduit une méthode pour extraire les contours des objets en mouvement. Ceux-ci sont extraits à l'aide d'un opérateur basé sur la différenciation des gradients spatiaux de trois images successives. Cet opérateur est robuste à l'amplitude du déplacement des objets et au bruit aléatoire présent dans les images. Nous montrons ensuite comment une segmentation hiérarchique de l'image peut être utilisée afin de déduire les régions en mouvement, à partir des gradients en mouvement avec un faible coût de calcul. Afin d'obtenir les régions en mouvement deux critères sont combinés :

i) un critère de contraste qui permet d'extraire les régions les plus représentatives de l'image.

ii) un critère permettant de mesurer les mouvements au niveau des contours de l'image, utilisé pour sélectionner les régions mobiles dans la hiérachie.

La segmentation hiérarchique permet non seulement une analyse de l'image à divers niveaux de résolution mais ré-

duit aussi le temps de calcul de l'algorithme, qui est le facteur limitant dans les applications de vidéosurveillance.

Un autre point fort de la méthode est qu'elle permet d'extraire les cibles en mouvement sans aucun calcul de flot optique, ni aucune information *a priori* sur la scène ou les objets à détecter. De plus les objets en mouvement sont extraits sous la forme d'agrégats de régions homogènes, de tailles et de contrastes différents. Grâce à la structure de graphes hiérarchiques sous-jacente, leur relations d'adjacence et d'inclusion sont connues. Ces considérations sont très utiles dans le but de modéliser les objets détectés. Ce modèle peut ensuite être exploité dans des étapes ultérieures de l'interprétation de scènes, comme le suivi, la gestion des occultations lorsque deux cibles se croisent, ou la reconnaissance d'objets. Par conséquent, l'étape suivante de notre travail consistera en l'étude d'une description basée sur des hiérachies de graphes pour l'interprétation de la scène, et le suivi d'objets dans le domaine de la vidéosurveillance.

Références

[1] C. Vachier et F. Meyer, Extinction Values: A New Measurement of Persistence, IEEE Workshop on Non Linear Signal/Image Processing, p 254-257, Juin 1995.

[2] P. Salembier et L. Garrido, Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval, IEEE Transactions on Image Processing, volume 9(4), p 561-576, Avril 2000.

[3] F. Zanoguera, B. Marcotegui et F. Meyer, A Toolbox for Interactive Segmentation Based on Nested Partitions, ICIP, 1999

[4] V. Agnus, Segmentation spatio-temporelle de séquences d'images par des opérateurs de morphologie mathématique, thèse de doctorat, LSIIT, Université Louis Pasteur, Strasbourg, Octobre 2001

[5] S. Beucher, Segmentation d'images et morphologie mathématique, thèse de doctorat, Ecole des Mines de Paris, Cahiers du centre de Morphologie Mathématique, Fascicule 10, Juin 1990.

[6] S. Sclaroff et L. Liu, Deformable Shape Detection and Description via Model-Based Region Grouping, IEEE-PAMI, volume 23, p 475-489, 2001.

[7] D.S. Zhang et G. Lu, Segmentation of Moving Objects in Image Sequence: A Review, Circuits, Systems and Signal Processing (Special Issue on Multimedia Communication Services), volume 20(2), p 143-183, 2001.

[8] S. Andra, O. Al-Kofahi, R.J. Radke, et B. Roysam, Image Change Detection Algorithm : A systematic Survey, IEEE Transactions on Image Processing, Juin 2003.

[9] M. Piccardi, Background subtraction techniques: a review, IEEE SMC 2004 International Conference on Systems, Man and Cybernetics, Octobre 2004.

[10] M.J. Black et P. Anandan, The robust estimation of multiple motions : Parametric and piecewise-smooth flow fields, CVIU, volume 63, p. 75-104, Janvier 1996.

[11] R. Deriche, Fast algorithms for low-level vision, IEEE-PAMI, volume 12(1), p 78-87, 1990.

[12] T. Viéville et O.D. Faugeras, Robust and fast computation of unbiased intensity derivatives in images, In Proc. EECV'92, p 203-211, Mai 1992.

[13] N. Paragios et R. Deriche, A PDE-based Level Set Approach for Detection and Tracking of Moving Objects, ICCV, Janvier 1998.

[14] S. Jehan-Besson et M. Barlaud, DREAM2S: Deformable Regions Driven by an Eulerian Accurate Minimization Method for Image and Video Segmentation, IJCV, volume 53(1), p. 45-70, 2003.

[15] J. Shi et J. Malik, Motion segmentation and tracking using normalized cuts, University of California, Berkeley report UCB/CSD-97-962, 1997.

[16] R. Cucchiara, C. Grana, M. Piccardi et A. Prati, Detecting Moving Objects, Ghosts and Shadows in Video Streams, IEEE-PAMI, volume 25(10), p. 1337-1342, 2003.

[17] K. Toyarea, J. Krumm, B. Brumitt et B. Meyers, Wallflower : Principles and practice of background maintenance, ICCV, p. 255-261,1999

[18] K. Yoshinari et M. Michihito, A human motion estimation method us using 3-successive video frames, Intl. Conf. on Virtual Systems and Multimedia, p. 135-140, 1996.

[19] T.H. Kim, Y.S. Moon, A New Flat Zone Filtering Using Morphological Reconstruction Based on the Size and Contrast, VLBV, 1999

[20] S.M. Haynes et R.C. Jain, Time Varying Edge Detection, ICPR82, 1982.

[21] A. Elgammal, David Harwood et Larry Davis, Nonparametric Model for Background Subtraction, ECCV, 2000.

[22] P. Rosin, Thresholding for Change Detection, CVIU, volume 2, p. 79-95, Mai 2002.

[23] S.H. Kim, D.O Kim, J.S. Kang, J.H. Song et R.H. Park, Detection of moving edges based on the concept of entropy and cross-entropy, SPIE High-Speed Imaging and Sequence Analysis III, volume 4308, p. 59-66, Janvier 2001.

[24] V.T. VU, F. Bremond et M. Thonnat, Automatic Video Interpretation: A Recognition Algorithm for Temporal Scenarios Based on Pre-compiled Scenario Models, CVS03, p. 523, 2003.

[25] Projet SAMSIT : Système d'Aide à la Maîtrise de la Sécurite Individuelle dans les Transports publics - Alstom, CEA, INRETS, INRIA, SNCF - PREDIT 2003

FIG. 9 – *Première colonne : images originales, deuxième colonne : résultats superposés.*

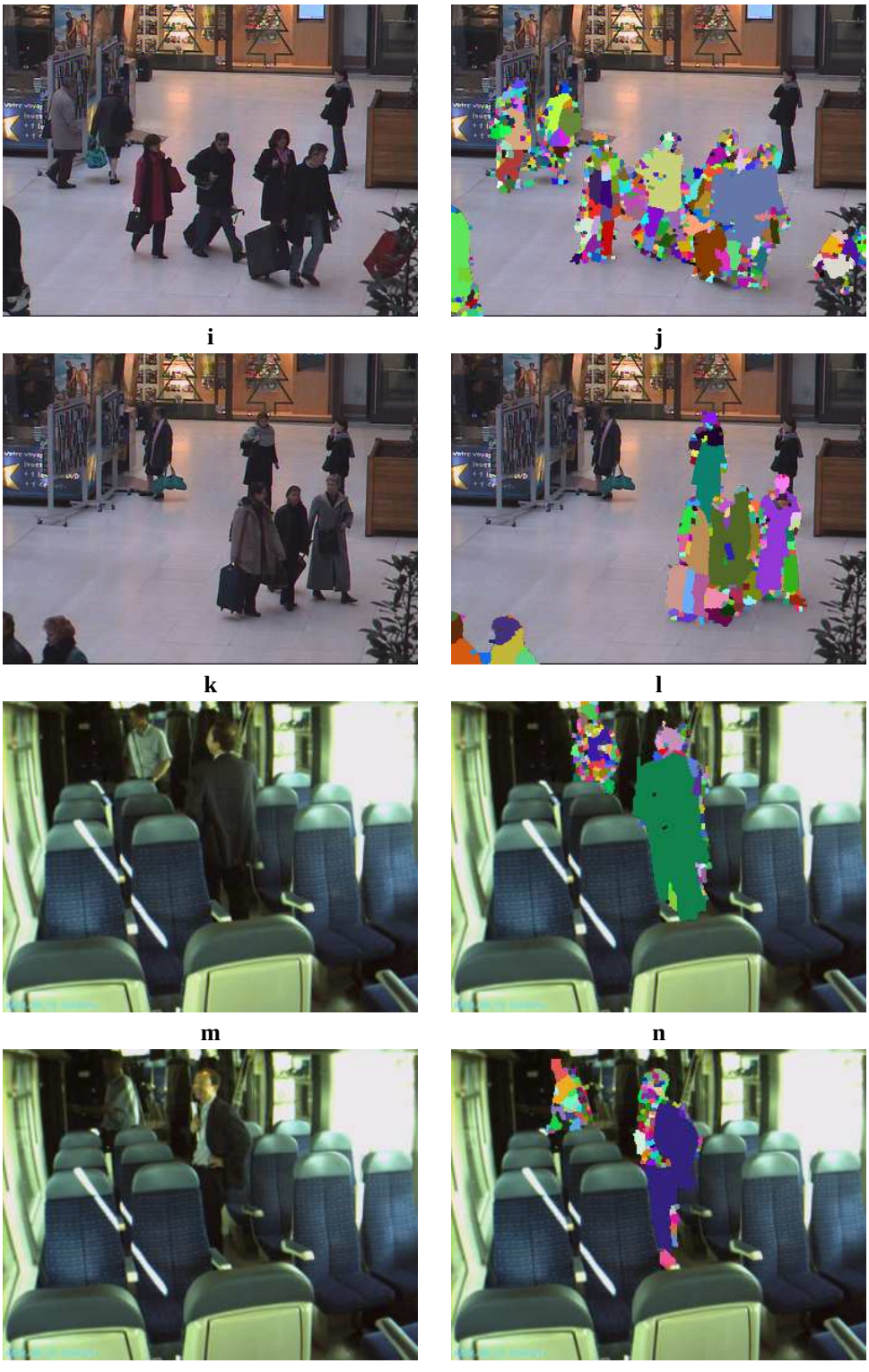


FIG. 10 – *Première colonne : images originales, deuxième colonne : résultats superposés (les images m à q sont extraites du projet SAMSIT [25]).*