

DETECTION DES OBJETS MOBILES DANS LES SCENES NATURELLES.

Maciej ORKISZ (POLOGNE)

Laboratoire TIRF
INPG 46, av. Félix Viallet
38031 GRENOBLE cedex

RESUME

Nous abordons le problème de détection des objets en mouvement dans des images prises à l'aide d'une caméra fixe. Il n'y a pas de restriction quant à la vitesse de déplacement de l'objet étudié. En revanche, d'éventuels changements d'éclairage sont supposés progressifs. Aucune connaissance préalable de la scène, pas plus que son image prise en absence de l'objet mobile, n'est disponible. Ainsi, la différence d'images, qui permet d'éliminer le fond stationnaire, doit s'associer à une autre opération, pour distinguer la position actuelle de l'objet en mouvement. Quelques méthodes intéressantes sont passées en revue. Les algorithmes accumulatifs permettent de reconstruire l'image de référence représentant uniquement les objets stationnaires, et donc de localiser l'objet en mouvement par la différence absolue entre l'image courante et l'image de référence. Mais le résultat n'est disponible qu'après l'analyse de nombreuses trames. La coïncidence des contours fournit un résultat dès la seconde trame, mais le fond désocculé est interprété comme "en mouvement", même s'il était visible au début de la séquence d'images. Les méthodes proposées visent à concilier les avantages des deux catégories citées et à améliorer la détection des contours coïncidents. Grâce à une opération originale, la coïncidence est déterminée pour les gradients, ce qui permet de localiser l'objet en mouvement dès la seconde image. Les contours statiques visibles dans au moins deux trames consécutives sont accumulés, pour former une image historique du fond et éviter ainsi leur mauvaise interprétation lors des désocclusions. Les techniques développées pour les contours non binaires sont extrapolées aux contours binaires détectés en temps réel. Les applications possibles sont brièvement discutées.

SUMMARY

We discuss the problem of moving objects detection in pictures taken by means of a fixed camera. There is no restriction relative to the studied object's displacement velocity. On the other hand, eventual changes of brightness are supposed gradual. Nor prior knowledge of the scene, neither its image taken without any moving object in it, is available. Hence image difference, which eliminates stationary back- and foreground, needs to be followed by another operation, in order to distinguish the moving object's current position. A survey of some interesting methods is given. Accumulative algorithms allow reference frame reconstruction. Since this reference picture contains only stationary objects, moving ones can be displayed by an absolute difference between it and the current frame. But the result is available only after an analysis of numerous frames. Edge coincidence gives a result from the second frame, but disoccluded background is interpreted as "moving", even if it was visible in the beginning of the image sequence. The proposed methods attempt to conciliate the advantages of the cited two categories and to improve the detection of the coincidence edges. Thanks to an original operation, coincidence is found for gradients. Moving objects are thus displayed from the second frame. Static edges visible at least in two consecutive images are accumulated in order to form a historic frame of the background and, consequently, to avoid their misinterpretation in the case of disocclusion. The techniques developed for non-binary edges are extrapolated on binary real-time detected edges. Possible applications are briefly discussed.

Introduction.

Les images représentant des scènes naturelles sont généralement très complexes. Interpréter tous les objets visualisés, alors que seulement très peu doivent être étudiés, serait une perte de temps. Il faut donc extraire l'information pertinente à l'aide d'un prétraitement approprié. Les objets en mouvement, filmés par une caméra fixe, peuvent être repérés grâce à leur déplacement par rapport au fond immobile. Dans la première section de ce texte nous passons en revue quelques techniques intéressantes, qui ont inspiré notre travail, ainsi que nos résultats antérieurs. La seconde partie présente le fruit de

nos recherches les plus récentes.

Méthodes connues.

Intuitivement, un objet en mouvement peut être détecté par une simple opération de soustraction pixel par pixel, effectuée à partir de deux images consécutives. En effet, pour les parties immobiles de la scène, la différence entre les deux images est proche de zéro, les écarts étant provoqués uniquement par les changements d'éclairage et le bruit de quantification. A l'objet étudié, en revanche, correspond une différence plus importante. Cependant, il ne suffit pas de seuiliser le résultat obtenu, pour localiser



l'objet en mouvement. Dans la zone correspondant à sa position précédente, la différence d'images est très semblable à celle calculée pour la position courante. Le signe ne peut pas nous aider à déterminer quelle partie de l'image différentielle représente l'objet en mouvement dans sa position actuelle, car pour les scènes complexes il n'est pas possible de savoir a priori si l'objet étudié sera plus clair ou plus foncé que le fond. De plus, la soustraction donne un résultat proche de zéro non seulement pour les parties immobiles de la scène, mais également dans la zone correspondant à l'intersection des positions courante et précédente de l'objet en mouvement. Par conséquent, ce dernier n'est visualisé que partiellement (Fig.1.).

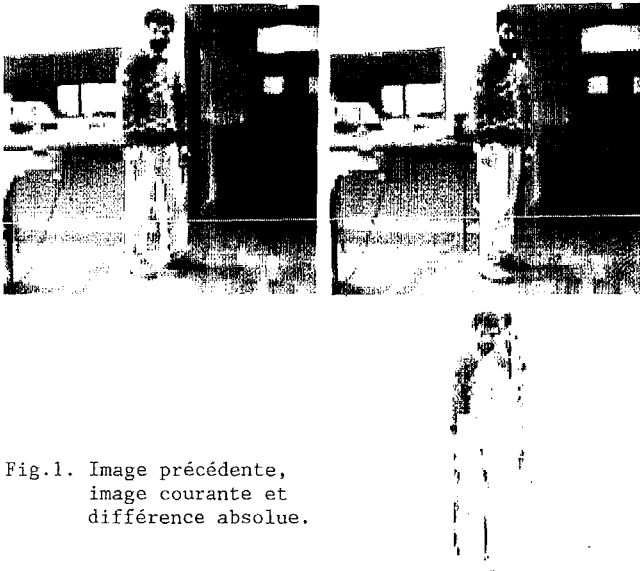


Fig.1. Image précédente, image courante et différence absolue.

1. Techniques accumulatives.

YACHIDA et al.[1] proposent une méthode permettant de reconstruire l'objet entier par concaténation de ses fragments détectés dans les différences entre l'image analysée et un certain nombre d'images consécutives. Cet algorithme, très simple de part sa conception, est très long à exécuter. Il nécessite la connaissance a priori des caractéristiques lumineuses de la scène, car seul le signe de la différence d'images est utilisé pour distinguer la position actuelle de la position suivante de l'objet en mouvement.

Le seul cas dans lequel cette connaissance initiale n'est pas nécessaire est celui de l'image de référence [2]. En effet, si l'on dispose d'une image qui représente la scène sans objets mobiles, la différence entre cette image et la trame étudiée suffit pour localiser correctement l'objet en mouvement, quelque soit la complexité du fond. Il faut cependant signaler que dans certains cas il n'est pas possible de filmer la scène en absence d'objets mobiles, et que dans bien d'autres l'éclairage change après la prise d'image de référence - suffisamment pour qu'elle ne soit plus compatible avec la séquence filmée par la suite. JAIN et NAGEL [3] ont développé une technique accumulative permettant de reconstruire cette image-étalon par suppression de l'objet en mouvement. Les valeurs des niveaux de gris dans la zone correspondant à cet objet sont remplacées par celles du fond, prises dans l'image dans laquelle l'objet s'est déplacé assez loin pour ne plus avoir d'intersection avec sa position initiale. Ceci signifie que le résultat ne peut être obtenu qu'après l'analyse de plusieurs trames. De ce fait, ainsi qu'à cause de la complexité des calculs effectués pour chaque image, cette méthode est très lente.

La nécessité de traiter un nombre considérable d'images avant que le premier résultat ne soit accessible est le défaut commun des méthodes purement accumulatives. Généralement, on essaie d'extraire

l'information pertinente en utilisant deux images consécutives. Mais dans ce cas, l'élimination des parties immobiles par la différence d'images doit être suivie d'une autre opération permettant de distinguer la position actuelle de l'objet en mouvement. La plus appropriée semble être la coïncidence de contours.

2. Coïncidence de contours.

Lorsque l'on utilise les contours, peu importe quels objets dans la scène sont plus clairs et lesquels plus foncés. Les contours de l'objet étudié se trouvent aussi bien dans l'image courante que dans la différence entre elle et la précédente. Ils sont les seuls à avoir cette propriété. En effet, les autres contours présents dans l'image courante correspondent au fond, alors que dans la différence d'images les parties immobiles sont absentes. La détection de l'objet en mouvement peut donc se résumer à la recherche des contours coïncidents dans l'image courante et dans la différence entre elle et la précédente.

Les méthodes fondées sur ce principe diffèrent par la façon dont les contours sont extraits et par l'opération utilisée pour détecter la coïncidence. LEUNG et YANG [4][5] appliquent à l'image courante une segmentation statique. Les limites entre les régions ainsi déterminées sont considérées comme contours. En revanche, la différence d'images est d'abord seuillée. Ainsi les contours sont localisés à la limite entre les zones dans lesquelles sa valeur est supérieure à un seuil fixé et celles où elle est inférieure. Ces contours binaires peuvent se prêter aux opérations logiques, et en particulier, la coïncidence peut être détectée par un simple ET. Cependant, la segmentation statique de l'image entière et le calcul de l'histogramme pour fixer le seuil adaptatif ne favorisent pas la rapidité de la méthode citée.

STELMASZYK et al.[6] n'utilisent pas les contours binaires, mais les gradients calculés à l'aide de l'opérateur de Sobel. Cette méthode est élégante et homogène, car le même opérateur est appliqué à l'image courante et à la différence d'images. La coïncidence est détectée par la multiplication des gradients obtenus. Malheureusement, seules les zones à contraste fort sont ainsi visualisées. Nous avons proposé de remplacer le seuillage du produit par la poursuite de son maximum local [7]. Cette démarche a pour objectif de visualiser également les parties faiblement contrastées de l'objet mobile, et d'éviter de faire tous les calculs systématiquement pour l'image entière. Les gradients et leur produit ne sont calculés que là où la différence entre les deux images est suffisamment grande. La poursuite du contour commence lorsque le produit est supérieur à un seuil, et s'arrête lorsque ce même produit est inférieur à un autre seuil - ce qui signifie que le contour a disparu dans le bruit - ou lorsque le contour est fermé. Le nombre de calculs est ainsi réduit, à condition que l'objet en mouvement n'ait pas de texture nette.

Aucune de ces trois méthodes ne réutilise les contours détectés dans l'image courante pour analyser l'image suivante. Cependant, l'utilisation des gradients donne la possibilité de le faire, ce qui permet un gain de temps considérable. En effet, le gradient de la différence est mathématiquement équivalent à la différence des gradients. Il n'est donc pas nécessaire de mémoriser l'image courante, pour la soustraire à la suivante et pour calculer le gradient du résultat. A sa place, on stocke son gradient qui sera soustrait directement à celui de l'image suivante.

Néanmoins, l'utilisation des gradients a jusqu'à présent un défaut considérable que nous avons déjà évoqué: leur multiplication, ayant pour objectif de détecter les contours coïncidents, met en évidence uniquement les parties bien contrastées. Effectivement, à ces endroits le résultat est grand, car les valeurs du gradient autant dans l'image courante

que dans la différence d'images sont élevées. Par contre, là où le contraste est faible la multiplication de deux petites valeurs du gradient donne souvent un résultat inférieur au produit d'un gradient fort par le bruit dans les zones correspondant au fond. Dans ces cas, même la poursuite du maximum local du produit peut être déroutée. Nous avons mis au point une autre technique de détection de la coïncidence des contours.

Méthodes proposées.

En décrivant l'utilisation des contours binaires déterminés en temps réel [7], nous avons remarqué que l'opération logique ET était le détecteur de coïncidence par excellence. Malheureusement, la méthode citée est très sensible au bruit à cause de deux opérations non-linéaires: seuillage et affinement, effectuées par le processeur câblé pour localiser les contours. L'utilisation des gradients repousse ces opérations à la fin du procédé, mais cela nécessite de trouver la coïncidence des contours dont les valeurs ne sont pas "tout ou rien". La technique proposée s'inspire directement de ET logique. Elle est fondée sur les observations suivantes:

- la soustraction du gradient de l'image précédente par rapport à celui de l'image courante élimine les contours des objets stationnaires, et affecte peu ceux de l'objet en mouvement;
- par conséquent, pour les contours des objets immobiles, le résultat de cette soustraction est très différent par rapport aux valeurs du gradient de l'image courante;
- en revanche, là où l'écart entre ces valeurs est faible il s'agit de l'objet en mouvement, ou éventuellement, de zones sans contraste. Mais le gradient dans ce dernier cas est quasiment nul.

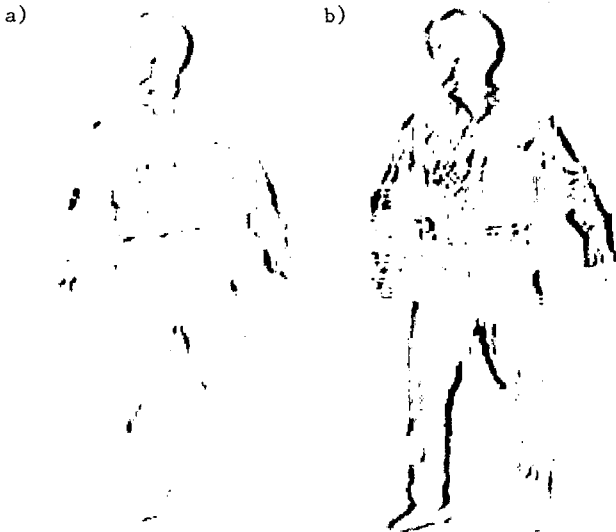


Fig.2. Coïncidence des contours non-binaires:
a) produit des gradients, b) heuristique proposée.

Pour détecter l'objet en mouvement, il faut donc chercher les endroits où le gradient de l'image courante est suffisamment grand et à peu près égal à la différence des gradients des images courante et précédente (Fig.2.). Le seuillage est donc précédé par l'opération suivante:

SI $\min(G_j, \text{abs}(G_j - G_i)) > \max(G_j, \text{abs}(G_j - G_i)) / \text{coef}$
ALORS résultat := $\max(G_j, \text{abs}(G_j - G_i))$
SINON résultat := 0

où G_i, G_j sont les gradients des images: précédente et courante. Le coefficient qui exprime ce que nous entendons par "à peu près égal" et qui sert de discriminateur entre le fond et l'objet en mouvement a été expérimentalement fixé à 2. En effet, les résultats obtenus pour cette valeur sont satisfaisants et la division peut être remplacée par un décalage à

droite qui est plus rapide. Pour les coefficients plus grands ayant cette propriété, l'élimination du fond est moins bonne, alors que les valeurs comprises entre 1 et 2 entraîneraient la division à virgule flottante.

La méthode présentée permet de détecter la coïncidence de contours non-binaires. Mais, comme tout algorithme fondé sur la comparaison de seulement deux images consécutives, elle interprète mal les objets appartenant au fond lorsqu'ils réapparaissent après une occultation. Effectivement, puisqu'ils sont présents dans l'image courante là où ils étaient absents dans la précédente, ils sont classés parmi les objets en mouvement. Nous avons proposé d'accumuler l'information, afin de construire une image de référence représentant tous les objets stationnaires qui ont été visibles depuis le début de la séquence d'images [7]. Cette image "historique" est initialisée avec les contours immobiles détectés dans la première trame. La contribution de chaque trame suivante se fait par une somme pondérée de ses contours statiques et de l'image de référence. Le choix des poids constitue un compromis entre l'effet mémoire - permettant de sauvegarder les contours occultés pendant assez longtemps - et l'effet d'apprentissage qui a pour objectif d'enrichir l'image historique avec de nouveaux éléments. Ce compromis est une limitation de la méthode citée, car les objets occultés finissent par disparaître de l'image de référence s'ils sont cachés trop longtemps. De même, les éléments nouveaux sont interprétés comme "en mouvement" pendant plusieurs trames, car l'apprentissage est progressif.

Ces remarques nous conduisent à proposer une technique nouvelle, fondée sur les principes suivants:

- 1° les valeurs du gradient correspondant au fond occulté ne doivent pas être modifiées dans l'image historique;
- 2° les nouveaux éléments stationnaires détectés dans l'image courante doivent être immédiatement rajoutés sans atténuation;
- 3° les valeurs du gradient correspondant au fond visible évoluent suivant les changements d'éclairage.

Les contours peuvent être classés dans une des catégories énumérées, par comparaison entre l'image historique H_i et les contours stationnaires F_j détectés dans l'image courante. Ainsi, les catégories citées se caractérisent respectivement par:

- 1° $H_i \gg F_j$
- 2° $H_i \ll F_j$
- 3° $H_i \approx F_j$

où F_j peut être déterminé comme $\min(G_i, G_j)$, ainsi que nous l'avons démontré dans [7]. Une autre façon de détecter dans l'image courante les contours qui n'ont pas bougé par rapport à l'image précédente, consiste à trouver dans les deux images les endroits où leurs gradients ont à peu près la même valeur, ce qui peut être formalisé par l'expression suivante:

SI $\min(G_i, G_j) > \max(G_i, G_j) / 2$ ALORS $F_j := G_j$
SINON $F_j := 0$

Cette version est plus lente, mais a l'avantage de donner $F_j = 0$ pour l'objet en mouvement. L'évolution de l'image de référence suivant les changements d'éclairage - supposés progressifs - peut être assurée de diverses façons. La nouvelle valeur H_j peut être une somme pondérée de H_i et de F_j ou peut, tout simplement, prendre la valeur de F_j . Nous avons choisi de prendre la moyenne de H_i et de F_j . La mise à jour de l'image historique s'effectue donc comme suit:

SI $\min(H_i, F_j) > \max(H_i, F_j) / 2$ ALORS $H_j := (H_i + F_j) / 2$
SINON $H_j := \max(H_i, F_j)$

Etant donné que la détection de l'objet en mouvement



se fait par rapport à l'image de référence, l'opération décrite doit s'effectuer avant cette détection. Sinon, chaque nouvel élément du fond serait interprété comme un objet en mouvement deux fois, au lieu d'une, avant d'être incorporé dans l'image historique. Les opérations décrites forment l'algorithme suivant:

```
* initialisation de l'image historique: H:=0;
* acquisition de la première image I1;
* calcul de son gradient G1;
* REPETER
  * acquisition de l'image suivante I2;
  * calcul de son gradient G2;
  * détection de la partie immobile: F:=min(G1,G2);
  * mise à jour de l'image historique:
    SI min(H,F) > max(H,F)/2 ALORS H:=(H+F)/2;
    SINON H:=max(H,F);
  * détection et affichage de l'objet en mouvement:
    SI min(G2,abs(G2-H)) > max(G2,abs(G2-H))/2
      ALORS R:=max(G2,abs(G2-H));
    SINON R:=0;
  * copie G1:=G2;
JUSQU'A critère d'arrêt;
```

Pour les contours binaires, il prend évidemment une forme plus simple. La mise à jour de l'image historique et la détection de l'objet en mouvement se font respectivement par les opérations suivantes:

```
H := H ou (G1 et G2);
R := G2 et (G2 ou exclusif H);
```

Ces formules binaires sont exécutées pour les octets entiers, donc pour huit pixels à la fois. Ceci et, plus encore, l'utilisation du détecteur de contours en temps réel (GTR) conçu dans notre laboratoire par CHEHIKIAN et al.[8], assurent une grande rapidité de cette méthode. Comparé à la coïncidence binaire sans accumulation, ce dernier algorithme non seulement résout le problème des désocclusions, mais est également moins sensible au bruit. En effet, grâce à l'apprentissage, les variations des contours dues aux erreurs de quantification sont progressivement enregistrées, ce qui est visible sur la figure 3 sous forme de traits plus épais dans l'image historique que dans l'image courante. Par conséquent, il arrive de moins en moins fréquemment qu'un contour d'un objet stationnaire ne coïncide pas avec l'image de référence et donc soit considéré comme "en mouvement".



Fig.3. Contours binaires détectés en temps réel : a) image précédente, b) image courante, c) objet en mouvement, d) image historique.

Les méthodes présentées concilient les avantages de deux techniques: coïncidence des contours et accumulation de l'information. Elles donnent un résultat dès la seconde trame et construisent en même temps une image de référence. Grâce à cette image,

l'opération de détection des contours coïncidents peut être supprimée au bout d'un certain temps, car alors la différence absolue suffit pour localiser l'objet en mouvement.

Conclusion.

Nous avons discuté ici quelques aspects de la détection d'objets en mouvement dans des scènes naturelles. La différence d'images élimine le fond stationnaire. La technique de détection des contours coïncidents permet de localiser l'objet étudié sans connaissance préalable des niveaux de gris de la scène. Enfin, l'accumulation de l'information sert à reconstruire l'image de référence et éviter ainsi les problèmes d'interprétation du fond désoculté. Les prétraitements décrits peuvent constituer le premier pas de nombreuses analyses d'images dynamiques, allant du contrôle du trafic routier aux études microbiologiques. L'application qui nous intéresse particulièrement concerne le mouvement humain. La détection du corps en mouvement devrait être suivie d'une segmentation appropriée permettant de repérer les points caractéristiques: les articulations, par exemple. La poursuite de ces cibles dans les trames suivantes pourrait se faire en utilisant, par exemple, la méthode décrite par SAGE [9]. Cependant, la détection des articulations sans port de marqueurs spéciaux par la personne filmée est un problème difficile dont l'étude doit être encore approfondie.

REFERENCES

- [1] M.YACHIDA, M.ASADA, S.TSUJI, "Automatic Analysis of Moving Images", IEEE Trans. on Pattern Anal. and Machine Intell., Vol. PAMI-3, No 1, January 1981, pp 12-20.
- [2] T.TSUKIYAMA, Y.SHIRAI, "Detection of the Movements of Persons from a Sparse Sequence of TV Images", Pattern Recognition, Vol.18, No 3/4, 1985, pp 207-213.
- [3] R.JAIN, H.-H.NAGEL, "On the Analysis of Accumulative Difference Pictures from Image Sequences of Real World Scenes", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-1, No 2, April 1979, pp 206-214.
- [4] M.K.LEUNG, Y.-H.YANG, "Human Body Motion Segmentation in a Complex Scene", Pattern Recognition, Vol.20, No 1, 1987, pp 55-64.
- [5] M.K.LEUNG, Y.-H.YANG, "A Region Based Approach for Human Body Motion Analysis", Pattern Recognition, Vol.20, No 3, 1987, pp 321-339.
- [6] P.STELMASZYK, P.BONNET, J.C.POSTAIRE, "Analyse de Scènes Dynamiques par Recherche des Contours en Mouvement", 5ème Congrès Reconnaissance de Formes et Intelligence Artificielle AFCET/ADI/INRIA, Grenoble (France), Nov.1985.
- [7] M.ORKISZ "Image Difference Based Methods for Moving Objects Detection in Real World Scenes", Proceedings of 7th International Symposium IASTED Applied Informatics, Grindelwald, Feb. 1989, pp 80-83.
- [8] A.CHEHIKIAN, J.HERAULT, J.E.MENDEZ, "Un Processeur de Détection de Contours en Temps Réel dans une Image de Télévision", Annales des Télécommunications, tome 38, No 3/4, Mars-Avril 1983.
- [9] D.SAGE, "Poursuite d'Objets par Analyse d'une Séquence d'Images Couleur", thèse de docteur de l'INPG, LTIRF, Grenoble - France, 1989.