

# Segmentation Spatio-temporelle de Séquences d'Images

Jean-Pierre Gambotto

MS2I

3, Avenue du Centre  
78182 St-Quentin-en Yvelines Cedex

## RESUME

Cet article décrit une méthodologie pour segmenter une séquence d'images à l'aide de l'information de mouvement. Un paramètre important pour l'analyse du mouvement à partir d'une séquence d'images est la continuité spatio-temporelle. Lorsque la fréquence trame est élevée, la séquence d'images, considérée comme un volume tridimensionnel, peut être segmentée en domaines spatio-temporels homogènes.

Nous proposons l'utilisation d'un algorithme de regroupement de régions pour effectuer cette segmentation. Nous analysons ensuite le cas où les régions sont petites et nous décrivons un algorithme pour la segmentation de cibles ponctuelles en mouvement. Cet algorithme utilise une pyra-

mide spatio-temporelle pour définir le voisinage 3D dans la séquence et exploite la continuité locale des trajectoires pour segmenter les cibles.

## SUMMARY

This paper describes an approach for the segmentation of an image sequence using motion information. The spatio-temporal continuity of the image sequence is an important parameter for motion analysis. When the frame rate is high, the segmentation of the image sequence into spatio-temporal regions appears to be feasible.

This segmentation can be performed with a region growing algorithm. We analyze the case of small regions in details and present an algorithm for the segmentation of moving point targets in an image sequence. A spatio-temporal pyramid is used to describe the 3D neighborhood in the image sequence

and the segmentation is based on the local continuity of the trajectories.



## 1 Introduction

Cet article propose une méthodologie pour segmenter une séquence d'images à l'aide de l'information de mouvement. Le but de l'étude est le développement d'une chaîne algorithmique permettant à un système de vision d'analyser son environnement à partir d'une séquence d'images. Le mouvement de la camera combiné au mouvement propre de chaque objet, fournit en général une image de la scène présentant des évolutions complexes.

Un paramètre important pour l'analyse du mouvement à partir d'une séquence d'images est la continuité spatio-temporelle. Du fait du bruit et des discontinuités inter-images, l'analyse du mouvement réalisée à partir de deux images consécutives est souvent délicate. L'extraction d'information à partir d'un grand nombre d'images présente plusieurs avantages. Tout d'abord elle permet d'éliminer les cas d'ambiguïté. En effet, Maybank [1] a montré que le problème de la détermination du mouvement à partir de deux vues n'a pas toujours une solution unique. D'autre part, la prise en compte de la continuité temporelle permet d'obtenir des informations plus robustes au bruit.

Des techniques de filtrage spatio-temporel ont été proposées pour la restauration de séquences d'images [3] et pour la détection des contours tridimensionnels (3D). La technique développée par Bolles [7] utilise des contours extraits à partir de coupes du volume spatio-temporel pour reconstruire les paramètres 3D de l'environnement. Des formulations théoriques de la segmentation spatio-temporelle sont proposées dans [2] et [6].

Dans cet article, nous abordons le problème de la segmentation d'une séquence d'images en domaines spatio-temporels homogènes. L'extension des techniques de regroupement de régions au domaine 3D est une voie intéressante à explorer, car la mise en correspondance est alors effectuée simultanément à la segmentation. Nous décrivons ensuite un algorithme pour la segmentation de cibles ponctuelles à partir de leurs mouvements.

## 2 Segmentation en domaines spatio-temporels

### 2.1 La cohérence spatio-temporelle

La perception visuelle d'objets en mouvement implique une certaine cohérence à la fois spatiale et temporelle de la scène. Deux points voisins dans l'image et appartenant au même objet auront des propriétés radiométriques voisines. D'autre part, les changements entre deux images consécutives doivent être progressifs. Cette notion, bien que intuitive, est aussi importante pour la vision par ordinateur. Elle se traduit par des contraintes sur le mouvement des objets:

1. *Vitesse maximale*: la vitesse d'un objet dans l'image est limitée. Elle est donnée par  $(\Delta u, \Delta v) / \Delta t$  où  $\Delta t$  et  $(\Delta u, \Delta v)$  sont respectivement l'intervalle de temps et le déplacement maximal entre 2 trames.
2. *Mouvement commun*: la cohérence spatiale implique que les points d'une même région de l'image ont approximativement le même mouvement.
3. *Évolutions lentes*: les lois de la physique et une fréquence trame élevée impliquent des variations lentes pour la plupart des paramètres caractéristiques des objets. En particulier, les changements de forme et de vitesse (amplitude et direction) restent faibles. D'autre part, les objets ne peuvent apparaître ou disparaître de façon brutale.

La plupart des techniques développées pour l'analyse du mouvement dans les images utilisent une ou plusieurs de ces contraintes. Les techniques fondées sur le calcul local du mouvement [9] nécessitent bien sûr une fréquence trame élevée. Par contre, cette contrainte est à priori moins importante pour les méthodes utilisant une mise en correspondance d'objets [5, 8, 14, 20].

Considérons le problème de la mise en correspondance de régions. Il y a cohérence lorsque deux régions appartenant à des images consécutives et qui se correspondent, s'intersectent (cf. figure 1). Cette propriété est utilisée dans un grand nombre de techniques. Une approche élémentaire consiste à effectuer une corrélation inter-images dans des fenêtres prédéfinies. Deux approches plus élaborées permettent de segmenter des régions de formes quelconques à partir de leurs mouvements. La première est fondée sur une détection des changements entre images consécutives de la séquence. Les zones de changement sont tout d'abord obtenues par soustraction d'images, puis, une analyse de l'image résultat permet de retrouver les objets et leurs mouvements [10, 11]. La seconde approche consiste à analyser les zones inchangées dans l'image. Afin de détecter ces zones, la seconde image est progressivement décalée par rapport à la première avant d'effectuer la soustraction. Cette méthode a été proposée par Anstis [12] puis exploitée par Jayaramamurthy et Jain [13] pour la détection d'objets texturés en mouvement.

Une approche similaire est aussi développée dans [14]. Elle consiste à analyser les intersections entre les régions de deux images consécutives. L'algorithme détecte de façon récursive chaque composante du mouvement dans le plan image. Ensuite, une analyse arborescente permet d'identifier différents types de mouvement (translation, rapprochement, éloignement...) ainsi que certaines transformations spécifiques (masquages progressifs, changements de forme...). Dans cette approche, le calcul du mouvement et la mise en correspondance sont réalisés simultanément.

### 2.2 Regroupements spatio-temporels

Les techniques de regroupement ont été développées pour la segmentation d'une image en régions homogènes [15]. Les pixels de l'image sont regroupés de façon récursive en régions de tailles croissantes. La segmentation finale est obtenue en utilisant des critères basés sur l'homogénéité des régions et sur le gradient à leurs frontières.

L'extension de ces techniques au domaine tridimensionnel (espace+temps) semble envisageable lorsque la cohérence spatio-temporelle est vérifiée. Considérons une région  $R_i$  dans l'image  $I_n$ ; son voisinage spatio-temporel comprend toutes les régions de l'image  $I_n$  qui lui sont adjacentes ainsi que toutes

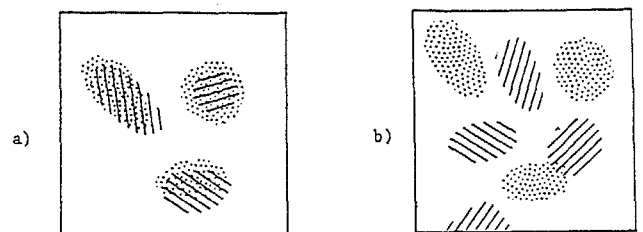


Figure 1: Influence de la fréquence trame sur la cohérence spatio-temporelle: (a) fréquence élevée, (b) fréquence faible. (les objets hachurés sont dans l'image  $n$ ; les objets en pointillés sont dans l'image  $n + 1$ ).

les régions des images  $I_{n-1}$  et  $I_{n+1}$  qui intersectent  $R_i$ . Les regroupements seront donc effectués à la fois dans le plan image et suivant l'axe temporel. Cependant, dans le cas général, l'algorithme de regroupement fournira deux types d'objets:

- des domaines spatio-temporels lorsque l'hypothèse de cohérence est vérifiée,
- des régions homogènes sinon.

En effet, lorsque la région est trop petite, la cohérence spatio-temporelle telle qu'elle a été définie n'existe plus. Dans ce cas, il est nécessaire d'analyser de façon plus précise la continuité temporelle.

### 3 Détection de cibles ponctuelles par analyse spatio-temporelle

Le problème de la détection de cibles ponctuelles se recontrent dans un grand nombre d'applications telles que la poursuite de cible en imagerie infrarouge [18], la surveillance du champ de bataille [21] ainsi que la surveillance du ciel [16]. En général, les méthodes utilisées se composent d'une segmentation de chaque trame de la séquence suivie d'un algorithme de mise en correspondance (souvent appelé algorithme de formation de pistes).

Des techniques similaires sont aussi développées pour la vision des robots mobiles; des points caractéristiques ( par exemple points anguleux des contours) sont tout d'abord extraits de chaque trame puis appariés au cours du temps. La méthode proposée dans [17,19] est fondée sur la continuité du mouvement. La continuité locale est tout d'abord testée sur trois trames consécutives à l'aide d'un algorithme de relaxation, puis un processus global propage la trajectoire trouvée de trame à trame. Dans ces travaux, la mise en correspondance est effectuée uniquement en utilisant l'hypothèse de continuité des trajectoires dans le plan image. Une approche voisine est développée dans [20]; la méthode utilisée exploite directement la continuité du mouvement 3D pour mettre en correspondance des segments de droites.

Nous décrivons dans ce qui suit un algorithme permettant d'effectuer une segmentation spatio-temporelle de cibles ponctuelles et utilisant l'hypothèse de continuité des trajectoires. Considérons une cible ponctuelle en mouvement dont le déplacement inter-trame est supérieur au pixel. Bien que le mouvement de la cible soit continu, la connexité inter-trame n'existe plus. Cette connexité peut être rétablie en reconstruisant le voisinage temporel à l'aide d'une pyramide.

#### 3.1 La pyramide spatio-temporelle

La pyramide spatio-temporelle est représentée sur la figure 2. Elle est formée à partir de P trames consécutives  $I_n$  à  $I_{n+P-1}$ . Le sommet de la pyramide est centré sur le pixel  $(i, j)$  de la première image et la base de la pyramide est un rectangle de dimensions  $U, V$ , où  $U$  et  $V$  sont fonctions du vecteur de déplacement maximal  $(\Delta u, \Delta v)$ . Toute cible située au point  $(i, j)$  au temps  $t = n$  aura sa trajectoire incluse dans le volume de cette pyramide. Cette trajectoire peut être reconstruite en utilisant une structure de voisinage particulière.

#### 3.2 La structure de voisinage

La structure de voisinage utilisée permet d'associer à chaque point  $I_t(i, j)$  de la trame  $m$  un voisinage  $V_t(i, j, m)$  dans la trame  $m - 1$ . Ce voisinage est obtenu en supposant la trajectoire de la cible localement rectiligne. Considérons la droite joignant

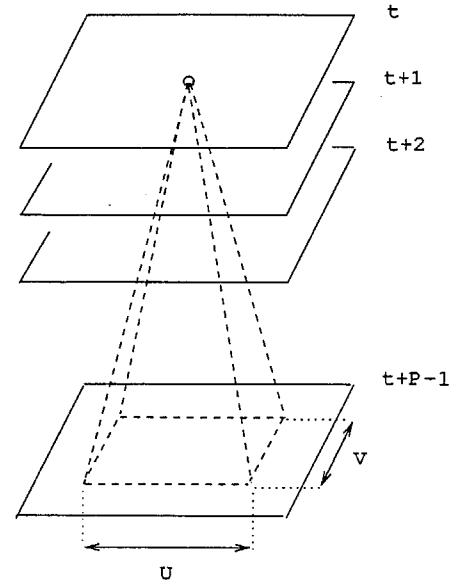


Figure 2: Pyramide spatio-temporelle formée à partir de P images consécutives de la séquence.

le sommet de la pyramide  $P_0$  au pixel  $I_m(i, j)$ . Le vecteur directeur associé à cette droite est donné par:

$$vec(i, j, m) = [I_m(i, j) - P_0] \quad (1)$$

et cette droite intersecte le niveau  $m - 1$  de la pyramide au point:

$$P_{m-1} = [I_m(i, j) - vec(i, j, m)]$$

Le voisinage  $V(i, j, m)$  est alors défini en considérant les pixels voisins du pixel intersecté par le vecteur (1). Ce voisinage comprend suivant les cas 1, 2 ou 4 pixels.

#### 3.3 Détection des cibles

La structure de voisinage permet de définir un ensemble de chemins possibles pour le déplacement de la cible. La détection consiste alors à trouver le chemin optimal, c'est-à-dire celui qui maximise le contraste global. Le chemin optimal peut être obtenu par une méthode de programmation dynamique. En effet, on montre facilement que la structure de voisinage induit des chemins monotones croissants suivant l'axe temporel. Tout sous-chemin du chemin optimal est donc lui-même optimal. Soit  $C_{m-1}^*(k, l)$  le contraste global sur le sous-chemin de  $P_0$  à  $I_{m-1}(k, l)$ , le contraste global au point  $I_m(i, j)$  est donné par:

$$C_m^*(i, j) = \max_{(k,l) \in V(i,j,m)} [C_{m-1}^*(k, l) + CL(i, j)]$$

où  $CL(i, j)$  est le contraste local au point  $(i, j)$ .

#### 3.4 Une application

Nous considérons ici le problème de la détection de cibles ponctuelles brillantes en mouvement dans un fond bruité. Chaque cible est caractérisée par son intensité et sa vitesse. Nous supposons que ces deux paramètres restent constants pendant l'intervalle d'analyse  $[t, t + P - 1]$  et que l'intensité du fond varie lentement dans le plan image. Pour ce problème, le contraste est défini comme étant la différence entre le niveau du pixel et la moyenne locale.

L'algorithme de détection comprend les étapes suivantes:

1. Acquisition de N pixels brillants dans la première image,
- 2. segmentation spatio-temporelle de chaque cible:



- recherche du chemin optimal issu de chaque point brillant dans une pyramide de hauteur P,
- classification des chemins trouvés en utilisant une règle de décision Bayésienne; cette étape permet d'éliminer les fausses alarmes.

Cet algorithme a été appliqué à une séquence d'image synthétique comprenant six images 128\*128. Dans chaque image, le fond est généré à partir d'un bruit blanc gaussien. Des cibles ponctuelles sont ensuite additionnées au fond synthétique. Un tirage aléatoire est utilisé pour déterminer la position initiale ainsi que la vitesse de chaque cible.

Les premières expériences ont montré que le nombre de fausses alarmes diminue lorsque la profondeur d'analyse augmente. Cependant, dans le cas d'un faible rapport signal à bruit, il est nécessaire d'acquérir un grand nombre de points brillants à la première étape de l'algorithme. Une alternative est d'utiliser l'algorithme de détection de façon récursive. Les trajectoires des cibles sont alors initialisées à partir d'un faible nombre de points brillants dans chaque image.

### Remerciements

Cette étude est financée par le contrat DRET No. 88.329

### References

- [1] **Maybank S.**, 'A theoretical study of optical flow', *Phd Thesis*, Birbeck College, Univ. Londres, (1987).
- [2] **Flinchbaugh B.E. and B. Chandrasekaran**, 'A theory of spatio-temporal aggregation for vision', *Artificial Intelligence*, Vol. 17, (1981), pp. 387-407.
- [3] **Samy R.**, 'Adaptive image sequence filtering based on motion detection', *Proc. SPIE Conf. on Optical and Electro-optical Applied Science and Engineering*, Cannes, (Dec. 1985).
- [4] **Monga O. and R. Deriche**, 'A new three dimensional boundary detection', *Proc. Int. Conf. on Pattern Recognition*, Paris, (1986).
- [5] **Pocztar L. and H. Maitre**, 'Restitution d'un flot de particules par mise en correspondance spatio-temporelle d'images', *Congrès AFCET*, Antibes, (Nov. 1987).
- [6] **Huntsberger T.L. and S.N. Jayaramamurthy**, 'Determination of optic field using the spatio-temporal deformation of region properties', *Pattern Recognition Letters*, Vol. 6, (1987), pp. 167-177.
- [7] **Bolles R.C. et al.**, 'Epipolar plane image analysis: an approach to determining structure from motion', *Int Journal of Computer Vision*, Vol. , (1987), pp. 7-55 .
- [8] **Peng S.L. and Medioni G.**, 'Spatio-temporal analysis for velocity estimation of contours in an image sequence with occlusion', *Proc. Int. Conf. on Pattern Recognition*, Rome, (1988), pp. 236-241.
- [9] **Schunck B. G.**, 'Image flow: fundamentals and future research', *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San-Francisco, CA, (1985), pp. 560-571.
- [10] **Jain R. and 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. 1, No. 2, (1979), pp. 206-214.
- [11] **Yalamanchili S. et al.**, 'Extraction of moving object description via differencing', *Computer Graphics and Image processing*, Vol. 18, (1982), pp. 188-201.
- [12] **Antsis S. M.**, 'Phi movement as a substraction process', *Vision Research*, Vol. 10, (1970), pp. 1411-1430.
- [13] **Jayaramamurthy S.N. and R. Jain**, 'Segmentation of textured dynamic scenes', *Proc. IEEE Conf. Pattern Recognition and Image Processing*, Dallas, (1981), pp. 206-214.
- [14] **Gambotto J.P.**, 'Segmentation and interpretation of infrared sequences', in *Advances in Computer Vision and Image Processing*, Vol 3 Huang T.S. Ed., JAI Press (1988), pp 1-38 .
- [15] **Gambotto J.P.**, 'Estimation récursive de la moyenne de signaux bidimensionnels: vers une approche parallèle de la segmentation d'images' *Dixieme Colloque GRETSI*, Nice, (Mai 1985).
- [16] **Mohanty N. C.**, 'Computer tracking of moving point targets in space', *IEEE Trans. on Pattern Recognition and Image Intelligence*, Vol.3, No. 5, (Sep. 1981), pp. 606-611.
- [17] **Sethi I.K. and R. Jain**, 'Finding trajectories of feature points in a monocular image sequence', *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vo.9, No. 1, (1987), pp 56-73.
- [18] **Thomas D. and V. Salari**, 'Motion filtering for target extraction and tracking in infrared images', *Proc. Conf. SPIE Application of Artificial Intelligence*, Vol. 937, (1988) pp. 505-511.
- [19] **Sethi I.K., V. Salari and S Vemuri**, 'Feature point matching in an image sequence', *Pattern Recognition Letters*, Vol. 7, (1988), pp. 113-121.
- [20] **Gambotto J.P.**, 'Tracking points and line segments in image sequences', *Proc. IEEE Workshop on Visual Motion*, Irvine, California, (Mar. 1989).
- [21] **Firschein O. et al.**, 'Track assembly and background suppression using an array processor and neighborhood coding', *Proc. SPIE Real-Time Signal Processing III*, Vol. 241, (1980), pp. 258-260.