



SEGMENTATION SPATIO-TEMPORELLE D'IMAGES ANIMEES BASEE SUR LE FUSIONNEMENT DE REGIONS HOMOGENES SPATIALEMENT

Jenny BENOIS - Ling WU - Dominique BARBA

LATI/IRESTE - La Chantrerie - CP 3003 - 44087 NANTES CEDEX 03

RÉSUMÉ

Dans ce papier, nous proposons une nouvelle approche pour la segmentation spatio-temporelle des séquences d'images animées en zones homogènes au sens du mouvement. Les zones sont construites initialement par fusionnement de régions homogènes spatialement qui elles-mêmes sont construites par segmentation faisant coopérer contours et régions. Ainsi les bords des régions en mouvement s'appuient sur la structure des contours spatiaux et engendrent une bonne qualité de segmentation. On effectue également au cours du temps, le traitement et l'affectation des zones de découvrément importantes suivant une procédure analogue.

1. INTRODUCTION

L'analyse des images animées connaît un essor considérable depuis ces dix dernières années en raison du champ extrêmement vaste des applications potentielles envisageables. Cela concerne d'abord la détection d'objets en mouvement, la caractérisation du mouvement de ces objets en vue de leur suivi dans la scène. De façon plus large, cela concerne la segmentation des séquences d'images animées en vue de leur analyse par extraction et mesure de paramètres informationnels importants par exemple (analyse d'images de cycles cardiaques, temps à collision pour l'aide à la conduite de véhicules).

De la même façon qu'en analyse d'images fixes pour laquelle la segmentation est l'une si ce n'est la phase clé des traitements, la segmentation d'images animées en régions homogènes au sens du mouvement est capitale pour beaucoup d'applications actuelles et futures de l'analyse des images animées. L'une de ces applications potentielles futures qui déborde le cadre strict de l'analyse d'image est celle du codage par analyse / synthèse des images animées. Cette nouvelle approche du codage d'images utilisant des procédés plus structurels voire plus sémantiques s'inscrit d'une certaine façon dans le prolongement direct du codage des images fixes par approche contour/texture [1].

C'est dans ce contexte que nous cherchons à généraliser le principe utilisé en incorporant la structuration liée au mouvement. Les primitives de la segmentation spatio-temporelles seront cohérentes au sens mouvement tout en

ABSTRACT

In this paper, we present a new approach to the segmentation of time varying images into homogeneous regions with respect to the motion. These homogeneous motion regions are built at the beginning by merging spatially homogeneous regions having the same motion. As these regions are segmented by the cooperation between contour and region like procedures, the frontiers of the macroregions correspond to the structure of the contours and provide a good quality of the segmented image. Moreover, we perform along the time the processing of occlusion zones according the same procedure for the large uncovered areas.

s'appuyant dans leur construction sur des primitives cohérentes au sens objet.

Après avoir présenté le principe général de la segmentation spatio-temporelle que nous proposons (section 2) nous décrivons de façon plus précise le principe de la construction des macrorégions homogènes au sens du mouvement (section 3). Nous présenterons les résultats obtenus sur deux séquences d'images animées (section 4).

2. PRINCIPE GENERAL

Le principe général d'une méthode de segmentation spatio-temporelle des images animées est de découper l'espace 3-D formé des deux dimensions spatiales (X,Y) et de celle du temps (T) en domaines spatio-temporels homogènes au sens du mouvement sur une tranche de temps (t_0, t_f) propre à chaque région. Ces régions homogènes dans le mouvement peuvent être constituées schématiquement de deux manières différentes.

On peut utiliser comme critère d'homogénéité un critère dépendant exclusivement de l'erreur de modélisation du mouvement (critère C_1). L'approche markovienne de segmentation d'image au sens purement mouvement est caractéristique de cette manière de procéder.

On peut utiliser à la fois un critère de type C_1 et un critère dépendant purement de l'homogénéité spatiale C_2 . Cela donne des régions homogènes à la fois spatialement et au sens du mouvement. Cela engendre pour des scènes complexes une segmentation spatio-temporelle très fine voire trop riche.

Enfin, on peut définir une segmentation spatio-temporelle s'appuyant sur un critère de type C_1 mais aussi sur



une segmentation spatiale pour construire les régions homogènes au sens du mouvement. L'approche que nous avons développé est de ce dernier type : les régions homogènes au sens du mouvement, que nous appellerons macrorégions MR, sont constituées du fusionnement de régions homogènes spatialement et animées du même mouvement :

$$MR_k^t = \{ \dots, R_{k,i}^t, \dots \}$$

où chacune des régions R est homogène spatialement. L'intérêt majeur de cette approche est que, par construction, les frontières des macrorégions sont, à chaque instant t, des frontières spatiales. Il suffit que dans la segmentation spatiale les contours soient pris en compte pour que les frontières des macrorégions s'appuient en partie sur les contours spatiaux définissant les grandes structures dans l'image animée. Ainsi, la cohérence au cours du temps des frontières des macrorégions est plus assurée. Ce point est très important, en particulier lors de l'initialisation de la segmentation spatio-temporelle et lors de la réactualisation de celle-ci au cours du temps. La méthode de segmentation se décompose en 3 phases successives de fusionnement de régions lors de l'initialisation de la segmentation spatio-temporelle. Celle-ci étant terminée, un suivi de la segmentation spatio-temporelle est effectué avec comme objectif la construction d'une segmentation $S_{t+1/t+1}$ de l'image suivante I_{t+1} . Nous allons de façon plus précise présenter la méthode de construction initiale de la segmentation spatio-temporelle.

3. INITIALISATION DE LA SEGMENTATION SPATIO-TEMPORELLE

Nous avons vu que la construction d'une région spatio-temporelle (macro-région) s'effectue par agglutination de régions spatiales adjacentes pour lesquelles le mouvement estimé est le même.

Ceci implique donc deux choses :

- la construction d'une segmentation spatiale initiale,
- le fusionnement de régions spatiales en régions spatio-temporelles.

Cette dernière suppose le choix d'un modèle de mouvement et l'estimation des paramètres du modèle considéré.

L'estimation du mouvement instantané lie deux images successives dans la séquence d'images. Du fait de l'objectif d'utilisation de la démarche proposée dans un schéma de codage par compensation de mouvement, celui-ci est de type arrière : le mouvement est estimé de t vers t-1. De ce fait, lors de l'initialisation, la segmentation spatio-temporelle porte sur la seconde image I_1 , la première I_0 étant purement codée en mode intra-image.

3.1. Segmentation spatiale

La segmentation spatiale de l'image I_1 est la base sur laquelle est construite la segmentation spatio-temporelle par fusion de régions spatiales ayant le même mouvement. Du fait de l'interaction entre segmentation spatiale et segmentation spatio-temporelle dans notre approche, la qualité de la segmentation spatiale doit être importante. Parmi le large

spectre des méthodes de segmentation spatiales, nous avons choisi de développer une approche originale faisant coopérer l'extraction des contours présents dans l'image et une extraction de régions par croissance de blocs à partir de germes distribués en fonction des contours détectés [2].

De ce fait, les contours les plus contrastés sont détectés précisément et la fermeture des contours par croissance de régions est régulière. Une approximation des contours résultants par segments de droites permet de construire ainsi des régions R de forme polygonale.

3.2. Fusionnement de régions spatiales en macrorégions spatio-temporelles

Le fusionnement des régions spatiales s'effectue au sens du mouvement selon un critère basé sur l'erreur de compensation de mouvement par un modèle de mouvement paramétré par un jeu \emptyset donné de paramètres. Nous nous sommes limités par souci de simplification à un seul modèle de mouvement linéaire affine simplifié à 4 paramètres pour toutes les régions [3] : $\emptyset = (t_x, t_y, \lambda, \theta)$

où : (t_x, t_y) est le vecteur de translation dans le plan image, λ le facteur de divergence et θ le paramètre de rotation.

Donc pour un point P de coordonnées (x, y) dans l'image appartenant à une région dont le centre C est de coordonnées (x_c, y_c) , on obtient un mouvement résultant donné par :

$$\begin{aligned} d_x &= t_x - \lambda (x - x_c) - \theta (y - y_c) \\ d_y &= t_y - \lambda (y - y_c) + \theta (x - x_c) \end{aligned}$$

L'estimation de paramètres de mouvement est faite en utilisant une méthode de gradient couplée à une méthode de relaxation déterministe de façon à éviter autant se faire que peut les minimums locaux.

Lors de l'initialisation, nous avons réalisé le fusionnement des régions par une procédure comportant trois phases :

- le fusionnement de régions adjacentes bien compensées et qui restent après fusionnement bien compensées,
- le fusionnement des petites régions restantes avec les régions adjacentes avec lesquelles elles sont le mieux compensées,
- le fusionnement général hiérarchisé de régions adjacentes respectant un critère un peu plus relâché de compensation de mouvement.

3.2.a - Fusionnement de régions adjacentes en régions bien compensées

Cette première phase a pour objectifs d'agglutiner les régions initiales homogènes spatialement en régions spatio-temporelles ayant le même mouvement et pour lesquelles la qualité de la compensation de mouvement est très bonne. Le critère de fusionnement d'une région R_k avec d'autres régions qui lui sont adjacentes est basé sur l'erreur quadratique moyenne de compensation de mouvement (classiquement appelée la DFD). Celle-ci s'exprime à l'instant t considéré par :



$$DFD(k) = \frac{1}{n_k} \sum_{p \in R_k} [I(p, t) - I(p + d(\emptyset_k, p), t-1)]^2$$

n_k est le nombre de pixels dans R_k .

Une région R_k bien compensée est telle que ceci est inférieur à un seuil fixé μ . Soit $\{R_j^k\}$ l'ensemble des régions adjacentes à R_k et qui sont bien compensées. Pour chaque union de 2 régions $R_{kj} = R_k \cup R_j^k$, on détermine les paramètres de mouvement \emptyset_{kj} optimaux. Ceci conduit pour chacun d'eux à l'optention de DFD (kj).

$$\text{Si } DFD(ki) = \text{Min}_j DFD(kj)$$

et si $DFD(ki) < SD$

alors on fusionne les régions R_k et R_j^k .

On réitère ce processus jusqu'à ce qu'aucune fusion ne puisse être effectuée par un seuil SD donné. On obtient ainsi des nouvelles régions R_k .

3.2.b - Fusionnement des petites régions

Dans les images complexes, la segmentation initiale en régions homogènes fournit parfois une segmentation trop fine produisant des petites régions. Nous avons considéré que les régions de taille inférieure à μ devaient être fusionnées car ne correspondant pas à des objets animés seuls. Le fusionnement de petites régions est nécessaire pour les deux raisons suivantes:

- la contribution de points-contours dans le nombre total de pixels de région est forte,

- l'erreur d'estimation du mouvement sur les contours spatiaux est toujours importante. C'est pourquoi les paramètres de mouvement obtenus pour une région ayant la taille de l'ordre de quelques pixels ne sont souvent pas valables et le fusionnement de telles régions est justifié. Les petites régions R_k sont fusionnées obligatoirement avec l'une des régions R_j^k adjacentes avec laquelle le critère de fusionnement est le meilleur. Pour cela, on calcule sur chaque région R_j^k une

"fonction de vraisemblance" f_j^k pour décider du fusionnement.

Elle est construite :

- à partir de l'erreur quadratique DFD_k^j de compensation de mouvement de la région R_k avec le vecteur de mouvement \emptyset_j^k de la région R_j^k ,

- et à partir de l'erreur quadratique DFD_j^k de compensation de mouvement de la région R_j^k par soit le vecteur de mouvement \emptyset_j^k de la région R_j^k (si celle-ci est de taille plus grande que R_k), soit le vecteur de mouvement réestimé de $R_k \cup R_j^k$ (si R_j^k est de taille plus faible que R_k). Alors on a :

$$f_j^k = DFD_k^j \left(1 + \alpha \frac{|DFD_k^j - DFD_j^k|}{DFD_j^k} \right)$$

Le choix de la région R_j^k par le fusionnement de R_k s'effectue suivant le minimum des f_j^k .

3.2.c - Fusionnement général hiérarchisé

Alors que les deux premières phases avaient pour but de fusionner par une méthode itérative assez simple les régions adjacentes ayant le même mouvement ou de petites tailles et de réduire ainsi le nombre de régions, la troisième phase correspond au coeur même de la méthode de fusionnement. De façon à obtenir de bonnes performances en qualité de segmentation spatio-temporel, elle est basé sur un fusionnement général hiérarchisé qui incorpore une approche relaxation basée région et non basée site comme c'est le cas généralement dans les approches markoviennes traditionnelles. Le critère d'homogénéité utilisé reste toujours basé sur l'évaluation de la DFD et la procédure de fusionnement revient à un test de rapport de vraisemblance généralisé. Si on appelle, $R_0 = R_1 \cup R_2$, on teste les deux hypothèses suivantes :

$$H_0 = \text{en tout point } p \text{ de } R_0 : \quad \varepsilon_{\emptyset_0}(p) = e_0(p)$$

$$\text{en tout point } p \text{ de } R_1 : \quad \varepsilon_{\emptyset_1}(p) = e_1(p)$$

$$\text{et en tout point } p \text{ de } R_2 : \quad \varepsilon_{\emptyset_2}(p) = e_2(p)$$

où $\varepsilon(p)$ est la différence entre la luminance vraie et celle obtenue par le mouvement estimé \emptyset avec l'image précédente.

$\varepsilon(p)$ sont des variables aléatoires gaussiennes centrées et considérées comme indépendantes.

Si f_0 et f_1 sont les vraisemblances associées à chaque hypothèse, le test suivant est effectué :

$$\Lambda = \log(f_1 / f_0) \begin{matrix} < & H_0 \\ & < \lambda \\ > & H_1 \end{matrix}$$

Pour stabiliser la segmentation vis à vis de l'ordre de prise en compte de régions examinées, la stratégie de fusion consiste à réaliser à chaque itération la fusion la meilleure parmi toutes les fusions possibles : celle correspondant à la valeur minimale du critère. En utilisant une liste ordonnée construite à partir de tous les couples de régions adjacentes et une structure de données adéquate, on obtient des temps de calcul inférieurs aux stratégies de type balayage séquentiel.

4- RESULTATS

Nous avons testé cette procédure d'initialisation de la segmentation spatio-temporelle à la fois sur une séquence d'images synthétiques SEQMILL comportant des objets géométriques texturés animés de mouvements plan divers (translation, rotation, divergence) sur un fond très limité et sur la séquence naturelle CAMPAGNE. Les images de la figure 1



montrent successivement l'image à $t=1$ de la séquence SEQMILL, sa segmentation en régions homogènes spatialement et l'initialisation de la segmentation spatio-temporelle obtenue à $t=1/t=0$. Le tableau I suivant donne le nombre de régions initial et celui obtenu à l'issue de chaque des trois phases de fusionnement au sens du mouvement et l'erreur quadratique moyenne de reconstruction finale.

NBR / Séquence	initial	phase 1	phase 2	phase 3	DFD
SEQMILL	49	30	29	20	30.27
CAMPAGNE	186	161	140	90	67.31

On remarque que cette méthode permet de conserver généralement une forme régulière aux frontières des macrorégions.

5- CONCLUSION ET PERSPECTIVES

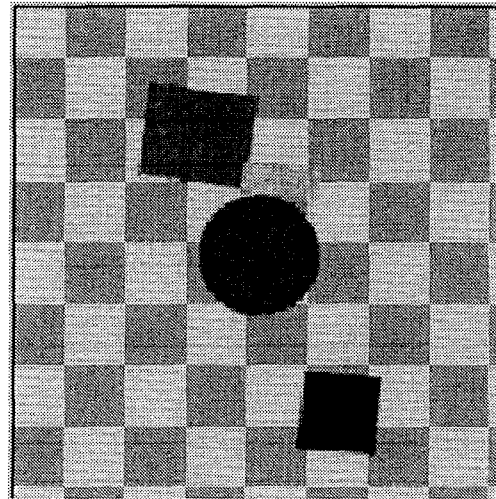
Dans ce papier, nous avons présenté une méthode d'initialisation de la segmentation spatio-temporelle d'une séquence d'images à partir de la segmentation spatiale. Cette méthode a l'avantage que, par construction les frontières des régions ainsi formées sont, lors de l'initialisation des frontières spatiales (contours). Dans le contexte du suivi temporel de cette segmentation spatio-temporelle, nous développons une méthode qui permet par prédiction puis ajustement de garder cette propriété au cours du temps. Le traitement des zones d'occlusion des régions en mouvement et plus particulièrement celui des régions de découverture est effectué quand elles sont larges suivant un principe similaire.

REMERCIEMENTS

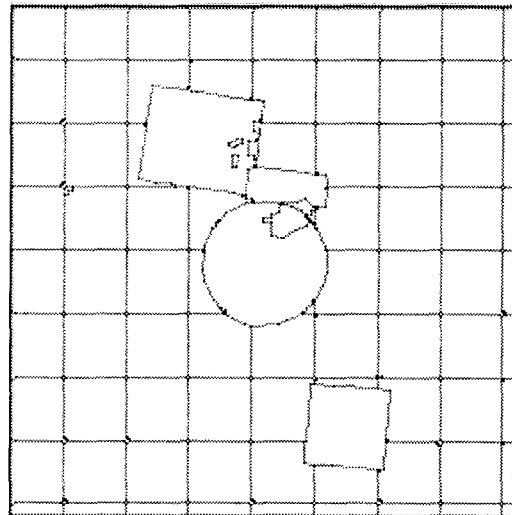
Cette étude s'effectue dans le cadre d'une convention de recherche avec le CCETT et en collaboration avec l'IRISA de Rennes (convention CCETT/IRESTE et CCETT/IRISA 92 ME 03).

BIBLIOGRAPHIE

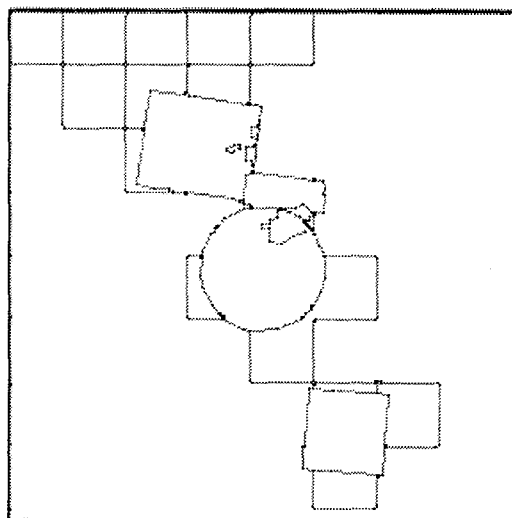
- [1] J.F. BERTRAND, D. BARBA : "Optimization of a monochrome picture coding scheme based on a two-component model", 9th ICPAR, Rome, 154 - 18 Nov 1988
- [2] J. BENOIS, D. BARBA : "Image segmentation by region-contour cooperation for image coding" ICPR 11th, La Haye, 1992, Vol c, pp 331-334
- [3] H. NICOLAS, C. LABIT : "Region-based motion estimation using deterministic relaxation schemes for image sequence coding", ICASSP'92, Vol 3, San-Fransisco, USA, Mars 1992, pp. III- 265-268



a) Image d'origine



b) Segmentation spatiale



c) Segmentation spatio-temporelle

Figure 1