

Représentation hiérarchique des séquences d'images

F. MORIER⁽¹⁾, J. BENOIS-PINEAU⁽¹⁾, D. BARBA⁽¹⁾, H. SANSON⁽²⁾

⁽¹⁾SEI, EP CNRS 63, IRESTE, rue Christian Pauc, BP 60601
44306 Nantes, France

⁽²⁾FRANCE TELECOM, CNET/DIH, CCETT, 4 rue du Clos Courtel, BP 59
35512 Cesson-Sévigne, France

RESUME

Ce papier propose une méthode de représentation hiérarchique des séquences d'images. Le niveau de base de la hiérarchie est une segmentation spatiale. Ensuite les régions qui ont le même mouvement ou la même texture sont progressivement fusionnées. Plus nous montons dans la représentation plus les entités globales deviennent accessibles. Cette représentation est suffisamment flexible pour fournir des segmentations manipulables par les utilisateurs et pour produire également un débit variable.

ABSTRACT

This paper presents a method for hierarchical representation of image sequences. The lower level of hierarchy is a spatial segmentation. Then the regions which have the same motion or the same texture are progressively merged. The more we go up in the representation the more global entities become accessible. This representation is sufficiently flexible to provide segmentations that users can manipulate and to provide a scalable bitrate too.

1 Introduction

La segmentation automatique des séquences d'images animées est un problème complexe. De nombreux travaux ont déjà vu le jour. Nous pouvons par exemple citer la segmentation par champs de mouvement [1], par morphologie mathématique [2] ou par fusion de régions spatio-temporelles [3].

La problématique abordée ici tente de répondre à ce but de segmentation automatique tout en donnant la possibilité de manipuler le contenu de ces images et de générer un débit à transmettre variable.

Une hiérarchie de segmentations des images animées est proposée ; elle permet de représenter leur contenu avec un niveau de détail progressivement relâché. L'utilisateur peut donc puiser dans cette panoplie de segmentations, chacune d'elles représentant un débit de transmission propre.

Ce travail bénéficie du soutien du CCETT au titre du marché 95ME20.

2 Schéma général

La hiérarchie proposée est purement ascendante. Elle démarre par une segmentation spatiale [4], puis les régions sont de plus en plus fusionnées. Le niveau haut de la hiérarchie représente les plus grosses entités discernables vis-à-vis du critère de fusion utilisé.

Le mouvement et la texture sont les deux caractéristiques utilisées séparément ou conjointement dans cette étude. Pour le mouvement le premier triplet de trames de la séquences

d'images animées est utilisé; pour la texture, seule l'image courante est nécessaire.

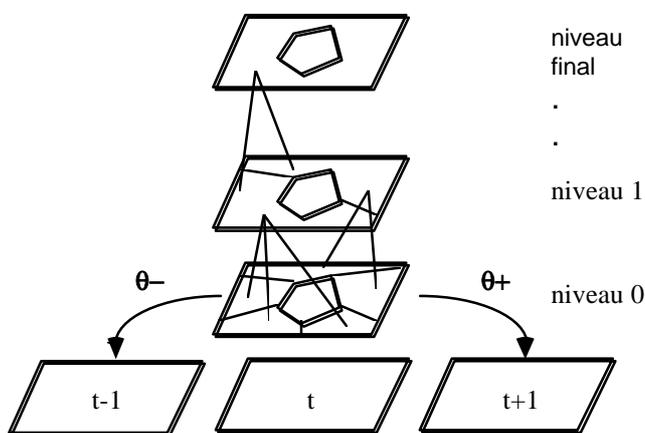


fig. 1: Schéma hiérarchique général

3 Segmentation hiérarchique basée mouvement

3.1 Estimation de mouvement

L'estimateur de mouvement employé utilise un modèle de mouvement linéaire simplifié (translation, divergence, rotation). Une initialisation par mise en correspondance de régions est employée. La méthode de descente du gradient à gain adaptatif ainsi qu'une relaxation déterministe [5] sont enchaînées.

Deux autres techniques d'estimation sont ajoutées également: la multirésolution [6] [7] et la bidirectionalité [8] (estimation en amont et en aval). La première permet de prendre en compte les mouvements de grandes amplitudes et la seconde de gérer les phénomènes d'occultation venant troubler l'estimation unidirectionnelle.

3.2 Critère de fusion de régions basé mouvement

Dans [3], un critère naturel de fusion basé sur l'erreur quadratique moyenne de compensation de mouvement a été proposé. Ce critère ne prend pas en compte la non uniformité de comportement de l'erreur de compensation entre les zones texturées et les zones lisses. Le procédé d'acquisition des images ainsi que celui d'interpolation dans l'estimation sub-pixellaire sont, entre autre, à l'origine de ce phénomène.

Pour pallier à ce défaut, dans le nouveau critère de fusion de régions l'erreur de compensation de mouvement est divisée par la projection du gradient spatial dans la direction du déplacement. Ce critère est appelé "critère spatio-temporel d'homogénéité" et est indépendant de la texture:

$$C_R(\theta) = \frac{1}{N_R} \sum_{j=1}^{N_R} F \left(\frac{DFD^2(p_j)}{\left(\frac{\nabla I_t(p_j) \cdot dp_j}{\|dp_j\|} \right)^2} \right) \quad (1)$$

avec

$$\begin{cases} F\left(\frac{x}{y}\right) = \frac{x}{y} & \text{if } y > s \\ F\left(\frac{x}{y}\right) = \frac{x}{s} & \text{if } y \leq s \end{cases}$$

Dans cette formule, θ est le meilleur vecteur de paramètre de mouvement entre le sens amont et aval de la région R considérée ; dp est le vecteur de déplacement élémentaire au point p . N_R est le nombre de points de la région R. Le seuil s correspond au contraste textuel perceptible (lié au bruit présent dans l'image).

3.3 Méthode hiérarchique de fusion de régions basée mouvement

Cette nouvelle méthode est appelée "fusion par similarité de bonne compensation de mouvement". Elle permet d'analyser séparément les régions des couples de régions adjacentes de sorte à éviter l'agglutination de régions. Pour chacune de ces régions, le critère spatio-temporel d'homogénéité est comparé avant et après fusion sur cette même région. Le coefficient k permet de régler l'évolution autorisée du critère:

$$C_{R_1}(\theta_{12}) \leq C_{R_1}(\theta_1) \times k \quad (2)$$

et

$$C_{R_2}(\theta_{12}) \leq C_{R_2}(\theta_2) \times k$$

La hiérarchie est donc créée en faisant évoluer k . Chaque niveau hiérarchique correspond ainsi à un niveau de compensation souhaité. L'évolution de k se rapproche d'une évolution logarithmique de sorte à permettre une fusion plus importante en terme de nombre de régions au début de la hiérarchie et une fusion plus lente ensuite, la taille des régions étant alors plus importante.

Dans cette méthode de fusion, un ordre de fusion a été déterminé. Un balayage linéaire de la liste des régions a lieu et chaque région est fusionnée avec la meilleure de ses régions voisines si le couple de régions vérifie le critère de fusion hiérarchique. La meilleure fusion possible correspond à une décision mutuelle pour le couple en question. Cette technique permet d'uniformiser les fusions sur toute la carte de segmentation.

3.4 Pré-traitement

Un pré-traitement a été employé dans notre construction hiérarchique de sorte à gagner du temps. Ce pré-traitement est utilisé après la segmentation spatiale et permet donc de constituer le premier niveau hiérarchique.

Pour chaque région faisant partie du couple considéré, le critère de fusion doit vérifier une certaine qualité de compensation:

$$C_R(\theta) \leq \lambda \quad (3)$$

Ce seuil est déterminé de façon visuel pour qu'aucune fausse fusion n'ait lieu.

4 Segmentation hiérarchique basée "texture"

L'introduction de la texture dans un schéma de représentation hiérarchique se justifie de plusieurs façons. Cet attribut complète la modélisation des régions et peut permettre de confirmer, infirmer ou modifier les décisions de fusion de régions prises avec un autre attribut (ici le mouvement). Certaines séquences sont également plus adaptées à une segmentation textuelle qu'à un autre type de segmentation.

Dans cette étude, le critère de fusion basé texture se limite à une comparaison des valeurs moyennes des luminances des couples de régions adjacentes. Ceci constitue un critère très simple mais générique pour des séquences naturelles variées:

$$C_t = \frac{(m_1 - m_2)^2}{m_1^2 + m_2^2} \quad (4)$$

La hiérarchie est réalisée sur les principes déterminés dans la section précédente (hiérarchie basée mouvement). Chaque niveau hiérarchique correspond à un pallier du critère de fusion:

$$C_t < 1 \quad (5)$$

Le seuil l évolue de manière logarithmique et la fusion est effectuée de façon ordonnée.

5 Segmentation hiérarchique basée "texture" et mouvement

Le mouvement et la texture étant introduit dans cette étude, il nous est maintenant possible de combiner ces deux attributs pour ainsi profiter de leurs avantages.

Le schéma général envisagé correspond à une mise en parallèle de ces deux critères (ou plus) et un contrôle de la priorité de tel ou tel critère.

Dans cette étude, un simple enchaînement d'une fusion basée texture et d'une fusion basée mouvement est réalisée. La première étape crée des régions suffisamment représentatives pour que l'estimation de mouvement présente dans la deuxième étape soit fiable.

6 Résultats

Les résultats sont établies ici sur la séquence INTERVIEW (format CIF) et COASTGUARD (format CCIR601, séquence ISO-MPEG4). Une analyse sur un triplet d'images a lieu pour la hiérarchie basée mouvement et sur une image pour la hiérarchie basée "texture". La valeur du seuil des algorithmes est indiquée ainsi qu'entre parenthèses, le nombre de régions correspondant.

Les hiérarchies ainsi présentées permettent une bonne séparation des différentes parties des objets tout en assurant une certaine régularité de ces représentations continûment emboîtées.

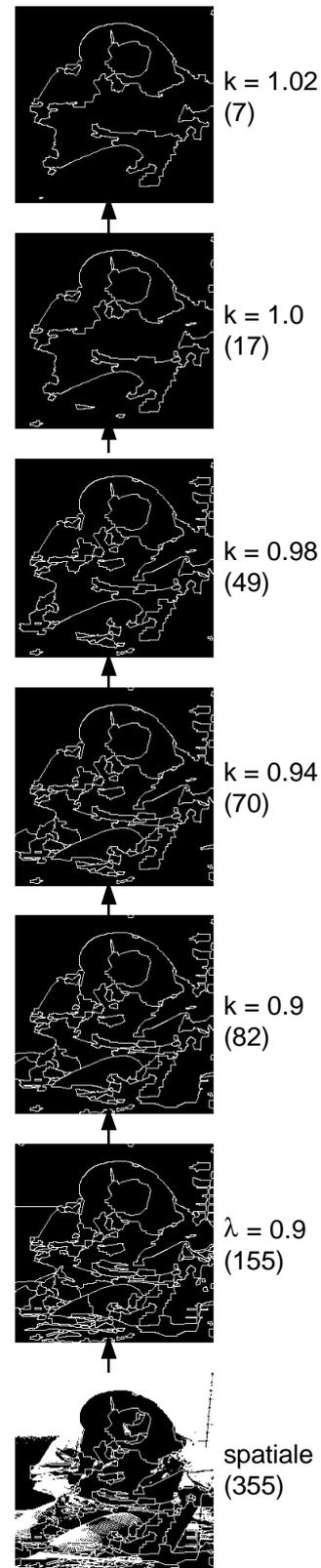


fig. 2: Hiérarchie basée mouvement

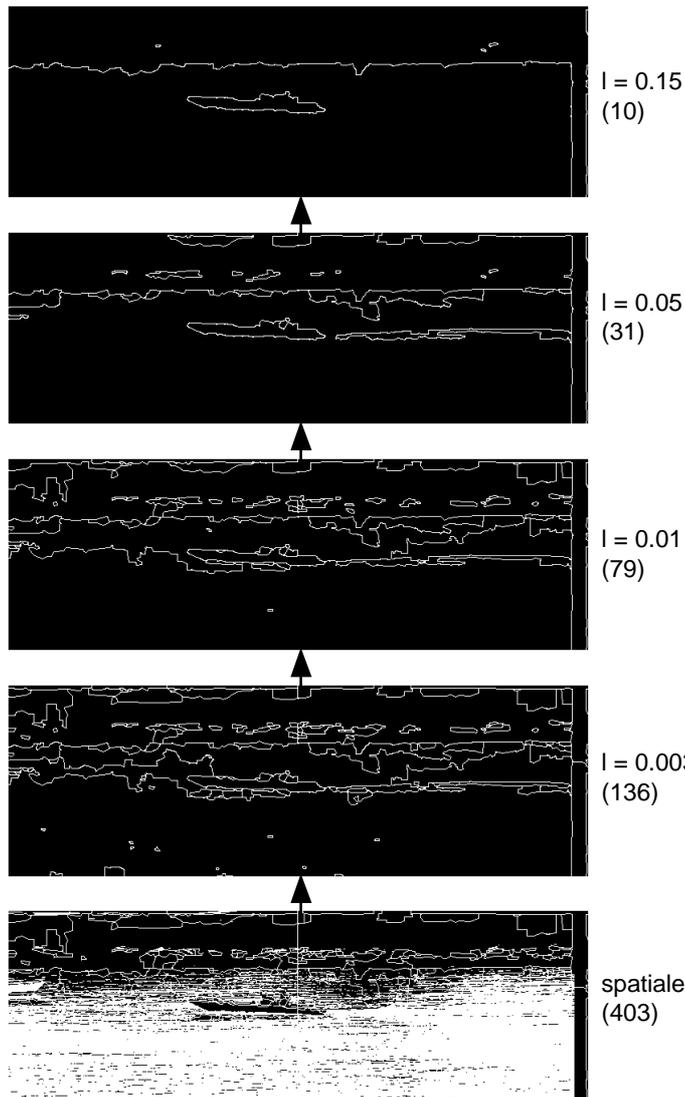


fig. 3: Hiérarchie basée "texture"

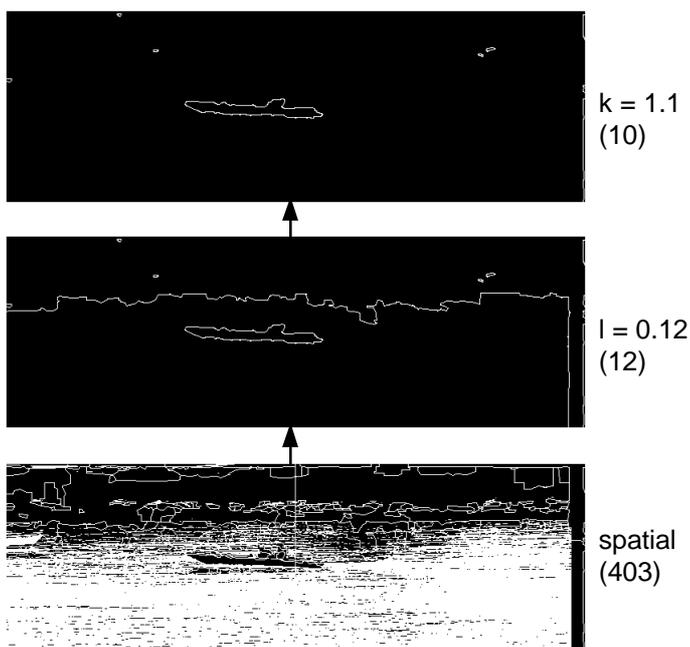


fig. 4: Hiérarchie basée "texture" et mouvement

7 Conclusion

La méthode hiérarchique proposée dans cette étude permet de réaliser des représentations d'une même image de manière suffisamment continue pour que l'utilisateur puisse avoir à sa disposition les segmentations qu'il désire. Les représentations ainsi générées permettront de produire un débit de transmission suffisamment variable pour s'adapter au canal de transmission.

D'autres caractéristiques telles que la profondeur devront être discernées de manière à permettre une meilleure modélisation de la scène représentée ainsi que de faciliter le suivi temporel de ces représentations.

8 Références

- [1] Heitz F., Bouthemy P., "Multimodal motion estimation and segmentation using markov random fields", Actes ICPR, pp. 359-386, 1994.
- [2] Salembier P., "Morphological multiscale segmentation for image coding", Signal Processing, volume 38, pp. 359-386, 1994.
- [3] Wu L., Benois J., Delagnes P., Barba D., "Spatio-temporal segmentation of image sequences for object oriented low bit-rate image coding", Signal Processing, volume 8, pp. 513-543, 1996.
- [4] Benois J., Barba D., "Image segmentation by region-contour cooperation as a basis for efficient coding scheme", Actes VCIP, pp. 1218-1225, 1992.
- [5] Nicolas H., Labit C., "Region-based motion estimation using deterministic relaxation schemes for image sequence coding", Actes ICASSP, pp. 265-268, 1992
- [6] Sanson H., "Estimation de mouvement basée sur des régions pour le codage des séquences télévisuelles", Actes GRETSI, pp. 859-862, 1993.
- [7] Sanson H., "Toward a robust parametric identification of motion on regions of arbitrary shape by non-linear optimization", Actes ICIP, pp. 203-206, 1995.
- [8] Saito T, Komatsu T., "Global motion analysis for mid-level moving image coding", Actes PCS, pp. 481-486, 1996.