

Codage avec pertes de mosaïques vidéo selon un critère MDL

Stéphane Pateux^(1,2) et Claude Labit⁽¹⁾

⁽²⁾IRISA/INRIA

Campus universitaire de Beaulieu

35042 Rennes Cedex

e-mail: spateux@irisa.fr, labit@irisa.fr

⁽¹⁾ *Ingénieur France Telecom mis à disposition de l'INRIA*

RÉSUMÉ

Cet article présente une technique de codage avec pertes de carte de segmentation dans le cadre d'un schéma de compression orientée-régions de séquences vidéo. Cette technique est basée sur une approximation polygonale des contours définissant les différentes régions. L'approximation polygonale effectuée est contrôlée à l'aide d'un critère MDL (Minimum Description Length) permettant de minimiser le coût de description global, prenant en compte le surcoût des erreurs induites dues à l'approximation effectuée. Cette technique permet ainsi d'obtenir des coûts de codage de l'ordre de 0.4 bits par éléments de contours.

ABSTRACT

In this article, we presents a lossy encoding technique for segmentation maps used in region-based video sequence coding scheme. This technique is based upon a polygonal approximation of the contours defining the various regions. The polygonal approximation is controlled by an MDL criterion (Minimum Description Length) able to minimize the global description cost, taking into account the potential increase in the coding cost of the DFD due to the polygonal approximation. This technique leads to an average cost of 0.4 bits per contour elements.

1 Introduction

Dans le cadre de la compression orientée-régions de séquence vidéos, le débit alloué à la définition du partitionnement en régions occupe une part non négligeable du débit total. Dans le cas du codage à très bas débit, il peut même occuper plus de 50% de la bande passante [9]. Afin d'atteindre des taux de compression plus élevés, il est alors intéressant de compresser cette information. Konrad [4] propose d'utiliser une approximation polygonale des contours définissant les régions, Labelle [5] propose quant à lui une approximation des contours à l'aide de splines. Ces deux approches reposent sur le même principe qui est de réaliser une approximation géométrique des contours ; mais cette approximation est basée sur un critère géométrique arbitraire, qui est de minimiser la distance par rapport au contour original, ou bien l'aire située entre le contour original et le contour approximé. Cette technique ne permet pas de contrôler l'erreur induite par ces approximations et peut donc amener à des artefacts de reconstruction, voire à une hausse du débit de codage de l'erreur de prédiction afin de compenser ces artefacts. Ces contraintes limitatives ne permettent d'atteindre que des coûts de l'ordre de 0.8 bits par éléments de contours [3, 4].

La technique que nous proposons, repose sur une approximation des contours selon un critère MDL (Minimum Description Length). Elle consiste, pour une qualité de reconstruction équi-

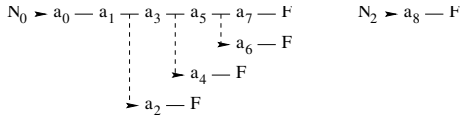
valente, de déterminer l'approximation optimale des divers contours afin d'obtenir le coût de description global (carte de segmentation, paramètres de mouvement et erreur de prédiction) minimum. Cette technique permet d'atteindre des coûts de codage de l'ordre de 0.4 bits par éléments de contours.

Dans un premier temps, la technique utilisée pour coder une carte de segmentation est présentée. Le critère MDL appliqué à l'approximation polygonale de contours est ensuite explicité. Enfin les résultats obtenus sur des séquences caractéristiques sont présentés.

2 Représentation d'une carte de segmentation à l'aide d'un graphe de contour

Afin de représenter la carte de segmentation, la technique couramment utilisée est une technique basée sur la représentation des régions par l'intermédiaire de leurs frontières. Ces frontières sont constituées de contours séparant deux régions d'étiquette distincte. Afin d'éliminer toute redondance dans la définition de ces contours, nous utilisons un graphe de contours représentant la segmentation. La représentation d'une carte de segmentation peut alors être faite via un parcours de ce graphe [7].

Ainsi pour la segmentation de la figure 1, on trouve les parcours suivants :



ce qui peut-être résumé par la notation ([7]) :

$$\begin{aligned} N_0 : a_0 - a_1 - T - a_3 - T - a_5 - T - a_7 - a_6 - a_4 - a_2 \\ N_2 : a_8 \\ Q \end{aligned}$$

où le symbole T représente les noeuds triples utilisés dans la représentation du graphe.

Dans le cas de contours polygonaux, les arcs a_i sont alors définis à l'aide d'une suite de segments codés de façon différentielle.

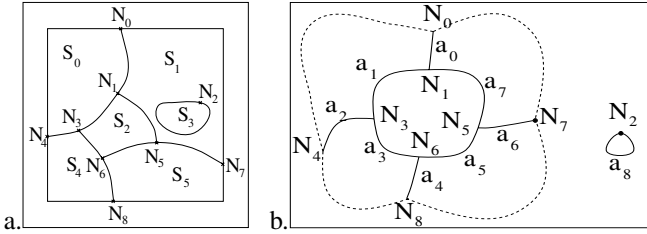


FIG. 1 — Exemple d'une carte de segmentation (a) et de son graphe de contour (b)

3 Approximation polygonale MDL

3.1 Formulation du critère

Afin de réaliser l'approximation polygonale des contours selon un critère MDL, il faut d'abord définir le coût de codage d'une image par une technique orientée-régions. En notant $DL(\chi)$ (Description Length), le coût de codage de l'entité χ , le coût de codage de l'image I^t est [7, 10] :

$$DL(I^t) = DL(S^t) + \sum_{n=1}^N DL(\theta(n)) + DL(\varepsilon^t)$$

où S^t est la segmentation en région utilisée, $\theta(n)$ sont les paramètres de mouvement utilisés sur une région n , et ε^t est l'erreur résiduelle de prédiction à transmettre. Le principe de l'approximation polygonale MDL, est alors de trouver l'approximation polygonale de la carte de segmentation minimisant le coût de codage de l'image I^t . En utilisant, un schéma de représentation de carte de segmentation tel celui présenté précédemment, on a :

$$DL(S^t) = \sum_{segments\ seg} DL(seg) + C_0$$

où C_0 est une constante dépendant uniquement de la topologie de la segmentation (coût de position du point de départ des graphes, coûts des noeuds triple, coût du symbole de fin de codage Q). $DL(seg)$ correspond au coût de codage d'un segment utilisé dans la définition des contours.

3.2 Codage des segments

Afin de coder efficacement les segments utilisés, un codage entropique de ceux-ci est effectué. Comme le nombre de segments acceptables est assez élevé, il est difficile d'avoir une statistique propre à chaque segment. Pour résoudre ce problème, nous utilisons alors une représentation de type longueur et angle afin de définir un segment ; cette modélisation a été retenue car il est assez facile d'obtenir une statistique fiable sur la longueur des segments, et l'orientation du segment est peu corrélée avec l'information de longueur (toutefois l'information de longueur du segment indique le nombre de valeur discrétisées admissibles). Les expériences menées sur les diverses segmentations traitées ont également montré que l'utilisation d'une statistique adaptée pour le codage des valeurs angulaires (prenant en compte une éventuelle corrélation entre deux segments consécutifs) n'amenait pas à des gains de codage significatifs par rapport à la complexité apportée. Étant donné que pour un segment de longueur L donnée, il y a $4L$ angles possibles, le coût de codage d'un segment est alors défini par :

$$DL(seg) = -\log_2(p(\|seg\|)) + 2 + \log_2(\|seg\|)$$

Remarque. En fait les noeuds triples du graphe des contours sont codés comme étant des segments de longueur nulle (il n'est pas alors nécessaire de définir leur orientation), ceci justifie l'hypothèse effectuée sur le coût fixe des noeuds triples.

3.3 Codage de l'erreur résiduelle

Afin de coder, les erreurs résiduelles, nous avons choisi d'utiliser une quantification scalaire de l'erreur associée à un codage entropique de l'erreur quantifiée. Le coût de codage de l'erreur ε^t quantifiée en $\hat{\varepsilon}^t$, est alors :

$$DL(\varepsilon^t) = \sum_{x \in I} -\log_2(P(\hat{\varepsilon}^t(x)))$$

3.4 Simplification du critère

L'approximation des contours ne modifie qu'un nombre limité de points en les faisant basculer d'étiquettes, et laisse inchangés les paramètres de mouvement sur chaque région ; en supposant alors que la loi de probabilité de l'erreur ne change pas, on peut ne prendre en compte que les points changeant d'étiquettes.

La variation de coût de codage d'un point x changeant d'étiquette est définie par :

$$C_{switch}(x) = -\log_2(P(\hat{\varepsilon}_{final}^t(x))) + \log_2(P(\hat{\varepsilon}_{init}^t(x)))$$

Le principe MDL peut alors se simplifier en la minimisation de :

$$\sum_{segments\ seg} DL(seg) + \sum_{x\ basculant} C_{switch}(x)$$

Afin de minimiser ce critère, il est possible de raisonner contour par contour, car le critère est additif et séparable

contour par contour. Dans ce cas, on peut alors faire l'hypothèse que le coût de basculement d'étiquette le long d'un contour varie peu, ce coût pouvant prendre deux valeurs correspondant aux deux basculements possibles le long de ce contour (cf. figure 2, étiquette droite vers étiquette gauche, et étiquette gauche vers étiquette droite). Cette hypothèse peut s'interpréter par le fait qu'un contour sépare deux régions, et que le long de ce contour, la texture de chaque région varie peu, la différence de vecteur mouvement engendre alors des variations de coûts suivant le sens du basculement.

Le critère à minimiser devient alors :

$$\sum_{seg} DL(seg) + \left\{ \begin{array}{c} N_{switch(g \rightarrow d)} C_{switch(g \rightarrow d)} \\ + \\ N_{switch(d \rightarrow g)} C_{switch(d \rightarrow g)} \end{array} \right\}$$

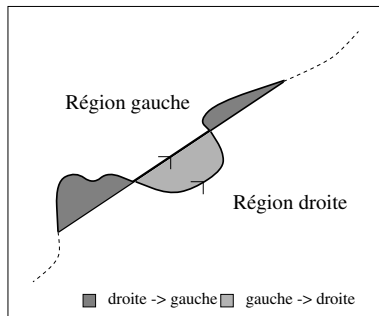


FIG. 2 — Zones de basculement d'étiquette le long d'un contour.

3.5 Minimisation du critère

La minimisation de ce critère est effectuée contour par contour à l'aide d'un algorithme de type division-relaxation (cf. figure 3). L'algorithme est d'abord initialisé à l'aide d'une approximation grossière du contour (un segment reliant les deux extrémités), puis une succession de phases de division-relaxation est effectuée. La division est effectuée sur le segment le plus défavorable au sens de notre critère ; le point pivot choisi pour la division est le point du contour à approximer le plus éloigné du segment approximant. Comme le point pivot choisi n'est pas forcément optimal au sens de notre critère, il est nécessaire de faire une étape de relaxation de la position des différentes extrémités de segments. A chaque étape, une mesure du critère est effectuée, et l'approximation polygonale rencontrée ayant engendrée le critère minimum est retenue comme approximation.

4 Résultats

Nous avons appliqué notre algorithme sur les séquences *Miss America* et *Interview*, préalablement segmentées, respectivement à l'aide d'un algorithme de segmentation de type MDL développé par Nzomigni [6] (mais utilisant une modélisation des contours à l'aide de chaînage de Freeman) et de

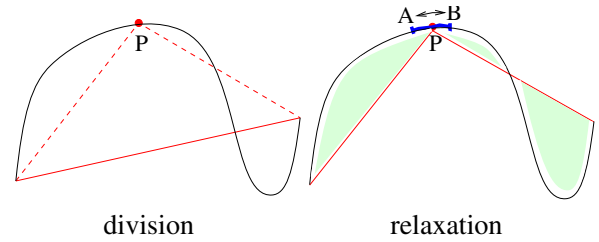


FIG. 3 — Approximation polygonale par division-relaxation

l'algorithme de Garduno [2]. Les résultats obtenus sont résumés dans le tableau suivant et comparés à une technique de codage de la carte de segmentation sans pertes effectuée à l'aide d'un chaînage de Freeman et d'un codeur arithmétique [4, 7]. Le nombre d'éléments de contours pour ces deux séquences est respectivement de l'ordre de 900 et 4000, ce qui fait que le coût moyen de codage d'un élément de contour, avec notre approche, est de l'ordre de 0.4 bits par éléments de contours pour ces deux séquences. Le gain en codage de la carte de segmentation est assez important (plus de 50%), sans pour autant apporter d'augmentation notable sur le coût de codage de l'erreur résiduelle (dans le cas de la séquence *Interview*, le coût de codage de la texture a même été diminué - ceci est dû au fait que la segmentation initiale a été obtenue à l'aide d'un critère cherchant à minimiser les valeurs prises par la DFD, et non le coût de codage de ces valeurs). On peut également observer que l'approximation polygonale des contours, bien que relativement importante pour certains contours, n'a pas amené à une dégradation des images reconstruites.

Sur la figure 4.b, on peut voir le caractère adaptatif de notre approximation. Les contours importants sont bien conservés tandis que les autres contours sont plus ou moins fortement approximatés. Ainsi pour la séquence *Interview*, les contours "externes" de la personne sont bien conservés tandis que tous les contours "internes" sont fortement approximatés (la précision de leur localisation n'étant pas réellement importante).

5 Conclusion

La méthode d'approximation que nous avons proposée permet d'abaisser fortement le coût de codage d'une carte de segmentation dans un schéma de codage orienté-régions. Nous pouvons ainsi atteindre des coûts de l'ordre de 0.4 bits par éléments de contours, sans pour autant dégrader la qualité de reconstruction. Bien que le critère présenté, ne porte que sur un type de codage de l'image erreur bien défini, des résultats similaires ont été obtenus pour des techniques de codage plus évoluées, basés sur une transformation DCT et une quantification vectorielle [8].

Le critère MDL proposé peut également être étendu à d'autres types d'approximations de contours comme par exemple celles à l'aide de splines.

Une perspective d'amélioration de la technique de codage proposée serait d'utiliser un suivi des contours afin d'utiliser les redondances temporelles existantes. A cette fin, il est alors nécessaire d'avoir une segmentation assez stable au cours du

coûts en bits/image	codage de la segmentation	coût de la segmentation	coût du codage des erreurs	coût total	PSNR
Miss America	sans pertes	885	2127	3012	40.7 dB
	avec pertes	355	2189	2544	40.8 dB
Interview	sans pertes	4392	58075	62467	34.2 dB
	avec pertes	1676	57088	58764	34.2 dB

TAB. 1 — Comparaison entre la méthode de codage sans pertes des contours et avec pertes dans le cadre de compression très bas débit.

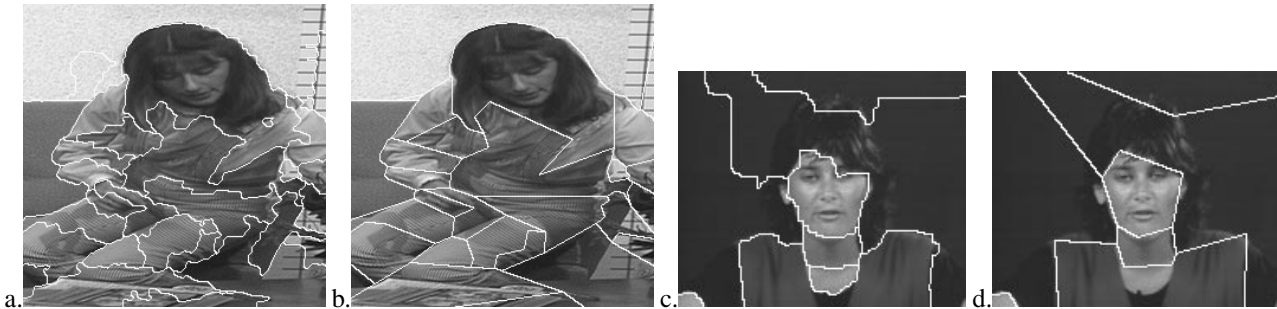


FIG. 4 — Segmentation originales et approximées pour les séquences Interview et Miss America.

temps (pour cela on peut utiliser des techniques telles celles présentées dans [1, 2]), ce qui sort du simple cadre que l'on s'était posé.

Références

- [1] Bonnaud (Laurent) et Labit (Claude). – Suivi d'objets multiples représentés par leurs frontières pour l'interpolation temporelle dans une séquence d'images. *In : Actes du GRETSI 97*. – Grenoble, septembre 1997.
- [2] Garcia-Garduno (Victor). – Une approche de la compression orientée-objets par suivi de segmentation basée mouvement pour le codage de séquences d'images numériques. – Thèse de doctorat de l'Université de Rennes I, mai 1995.
- [3] Gu (Chuang) et Kunt (Murat). – Contour simplification and motion compensated coding. *Signal Processing : Image communication* 7, 1995, pp. 279–296.
- [4] Konrad (Janusz), Mansouri (Abdol-Reza), Dubois (Eric), Dang (Viet-Nam) et Chartier (Jean-Bernard). – On motion modeling and estimation for very low bit rate video coding. *SPIE*, vol. 2501, 1995, pp. 262–273.
- [5] Labelle (Lillian). – Représentation adaptative d'images appliquée à un schéma de codage orientée régions de séquences tv. – Thèse de doctorat de l'Université de Rennes I, 1996.
- [6] Nzomigni (Victor). – Compression sans pertes de séquence d'images biomédicales. – Thèse de doctorat de l'Université de Rennes I, décembre 1995.
- [7] Pateux (Stéphane) et Labit (Claude). – *Codage Efficace de cartes de segmentation pour la compression orientée régions de séquences d'images*. – Rapport technique n 1073, IRISA, janvier 1997.
- [8] Ricordel (Vincent) et Labit (Claude). – Tree-structured lattice vector quantization. *In : Proc. of European Signal Processing Conference EUSIPCO'96. Trieste, Italie*, pp. 731–734.
- [9] Salembier (Philippe), Torres (Luis), Meyer (Fernand) et Gu (Chuang). – Region-based video coding using mathematical morphology. *In : Proceedings of the IEEE*, éd. par IEEE. IEEE, pp. 843–857. – Special Issue on Digital Television.
- [10] Zheng (Heyun) et Blostein (Steven D.). – Motion-based object segmentation and estimation using the MDL principle. *IEEE Transactions on Image Processing*, vol. 4, n 9, septembre 1995, pp. 1223–1235.