

Segmentation d'Image par Minimisation d'un Critère MDL dans une Pyramide de Segmentations

Laurent GUIGUES¹, Hervé LE MEN¹, Jean-Pierre COCQUEREZ²

¹Institut Géographique National / Laboratoire MATIS
2 Avenue Pasteur, 94165 Saint Mandé Cedex, France

²Université de Technologie de Compiègne / Heudiasyc UMR CNRS 6599
BP 20529 Compiègne Cedex, 60205, France

laurent.guigues@ign.fr, herve.le-men@ign.fr, cocquerez@hds.utc.fr

Résumé – Nous proposons une méthodologie originale pour la segmentation en régions des images qui repose sur la complémentarité de deux familles d'approches : les approches hiérarchiques (ou pyramidales) issues de la théorie des graphes et les approches par minimisation de fonctionnelles énergétiques globales. L'idée fondamentale est que les fonctionnelles séparables peuvent être minimisées *exactement* et *en temps linéaire* dans un sous-espace hiérarchisé de l'ensemble des partitions d'un ensemble. Nous justifions alors l'emploi de méthodes de groupement hiérarchique par priorité dynamique pour la construction de pyramides de segmentations. En effet, nous montrons qu'elles conduisent à des objets vérifiant une propriété globale de contraste. Nous proposons alors de minimiser une fonctionnelle de type Minimum Description Length (MDL) dans ces pyramides, Cette fonctionnelle comporte un terme de complexité géométrique de "haut" niveau, qui ne peut être obtenu que grâce à une représentation structurée sous forme de graphe des segmentations.

Abstract – We propose a new methodology for image segmentation. It is based on the complementarity of two families of approaches : graph-based hierarchical (or pyramidal) approaches and global energy minimization approaches. The main idea is that separable energies can be *exactly* minimized in *linear time* over a hierarchized subset of the space of all partitions of a set. We then argue the use of hierarchical clustering techniques based on dynamic priority queues to produce pyramids of segmentations. We show that they lead to clusters that verify a global property of contrast. We then propose to obtain the final segmentation by minimizing a Minimum Description Length (MDL) criterion in these pyramids. This criterion includes a "high-level" geometrical complexity term, that can only be computed thanks to the graph-based representation of the segmentation.

1 Introduction

Nous proposons une méthodologie originale pour la segmentation en régions des images. Elle repose sur la complémentarité de deux familles d'approches relativement disjointes de la segmentation : les approches hiérarchiques issues de la théorie des graphes : groupement (clustering) ascendant [5] ou descendant [8], pyramides irrégulières [3]; et les approches par optimisation globale d'une fonctionnelle de coût sur l'espace des partitions de l'image, par exemple formulée de manière variationnelle [6], ou sous forme d'un critère de type Minimum Description Length (MDL) [4].

Les méthodes hiérarchiques ont été introduites très tôt pour l'analyse des données, en particulier pour les problèmes de clustering et de classification. Dans ce type de problème, on dispose d'une topologie sur l'ensemble des objets élémentaires, sous la forme d'un graphe, et d'une mesure de ressemblance entre objets voisins, sous la forme d'une valuation des arêtes du graphe.

Pour la segmentation d'image, deux graphes duals sont pertinents : Le Graphe des Frontières de Régions et le Graphe d'Adjacence de Régions (RAG). Les régions élémentaires peuvent être tous les pixels de l'image ou données par une sur-segmentation de la scène. Des modèles ré-

cents de représentation des graphes planaires dynamiques, les *Cartes Combinatoires* [1], permettent alors d'effectuer des modifications et des requêtes - topologiques et géométriques - très efficacement dans ces deux graphes.

Les méthodes de clustering s'appuient sur le RAG. Parmi elles, on distingue les méthodes *agrégatives* qui consistent à regrouper itérativement les objets deux à deux, le couple à fusionner étant sélectionné selon une priorité dynamique, initialisée par les ressemblances entre couples. Selon le choix de la méthode de mise à jour des priorités, on aboutit à différents algorithmes, notamment au *Single*, *Average* et *Complete Linkage* (fusion par *min*, *moy* et *max*) [9].

Dans les années 90, d'autres méthodes de construction de pyramides de graphes ont été proposées [3]. Elle s'appuient sur un principe de réduction d'un facteur constant de la taille du graphe entre deux niveaux de la pyramide, (comme dans une pyramide multi-résolution) mais selon des groupements adaptatifs de pixels. L'idée est de parvenir à des décimations rapides (hauteurs de pyramides en $\log(n)$) mais sans imposer de trame de réduction a priori.

On peut également construire une hiérarchie par division (2 ou n -aire) itérée d'un graphe. Les solutions les plus récentes à ce problème de coupe s'appuient sur la théorie spectrale des graphes [8]. Une autre technique fréquemment utilisée est la dissection itérée de l'arbre de poids

minimal recouvrant le graphe [5].

Une pyramide (ou hiérarchie) est une représentation synthétique d'un ensemble de segmentations emboîtées. Cependant, pour en dégager une segmentation unique, le problème est alors de trouver une *Coupe* pertinente dans la hiérarchie. Des critères classiques de coupe dans une hiérarchie sont i) de couper à un niveau fixé par seuillage de l'ultramétrie associée, ii) de chercher la coupe maximale pour un critère d'homogénéité, iii) de définir des critères d'arrêt s'appuyant sur les ressemblances (e.g. variances inter/intra clusters). Les deux premiers critères nous apparaissent peu pertinents car ils reposent sur des définitions absolues et n'autorisent pas les coupes adaptatives. Le dernier type de critère, bien qu'adaptatif, ne s'appuie toujours que sur la matrice de ressemblances initiales. Or ces ressemblances i) ont été calculées entre les régions élémentaires ii) n'incluent pas de critère de formes sur les régions qui sont pourtant primordiales pour parvenir à une bonne interprétation. Il semble alors plus pertinent de revenir à des mesures plus globales, et indépendantes, pour juger de la qualité relative des segmentations contenues dans la pyramide.

La formulation du problème de segmentation comme un problème de minimisation d'une fonctionnelle de coût sur l'espace des partitions d'une image a vu le jour à la fin des années 80 [6][4]. Ce type de formalisation est attrayant car il permet d'exprimer ce qui est peut-être le paradigme central du problème d'inférence de modèle : *L'interprétation de données relève d'un compromis entre la complexité du modèle utilisé et l'adéquation du modèle aux données* (Formulation moderne du principe du "Rasoir d'Occam"). Le balance entre complexité et adéquation d'un modèle a essentiellement été traduite par l'introduction de termes de "régularisation" (e.g. longueur ou courbure minimale)[6], de "pénalisation" (nombre de paramètres minimal : AIC,BIC), ou formulée dans un cadre d'inférence bayésienne ou de longueur de codage minimal des paramètres (MDL,MML) [7] [4]. Les deux derniers formalismes nous semblent les plus pertinents car ils permettent de mettre en balance des coûts de nature hétérogène (e.g. géométrique et radiométrique), et s'affranchissent en théorie d'une pondération subjective.

Le problème majeur de ces méthodes est que la minimisation du critère choisi est difficile : l'espace de recherche est gigantesque et les énergies ont de nombreux minima locaux. De nombreuses méthodes de résolution approchée ont été proposées, e.g. variationnelles [6], par "continuation" [4], ou méthodes généralistes, tel recuit simulé.

2 Pyramides et Fonctionnelles Séparables

L'idée générale que nous proposons est de marier les approches pyramidales et celles d'optimisation globale afin de tirer parti des avantages des deux approches. elle s'appuie sur le résultat classique suivant :

Théorème 1 *Soit H une hiérarchie sur un ensemble X et C une fonction de coût sur l'ensemble $\mathbb{P}(X)$ des partitions*

de X . Si C est séparable, i.e. s'écrit

$$\forall P \in \mathbb{P}(X) \quad C(P) = \sum_{R \in P} D(R) \quad (1)$$

où D est une fonction de coût sur les parties de X , alors la coupe de H qui minimise C se calcule par programmation dynamique en remontant H . La complexité du calcul est $\mathcal{O}(m)$, où m est le nombre de liens père-fils de H , nombre qui est borné par $2(n-1)$, avec $n = |X|$.

Or, la plupart des fonctionnelles proposées dans la littérature sont séparables [6][4]¹. En effet, la formulation du coût de description d'une scène comme une somme de coûts sur les objets qui la composent est relativement naturelle, en particulier si l'on souhaite que les fonctionnelles soient relativement invariantes par restriction à des sous-parties d'image.

On voit donc que si l'on restreint l'espace des partitions à une famille de partitions hiérarchisées, l'optimisation de telles fonctionnelles peut être réalisée *exactement* et *en temps linéaire* par rapport au nombre de pixels de l'image.

Le principe général de notre méthodologie de segmentation est donc d'utiliser une technique de clustering de graphes pour construire une pyramide d'hypothèses de segmentations, de laquelle la segmentation finale est extraite par optimisation d'une fonctionnelle de coût séparable. Le rôle de la pyramide est d'élaguer l'espace de recherche, en proposant un ensemble structuré de segmentations possibles. Le rôle de la fonctionnelle de coût est de sélectionner une solution dans cet ensemble.

En conclusion, si les objets que l'on souhaite segmenter apparaissent à *un niveau quelconque* de la pyramide et que la fonction de coût est bien minimale sur ces objets, alors le résultat sera celui souhaité, et le même que si l'on avait pu conduire une minimisation exacte sur l'ensemble des partitions possibles de l'image.

3 Pyramides de Cocons

Dans [2], nous avons introduit une nouvelle classe de pyramides de graphes, qui repose sur la définition des sous-graphes remarquables d'un graphe, appelés *Cocons*. La définition y était restreinte à une famille particulière de cocons. Nous la généralisons ici.

Soit $G = [X, U, d]$ un graphe non orienté, attribué par une mesure de dissemblance d portée par les arêtes de G . Si A et B sont deux sous-ensembles de noeuds, on définit $U[A, B]$ comme l'ensemble des arêtes qui ont une extrémité dans A et l'autre dans B . On se dote de deux fonctions f et g à valeurs réelles sur les sous-ensembles d'arêtes de G , qui ont pour rôle de calculer des mesures "agrégées" des dissemblances portées par des ensembles d'arêtes. On définit alors le *Contraste Interne* I_f et le *Contraste Externe* E_g d'un ensemble de noeuds $Y \subset X$ par

$$\begin{aligned} I_f(Y) &= f(U[Y, Y]) \\ E_g(Y) &= g(U[Y, X \setminus Y]) \end{aligned}$$

¹. Contrairement à l'intuition, un certain nombre de formulations Markoviennes sont aussi séparables.

$I_f(Y)$ est une mesure synthétique des dissemblance portées par les arêtes internes à Y , i.e. connectant deux noeuds de l'ensemble, et $E_g(Y)$ est une synthèse des dissemblances entre l'ensemble Y et son voisinage $X \setminus Y$. On dit alors qu'un ensemble Y est contrasté (pour f et g) ssi $I_f(Y) \leq E_g(Y)$. Un *Cocon* est un sous-ensemble *Contrasté Connexe* de G . On note $C_f^g(G)$ l'ensemble des Cocons de G pour les fonctions f et g .

Dans [2], nous avons montré que $C_{max}^{min}(G)$ (privé de \emptyset) est une hiérarchie sur X , et que cette hiérarchie est une sous-hiérarchie de celle obtenue par *Complete Linkage*.

Dans un RAG, un cocon de $C_{max}^{min}(G)$ est un ensemble connexe de régions dont toutes les dissemblances internes sont inférieures aux dissemblances avec les régions voisines. Quand les dissemblances s'appuient sur les différences de contenu radiométrique des régions, cette notion traduit bien une idée de contraste. Cependant, les opérateurs *min* et *max* rendent cette notion de cocon peu résistante au bruit, et en pratique, nous avons besoin d'itérer le processus de coupe de hiérarchie, car les objets globaux n'étaient pas capturés en une seule remontée.

Nous considérons ici d'autres ensembles de cocons, qui utilisent des opérateurs plus robustes, comme la moyenne (*moy*) ou la médiane (*med*). Malheureusement, en général, les ensembles de cocons $C_{moy}^{moy}(G)$ et $C_{med}^{med}(G)$ ne sont pas hiérarchisés.

Intéressons-nous néanmoins à la catégorie des cocons tels que $f = g$. Envisageons un algorithme général de clustering ascendant par priorité dynamique, dit algorithme de *f-Linkage*. Pour f donnée, le *f-Linkage* est défini comme l'algorithme qui à toute étape de l'agrégation, associe à deux clusters A et B la priorité de fusion $f(U[A, B])$. Par exemple, l'algorithme de *moy-Linkage* est celui qui fusionne en priorité les deux clusters tels que la moyenne des dissemblances sur les arêtes qui les séparent est minimale. Moyennant quelques hypothèses minimales sur f (que vérifient *moy* et *med* par exemple), on obtient

Théorème 2 *Tous les éléments de la Hiérarchie construite par f-Linkage sur G sont des Cocons de $C_f^f(G)$.*

Ce résultat signifie que même si par *Average* ou "*Median*" *Linkage*, on n'obtient pas tous les ensembles contrastés pour la moyenne ou la médiane des dissemblances, on n'obtient cependant que des ensembles contrastés. En outre, comme l'agrégation se fait par dissemblances croissantes, on obtient naturellement la hiérarchie des cocons les plus homogènes. Mettons l'accent sur le fait que le *f-Linkage* est par essence une méthode de groupement sur un critère local mais qui conduit à des groupes possédant une propriété globale (de contraste). Ceci est à relier au fait que le *Single Linkage* conduit à l'arbre de poids minimal, qui est défini par une propriété globale. En outre, la définition d'une hiérarchie de *f-Linkage* ne dépend d'aucun paramètre.

4 Coupes sur un critère MDL

Nous décrivons maintenant la fonctionnelle que nous proposons de minimiser pour couper les hiérarchies de cocons. Il s'agit d'un critère de type MDL.

Considérons le codage d'une image I selon une partition $S = \{R_i\}$ du domaine de I en régions. La longueur du code obtenu peut s'écrire sous forme générale :

$$L(I\langle S \rangle) = L(S) + \sum_{R \in S} L(I_R|R) \quad (2)$$

où $L(\cdot)$ désigne la longueur de code nécessaire à la transmission d'une information donnée, dite *Longueur de Description (LD)*. $I\langle S \rangle$ désigne la description de I utilisant le modèle S . $L(S)$ est la *LD* de la partition utilisée pour le codage, et $L(I_R|R)$ est la *LD* de la partie d'image comprise dans la région R , connaissant sa géométrie. Si la géométrie des régions est codée par les frontières, alors

$$L(I\langle S \rangle) = \sum_{R \in S} \frac{1}{2} L(\delta R) + L(I_R|R) \quad (3)$$

où δR désigne la frontière de R . Le facteur $1/2$ provient du fait que chaque frontière élémentaire sépare deux régions. Une telle fonctionnelle est donc bien séparable.

Une approche classique pour coder δR est de considérer le code de Freeman des limites de régions. On a alors $L(\delta R) = k|\delta R|$, avec $k = 2$ pour des frontières interpixelaires (4 connexes), et $k = 2$ ou 3 pour des frontières pixelaires 4 ou 8 connexes, et $|\delta R|$ représente la longueur de la frontière. Du point de vue de l'objectif d'interprétation de l'image, ce type de codage comporte plusieurs inconvénients : i) La longueur discrète peut s'écarter significativement de celle euclidienne, et en particulier n'est pas invariante par rotation ; ii) La longueur n'est pas significative de la simplicité des formes des régions : deux lignes de même longueur ont la même *LD*, quelles que soient leurs formes.

Nous proposons donc de coder les frontières de régions en utilisant des approximations polygonales. Si $Pol(\delta R)$ est une telle approximation, sa *LD* est

$$L(\delta R\langle Pol(\delta R) \rangle) = L(Pol(\delta R)) + \sum_{p \in \delta R} L(p|Pol(\delta R)) \quad (4)$$

où l'on décrit d'abord le modèle polygonal utilisé puis les écarts observés par rapport au modèle. En théorie MDL, on cherche alors l'approximation polygonale qui minimise la *LD* donnée par 4. Il faut pour cela spécifier une méthode de codage des erreurs (une loi a priori). La ligne polygonale optimale se trouve alors par programmation dynamique, mais est en $\mathcal{O}(n^2)$ quand n points sont à approximer.

Nous proposons une autre approche qui consiste à rechercher une approximation avec contrainte d'erreur bornée (par exemple une distance à la corde), et à considérer que les erreurs inférieures à ce seuil ne sont pas pertinentes pour l'interprétation, et n'ont donc pas à être codées. Il s'agit typiquement des écarts entre une droite réelle et son approximation discrète. Dans ce schéma, seul la *LD* de l'approximation est alors à transmettre.

Pour obtenir une description invariante par déplacement, nous proposons de coder les polygones par un parcours en coordonnées polaires relatives. Si une polygone Pol dont les points s'appuient sur une grille discrète est constituée de n segments de longueurs l_i , on peut montrer que sa *LD* est alors approximativement

$$L(Pol) = \sum_{i=1}^n 2 + \log^*(l_i) + \log(l_i). \quad (5)$$

où $\log^*(i) \simeq 1 + \log(i) + \log \log(i) + \dots$ est le coût de codage « universel » de Rissanen d'un entier i [7]. Notons qu'au premier ordre, le codage 5 correspond à une LD proportionnelle au nombre de lignes nécessaire à une approximation de la forme. Ce principe de codage s'étend directement à des modèles plus complexes des formes, e.g. constitués de portions de courbes par morceaux.

Pour la description des radiométries des régions, nous ne considérons actuellement que des modèles gaussiens, qui se révèlent généralement suffisants pour notre problématique : la segmentation d'images aériennes. Nous utilisons pour cela une extension à n dimensions (3 en couleur) des longueurs minimales de description données dans [4].

5 Mise en Application

En pratique, notre algorithme part d'une sur-segmentation, obtenue par calcul d'une ligne de partage de eaux sur le module du gradient [5]. Nous construisons alors le RAG correspondant et l'attribuons par une mesure de dissemblance s'appuyant sur un test de Fisher entre distributions de radiométrie des régions. Les dissemblances obtenues sont modulées par l'intensité du gradient sur la frontière. La mesure de dissemblance est donc mixte régions/contours. Pour les images couleurs, nous avons obtenu de meilleurs résultats dans l'espace colorimétrique $L^*u^*v^*$. La hiérarchie des cocons moyens est alors calculée et la fonctionnelle MDL optimisée dans la hiérarchie. Comme les hiérarchies sont construites de manière ascendante, l'optimisation de la fonctionnelle peut être conduite simultanément lors de la construction. Grâce à l'implantation du RAG sous forme de Carte Combinatoire, à chaque nouveau cluster proposé, nous pouvons extraire l'ensemble de ses frontières entourantes. Nous effectuons l'approximation polygonale à la volée pour calculer la LD géométrique de la nouvelle hypothèse de région. Le coût de description radiométrique s'obtient par ailleurs instantanément à partir de la moyenne et de la matrice de variance/covariance des radiométries, qui peuvent être mises-à-jour instantanément lors des agrégations.

Un exemple est présenté figure 1. On peut y voir que d'une part certaines régions relativement hétérogènes sont extraites en une pièce, car les limites internes en sont très sinueuses, et d'autre part, certaines limites entre régions de contenu très peu différent sont préservées, en raison de leur simplicité polygonale. Le calcul complet sur une image 256×256 prend 2 à 3 s. sur une machine à 500 MHz.

6 Conclusion

Pour conclure, mettons l'accent sur le fait qu'il s'agit d'une méthodologie générale de segmentation. En effet, la méthode elle-même est fixe dès lors que l'on se dote d'une mesure de dissemblance et d'un critère de coût global, additif sur les régions. Notre algorithme est donc applicable à toutes sortes d'images : texturées, images de profondeur, etc.

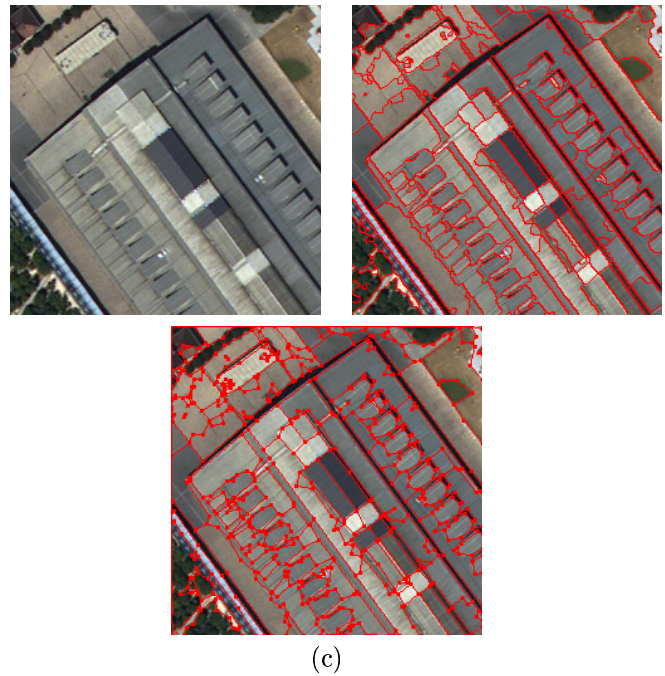


FIG. 1: (a) Image (b) Segmentation (c) Graphe sous-jacent.

Références

- [1] Jean-Pierre Braquelaire and Luc Brun. Image segmentation with topological maps and inter-pixel representation. *Journal of Visual Comm. and Image Repres.*, 9(1):62–79, March 1998.
- [2] L. Guigues, H. Le Men, and J-P. Cocquerez. Graphs, cocoons and image segmentation. In *Proc. of 3rd IAPR TC-15 Workshop on Graph-based Representation in Pattern Recognition*, pages 22–31, May 2001.
- [3] Jean-Michel Jolion and Annick Montanvert. The adaptive pyramid, a framework for 2d image analysis. *Comp. Vision, Graphics and Image Proc.*, 55(3):339–348, May 1992.
- [4] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. Journal of Computer Vision*, 3(1):73–102, 1989.
- [5] Fernand Meyer. Graph based morphological segmentation. In *Proc. 2nd IAPR TC-15 Workshop on Graph based Representations, Haindorf, Austria, 1999*.
- [6] David Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. on Pure and Applied Math.*, 17(4):577–685, 1989.
- [7] J. Rissanen. A universal prior for the integers and estimation by mdl. *Ann. of Statistics*, 11(2):416–431, June 1983.
- [8] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE trans. PAMI*, 22(8):888–905, August 2000.
- [9] Jean-Claude Simon. *La Reconnaissance des Formes par Algorithmes*. Masson, 1985.