

# Segmentation pyramidale et groupements perceptuels

Jérémy HUART, Pascal BERTOLINO

Laboratoire des Images et des Signaux  
BP 46, 38402 Saint Martin d'Hères, France

Jeremy.Huart@lis.inpg.fr  
Pascal.Bertolino@lis.inpg.fr

**Résumé** – Cet article présente une méthode de segmentation qui génère une partition perceptuellement pertinente sans connaissance *a priori* du contenu sémantique de l'image. Tout d'abord, une analyse locale de l'homogénéité de l'image permet de détecter les zones à segmenter. Ensuite, une segmentation locale est effectuée sur ces zones, sur un critère de similarité. Enfin, les régions obtenues sont regroupées en utilisant les critères issus de la théorie de groupement du Gestalt. Cette méthode est présentée en tant que traitement hiérarchique.

**Abstract** – In this paper, we present a segmentation method that provides perceptually relevant partitions without any *a priori* knowledge of the image semantic content: first a local homogeneity analysis detects the image areas that have to be segmented. Then segmentation using a similarity criterion is locally performed. At last, segmented regions are grouped using Gestalt criteria. The whole method is presented in a hierarchical framework.

## 1 Introduction

Nous traitons ici les problèmes relatifs à une méthode générique et automatique d'extraction d'objets dans les images. On distingue trois approches principales dans les méthodes de segmentation. Premièrement, les méthodes orientées sur les contours qui regroupent principalement les algorithmes de contours actifs (*e.g.* snake, balloon [1]). Deuxièmement, les méthodes orientées sur les régions telles que les algorithmes de croissance de régions ou du *split-n-merge*. Cette catégorie regroupe également les approches probabilistes optimisant une fonction d'énergie globale calculée sur les régions [2]. [1] propose une méthode originale qui mélange fonctions d'énergie et morphologie mathématique. Et enfin, les techniques hybrides qui fusionnent les deux types d'information tel que l'algorithme de compétition de régions [3]. [4, 5] proposent de combiner les techniques précédentes avec le *scale-space* afin d'ajouter l'aspect multirésolution qui est bien adapté au système de vision humain.

Afin de proposer une nouvelle méthode non fondée sur la connaissance *a priori* du contenu sémantique de l'image ou sur un modèle quelconque d'objet, plusieurs méthodes efficaces sont intégrées et interviennent successivement dans la pyramide de graphe irrégulière [6] : (1) Une **analyse locale de l'homogénéité** de l'image est effectuée pour initialiser une segmentation locale et ainsi éviter une sur-segmentation. (2) La pyramide de graphe réalise une **segmentation locale** des zones hétérogènes de l'image. En utilisant un critère de similarité, elle génère un empilement de partitions précises par croissance de régions en parallèle. (3) La pyramide est de nouveau utilisée sur les régions issues de la segmentation pour un traitement de **groupement perceptuel** selon des critères issus de la théorie du Gestalt. Ces critères sont bien adaptés à une méthode n'utilisant pas de modèle puisqu'ils prennent en compte uniquement la pertinence visuelle des régions.

## 2 La pyramide locale

La pyramide de graphe [6] est un outil puissant qui fournit une segmentation en multirésolution en un seul traitement. Son principe est d'initialiser un graphe d'adjacence, où chaque sommet correspond à un pixel de l'image. Utilisant un algorithme local effectué sur toute l'image, les voisins similaires peuvent fusionner, entraînant une décroissance du nombre de sommets qui représentent alors des régions. Les régions  $R_i$  et  $R_j$  sont similaires si la distance entre leur couleur moyenne, dans l'espace YUV, est plus faible qu'un certain seuil (cf. équation 1).

$$d(YUV[R_i], YUV[R_j]) < T \quad (1)$$

Ce traitement est effectué de manière itérative sur des graphes successifs jusqu'à convergence du nombre de régions.

Habituellement, le graphe d'adjacence est initialisé avec autant de sommets que de pixels contenus dans l'image pour effectuer une segmentation de l'image entière (fig. 1). Dans une pyramide locale, seul un sous-ensemble de pixels constituant une zone indéfinie de l'image est associé à des sommets. Tandis que le reste des pixels est associé arbitrairement à  $n$  sommets *racines* (régions qui appartiendront au résultat final) telles que le fond par exemple (fig. 2). Pendant le traitement de segmentation, les pixels des zones indéfinies fusionnent ensemble et forment des régions qui se rattachent à une des racines, afin d'affiner leur forme.

## 3 Segmentation orientée similarité

Pour initialiser automatiquement la pyramide locale, nous utilisons une méthode présentée dans [7] qui réalise une analyse de l'homogénéité dans les images couleur. Contrairement à [7] qui calcule l'image d'homogénéité (ou «H-image») selon les composantes RGB, nous utilisons l'espace couleur CIE L\*a\*b

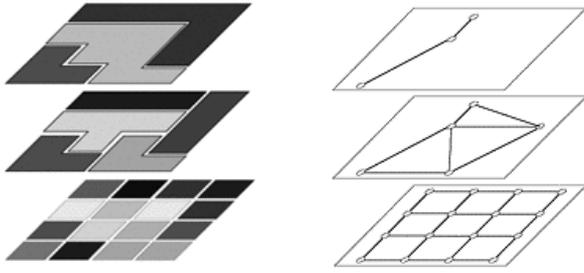


Figure 1: Pyramide de graphe construite sur une image 4×4 pixels : les partitions et leur graphe



Figure 2: Exemple d'initialisation de pyramide locale

puisqu'il permet de fournir plus d'homogénéités en texture (voir fig. 3.b et 4.b). La H-image est une image en niveaux de gris dont les valeurs des pixels représentent les discontinuités de l'image selon un certain critère d'homogénéité. Les fortes valeurs correspondent à d'éventuelles discontinuités tandis que les faibles valeurs correspondent à des régions homogènes. Trois H-images sont donc calculées correspondant aux trois composantes de l'espace couleur :  $H_L$ ,  $H_a$ ,  $H_b$ , et la H-image finale est obtenue en utilisant la formule suivante :

$$H = \sqrt{(H_L^2 + H_a^2 + H_b^2)} \quad (2)$$

Ensuite, une classification des pixels de la H-image produit une partition binaire comprenant des composantes connexes homogènes et hétérogènes. Les auteurs de [7] effectuent une croissance de régions à partir de germes sélectionnés dans les zones homogènes. En ce qui nous concerne, ce masque binaire initialise les racines et les zones indéfinies nécessaires à la segmentation locale (fig. 3.b).

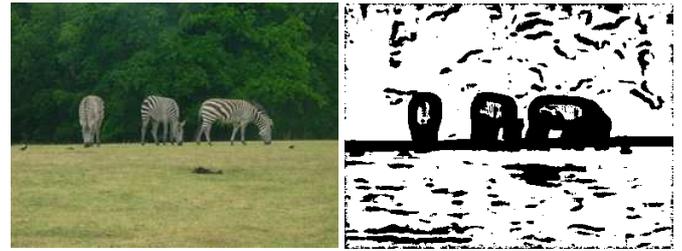
Les composantes connexes en blanc sont les racines. Elles représentent d'éventuels objets d'intérêts. Les composantes noires quant à elles représentent des hétérogénéités qui peuvent être d'éventuels contours d'objets. Les pixels de ces composantes noires sont amenés à fusionner ensemble et/ou avec une racine voisine selon le critère de similarité [8].

En comparaison du résultat généré par une pyramide classique (fig. 3.c), celui obtenu avec une segmentation locale initialisée par un masque d'homogénéité (fig. 3.d et 4.c) est plus approprié, puisqu'il comporte moins de régions mais conserve la même précision. Notons que dans les deux cas (local et global), un seuil identique a été utilisé.

## 4 Groupement orienté perception

### 4.1 Groupement de régions pyramidal

Un traitement de groupement de régions est maintenant utilisé pour simplifier d'avantage la partition issue de la segmentation. Deux contraintes doivent être respectées pour l'efficacité de ce traitement. Premièrement, seuls les meilleurs groupements locaux doivent être retenus. Ce qui signifie qu'un maximum de combinaisons de régions doit être étudié (parmi deux,



(a) Image originale

(b) Masque d'homogénéité RGB



(c) Segmentation par pyramide classique

(d) Segmentation par pyramide locale

Figure 3: Comparaison entre segmentation globale et segmentation locale

trois, quatre, ...,  $n$  voisins). Deuxièmement, le résultat ne doit pas être influencé par l'ordre des groupements.

La pyramide irrégulière a été choisie afin de réaliser l'étape de groupement pour trois raisons principales. Premièrement, sa structure de graphe est bien adaptée à la manipulation en parallèle (i.e. indépendante) de régions. Deuxièmement, les critères de groupement de régions sont facilement interchangeables. Enfin, les itérations du traitement sont simplement obtenues par génération de niveaux supplémentaires résultant des fusions entre régions.

Le graphe final de la pyramide de similarité constitue le graphe initial de la pyramide de groupement. En effet, la pyramide locale est étendue avec des niveaux supplémentaires induits par le groupement de régions.

Dans [2] les auteurs groupent seulement des paires de régions. Contrairement à leur travail, avec notre méthode, un nombre quelconque de régions peut fusionner simultanément en un seul groupement. Cela fournit plus de choix dans la stratégie de groupement et donc, plus d'adaptativité au contenu de l'image.

### 4.2 Critères de groupement

Les critères choisis pour effectuer le groupement sont dérivés de la théorie du Gestalt [9] qui n'utilise aucun modèle d'objet. La vision humaine effectue des groupements indépendants (appelés Gestalt) fondés sur cinq propriétés principales : la proximité, la similarité, la fermeture, la continuité et la symétrie [10].

Des énergies sont extraites de ces propriétés et sont calculées pour des régions ou des groupements de régions. Le but étant de sélectionner les groupements de plus faibles énergies représentant une forte pertinence visuelle. Le coût d'un

groupement est composé de plusieurs fonctions d'énergie proposées par [2].

$E_{fusion}$  est le coût de l'opération de fusion fondé sur la différence des moyennes des composantes Lab, et sur l'étude des jonctions (continuité des contours) des différentes régions du groupement.

$E_{region}$  est le coût de la région résultant d'une fusion. il peut être considéré comme le degré de pertinence du groupement potentiel (plus l'énergie est faible, plus le degré de pertinence est important). Ce coût est fondé sur la compacité, la convexité et l'aire du groupement.

La fonction d'énergie d'une région résultant d'un groupement est donnée par  $E = E_{fusion} + E_{region}$ . Une énergie faible indique un fort intérêt visuel. Au contraire, une forte valeur indique une région ou un groupement indésirable. Le but étant de réaliser le groupement qui assure la plus faible énergie localement.

### 4.3 Sélection du groupement

#### 4.3.1 Sélection du meilleur groupement local

Soit  $v_c$  un sommet,  $c \in \llbracket 1, N \rrbracket$  et  $n_c$  le nombre de ses voisins. Tous les groupements incluant  $v_c$  et les différentes combinaisons de ses voisins sont considérés. Le nombre de combinaisons est donné par la formule suivante :

$$C = \sum_{j=1}^{n_c} C_{n_c}^j \quad (3)$$

$C_{n_c}^j$  étant le nombre de combinaisons de  $j$  voisins parmi  $n_c$ .

$E_{fusion}$  et  $E_{region}$  sont calculées pour chacun de ces groupements.

Soit  $g_c$  le groupement incluant  $v_c$  ayant la plus faible énergie  $E(g_c)$ .  $g_c$  est un groupement potentiel si : (1)  $g_c$  améliore localement l'énergie de la partition, (2)  $E(g_c)$  indique une forte pertinence visuelle. Si ce n'est pas le cas,  $g_c$  n'est pas retenu.

Notons que pour les calculs expérimentaux du coût de la partition, le nombre maximum de voisins par combinaisons est limité à 5 ou 6, ce qui donne  $C \leq 31$  ou  $C \leq 63$ .

#### 4.3.2 Sélection du meilleur groupement global

Un ensemble  $G$  de groupements potentiels est à présent défini. Les groupements effectivement réalisés sont sélectionnés dans  $G$  par ordre croissant des énergies. Lorsqu'un groupement  $g_s$  est sélectionné, tout groupement de  $G$  qui intersecte avec  $g_s$  est exclu. Ainsi, la fusion de chaque groupement sélectionné peut être correctement réalisée. Ces fusions engendrent, dans la pyramide, un niveau supplémentaire correspondant à la nouvelle partition.

Cette sélection assure les meilleurs groupements dans l'image complète. Le traitement de groupement est réitéré jusqu'à ce que le nombre de sommets reste stable. Ces différents niveaux de la pyramide constituent un aspect multirésolution de la méthode.

## 5 Résultats

Les différents résultats obtenus avec notre méthode sont présentés dans les figures 4, 5 et 6. Pour des images complexes

dont la dimension est d'environ  $300 \times 300$  pixels, la segmentation locale génère habituellement une partition de 100 à 200 régions. De cette partition, l'étape de groupement donne une partition de moins de 20 régions. En général, l'étape orientée similarité converge en moins de 100 niveaux et l'étape orientée perception s'étend seulement sur 10-15 étages supplémentaires.

La figure 4.d représente une partition comportant 13 régions qui définit avec une bonne précision les animaux.

Dans la figure 5.d la méthode donne seulement 2 régions dont 1 définit l'animal.

La figure 6 illustre le fait qu'augmenter le nombre de voisins par groupement peut aider à faire de meilleurs choix dans les groupements. Mais cela augmente considérablement le temps de calcul et la partition finale contient approximativement le même nombre de régions.

L'aspect multirésolution de la pyramide constitue un grand avantage dans le cas où dans les derniers niveaux de la pyramide, des objets d'intérêts sont malheureusement perdus, puisque l'utilisateur peut alors facilement parcourir les différents étages afin de les récupérer.

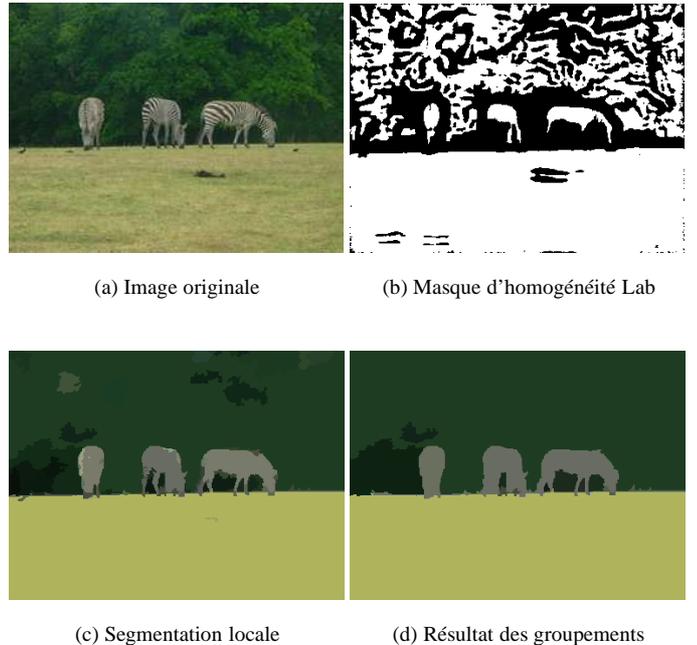


Figure 4: Différents niveaux de segmentation avec plusieurs objets d'intérêt

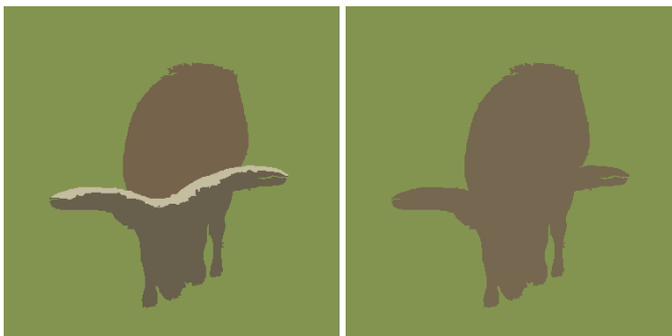
## 6 Conclusion

Nous avons présenté une méthode générique combinant une segmentation locale efficace qui prend en compte les informations couleurs de l'image, avec un traitement de groupement qui fusionne les régions pour engendrer des objets perceptuellement pertinents. Le résultat final est un empilement de partitions composées de quelques objets. Cette méthode est particulièrement appropriée pour décrire la structure d'un objet et/ou pour superviser la segmentation.



(a) Image originale

(b) Résultat de la pyramide locale



(c) Groupement intermédiaire

(d) Groupement final

Figure 5: Différents niveaux de la segmentation d'une image comportant un objet d'intérêt

Nous travaillons actuellement sur la qualification des interactions des différentes énergies pour comprendre plus en détails la pertinence de chacune d'elles dans le voisinage local. Nous voulons aussi étendre ces travaux à la vidéo afin d'inclure les informations temporelles dans ce traitement.

## References

- [1] Hieu Tat Nguyen, Marcel Worring, and Rein Van Den Boomgaard, "Watersnakes: Energy-driven watershed segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, March 2003, vol. 25 (3).
- [2] Jiebo Luo and Cheng en Guo, "Perceptual grouping of segmented regions in color images," in *Pattern Recognition*, April 16 2003, pp. 2781–2792.
- [3] S.C. Zhu and A.L. Yuille, "Region competition: Unifying snakes, region growing and bayes/mdl for multiband image segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, September 1996, vol. 18 (9), pp. 884–900.
- [4] Ana Petrovic, Oscar Divorra Escoda, and Pierre Vanderghyest, "Multiresolution segmentation of natural images: from linear to nonlinear scale-space representations," in *IEEE Transactions on Image Processing*, august 2004, vol. 13 (8).



(a) Image originale

(b) Utilisation de 2 régions maximum par groupement



(c) Utilisation de 4 régions maximum par groupement

(d) Utilisation de 6 régions maximum par groupement

Figure 6: Résultats obtenus avec des groupements locaux comportant un nombre de régions maximum différent

- [5] Ming Tang and Songde Ma, "General scheme of region competition based on scale space," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 2001, vol. 23 (12).
- [6] A. Montanvert, P. Meer, and A. Rosenfeld, "Hierarchical image analysis using irregular tessellations," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, April 1991, vol. 13(4), pp. 307–316.
- [7] F. Jing, M. Li, H. Zhang, and B. Zhang, "Unsupervised image segmentation using local homogeneity analysis," in *Proc. IEEE International Symposium on Circuits and Systems*, 2003.
- [8] J. Huart, G. Foret, and P. Bertolino, "Moving object extraction with a localized pyramid," in *7th International Conference on Pattern Recognition*, Cambridge, UK, august 2004.
- [9] M. Wertheimer, "Principles of perceptual organization," in *Readings in Perception*, 1958, pp. 115–135.
- [10] N. Zlatoff, B. Tellez, and A. Bazkurt, "Image understanding and scene models: a generic framework integrating domain knowledge and gestalt theory," in *ICIP*, Singapore, October 24-27, 2004.