Critères multi-échelles d'évaluation de la segmentation

Sylvie Philipp-Foliguet[1], Laurent Guigues[2]

[1]ETIS, CNRS UMR 8051, ENSEA, 6 avenue du Ponceau, 95014 Cergy-Pontoise Cedex, France

[2]CREATIS, CNRS UMR 5515, INSERM U 630, INSA, 7 rue Jean Capelle, 69621 Villeurbanne Cedex, France

philipp@ensea.fr, laurent.guigues@creatis.insa-lyon.fr

Résumé – Ce papier s'intéresse à la définition de nouveaux critères d'évaluation de la segmentation, dans le cas où l'on ne dispose pas de vérité-terrain. L'avantage de ces critères est de pouvoir comparer des images segmentées en fonction du niveau de détail recherché. Ces critères sont comparés aux principaux critères existants : les critères de Levine-Nazif [5], de Liu-Yang [6] et de Borsotti [2].

Abstract – This paper defines new criteria for segmentation evaluation when no ground truth is available. These new criteria are able to compare segmented images with respect to the level of detail needed by the application. They are compared with the main existing criteria : Levine-Nazif [5], Liu-Yang [6] and Borsotti [2] criterion.

1 Introduction

Devant le foisonnement de méthodes développées depuis plusieurs décennies pour la segmentation des images, le problème de l'évaluation est devenu crucial. Définir une bonne segmentation demeure difficile d'autant que la solution dépend du but recherché. La segmentation n'est pas une fin en soi et l'idée de qualité absolue, indépendante de toute application, est certainement vide de sens. Nous pensons donc que la seule façon réaliste d'aborder l'évaluation est de l'envisager comme une mesure de l'adéquation d'un résultat à un besoin donné. Or une des caractéristiques essentielles d'un besoin en segmentation - si ce n'est la caractéristique essentielle - est le niveau de détail attendu : l'application requiert-elle une description fine ou à grands traits de la scène ?

Nous considérons le cas où l'on ne dispose pas de vérité-terrain et où l'on cherche à évaluer la qualité relative de différentes segmentations. Tous les critères proposés à ce jour pour l'évaluation sans vérité-terrain se présentent sous la forme d'une unique mesure de qualité. Nous proposons pour notre part de *rendre explicite la dépendance de la qualité d'une segmentation à l'échelle requise par l'application*, en développant des critères qui sont des *fonctions* d'un paramètre réel qui se comporte comme un paramètre d'échelle.

2 Critères d'évaluation multi-échelles

D'une façon générale, le problème de segmentation d'image peut être envisagé comme un problème de modélisation par morceaux d'une image : modélisation constante, polynomiale, gaussienne, lisse par morceaux ... chaque région correspondant à un "morceau" du modèle. Une fois une classe de modèle sélectionnée (p.ex. constant par morceaux), la recherche du meilleur modèle peut toujours se formuler comme un problème d'optimisation : rechercher une partition P de l'image en régions et sur chaque région R de P le modèle M_R qui minimisent une certaine énergie totale $E(P)$. L'énergie doit évidemment prendre en compte la qualité de la modélisation en incorporant un terme $E_D(P)$ mesurant la distance entre le modèle et l'image. Toutefois, si l'on se contente d'une énergie d'adéquation du modèle aux données, la segmentation optimale est systématiquement la sur-segmentation absolue ou une partition proche. Par exemple, si l'on considère des modèles constants par morceaux, alors le modèle qui comporte une région par pixel, de valeur la valeur du pixel, est une solution exacte, de distance nulle à l'image. Pour obtenir un résultat utile, il faut alors incorporer un terme énergétique $E_C(P)$, dit de "simplicité", qui pénalise les segmentations trop fines, c'est-à-dire les modèles trop "complexes". Si l'on considère des modèles indépendants sur chaque région, on aboutit alors à des énergies qui prennent la forme générale :

$$E_k(P) = \sum_{R \in P} E_D(R) + k \times E_C(R) \qquad (1)$$

où k est un paramètre réel qui règle la contribution relative des deux termes énergétiques. Notons qu'outre contrôler la finesse de la solution, l'énergie de "simplicité" E_C permet de contrôler la régularité géométrique de la solution, en privilégiant par exemple les régions aux frontières peu tortueuses.

Dans ce cadre, le choix d'une segmentation relève d'un compromis entre adéquation aux données et simplicité du modèle et il n'y a pas intrinsèquement de meilleure solution : certaines applications peuvent avoir besoin d'un modèle précis, qui sera donc complexe, d'autres au contraire peuvent avoir besoin d'une description grossière, à grand traits, de l'image. Si l'énergie E_C est une fonction croissante avec la finesse de la partition, alors le paramètre k de l'équation (1) permet de contrôler la finesse de la

solution, c'est-à-dire se comporte comme un *paramètre d'échelle* : comme nous l'avons vu, si $k = 0$ on trouve un modèle très morcelé qui s'adapte parfaitement à l'image, à l'inverse, pour k assez grand, l'image est modélisée par une seule région. Suivant cette idée, il a été proposé dans [4, 3] de ne pas se contenter d'une seule solution de (1), pour une valeur de k fixée a priori, mais de rechercher une famille de segmentations sous la forme d'une séquence $\{P_k\}_{k\in R^+}$ de partitions de finesse décroissante avec k. On montre que c'est équivalent à rechercher une hiérarchie indicée dont les ensembles représentent les régions appartenant aux minima de (1) et dont l'indice représente la plus petite échelle pour laquelle une région de la hiérarchie appartient à une solution du problème de minimisation. Un algorithme efficace pour trouver une séquence de solutions localement optimales, dans un sens précis, a été proposé [4, 3].

Nous inspirant de ce travail, nous nous intéressons ici à un problème complémentaire du problème de segmentation : celui de l'évaluation de la qualité de résultats de segmentation.

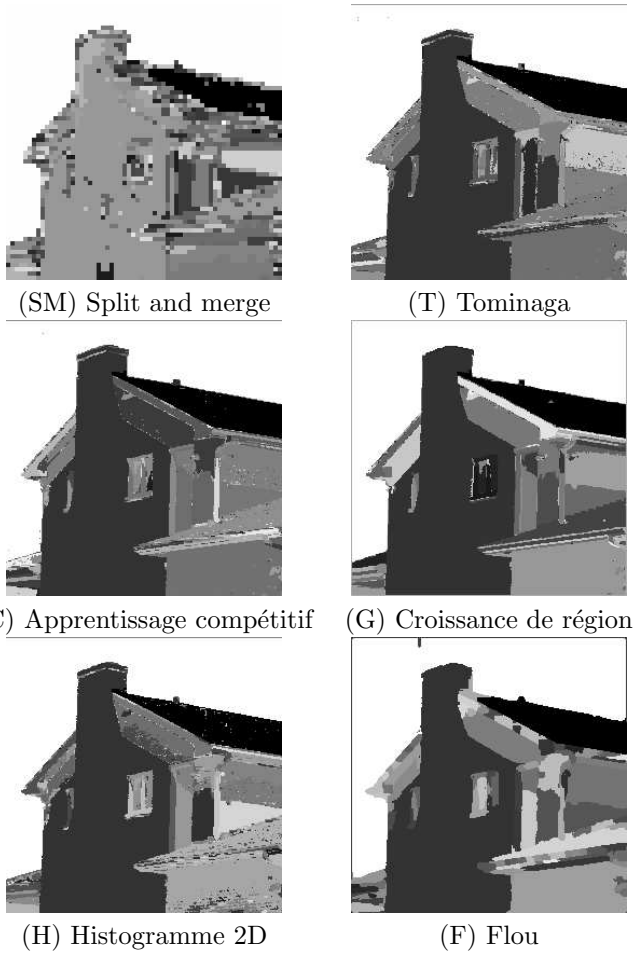
Nous reprenons pour celà l'expression énergétique à deux termes (1), attache aux données $E_D(P)$ et simplicité $E_C(P)$, et proposons de caractériser une segmentation P par la fonction $E_k(P)$ qui à k associe $E_D(P) + kE_C(P)$. Il s'agit d'une fonction affine, croissante si E_C est positive. Si l'on dispose de deux segmentations P_1 et P_2 d'une même image, deux cas se présentent : soit $E_k(P_1) < E_k(P_2)$ pour tout k (ou inversement) auquel cas P_1 est systématiquement meilleure que P_2, soit il existe k_0 tel que $E_{k_0}(P_1) = E_{k_0}(P_2)$, auquel cas une des deux segmentations est meilleure aux petites échelles et l'autre meilleure aux grandes échelles. Cette analyse se généralise au cas où l'on dispose d'un nombre arbitraire n de segmentations. On montre facilement que l'ensemble des échelles k pour lesquelles une segmentation est meilleure que toutes les autres est un intervalle, éventuellement vide, qui représente la gamme d'échelles sur laquelle cette segmentation est la plus pertinente. Notre approche permet donc d'ordonner les segmentations en échelle et d'évincer les segmentations qui ne sont pertinentes pour aucune échelle.

3 Comparaison de différentes énergies

Afin d'évaluer la pertinence de notre approche, nous avons envisagé différentes formes d'énergie.

Soit R_i une région contenant A_i pixels notés $(X_1, X_2, ..., X_{A_i})$ et soit X_p^j la j-ème composante couleur du pixel X_p. Soit μ^j la moyenne de la composante j et V la matrice de variance / covariance des X_p de terme général : $V(j,k) = \frac{1}{A_i} \sum_{p=1}^{A_i} \left(X_p^j - \mu^j\right)\left(X_p^k - \mu^k\right)$. Nous notons λ_j la j-ème valeur propre de V.

Le premier modèle considéré est un modèle constant par morceaux et la distance modèle-image, i.e. l'énergie E_D, est mesurée en norme L_2. Pour une région donnée, le vecteur qui minimise la distance à l'image est bien entendu le vecteur moyen, et la distance est $Q(R_i) = A_i \cdot Trace(V) = A_i \sum \lambda_j$.



(SM) Split and merge (T) Tominaga

(C) Apprentissage compétitif (G) Croissance de région

(H) Histogramme 2D (F) Flou

FIG. 1: *6 résultats de segmentation de l'image Maison*

Le deuxième modèle envisagé est un modèle probabiliste, qui suppose que l'échantillon des couleurs d'une région a été obtenu par tirage i.i.d. selon une loi gaussienne. L'énergie considérée est alors l'opposé du log de la vraisemblance des observations sachant le modèle. Les estimateurs optimaux pour la moyenne et la variance/covariance de la gaussienne sont alors les estimateurs empiriques et l'énergie d'une région s'écrit à une constante près : $G(R_i) = A_i \log(\det V) = A_i \sum_{j=1}^{3} \log(\lambda_j)$.

D'autres formes proches des deux formes précédentes peuvent être utilisées, nous emploierons dans nos tests $D(R_i) = A_i \det V = A_i \prod_{j=1}^{3} \lambda_j$.

La normalisation de l'énergie interne est différente pour les deux expressions : pour des images codées sur $2n$ valeurs par composante, n^2 est un majorant des valeurs de variance et covariance. La normalisation s'effectue donc en divisant par $n^2 \times 3 \times A \times 100$ pour l'énergie Q, par $n^6 \times 3 \times A \times 10000$ pour l'énergie D et par A pour l'énergie G.

En ce qui concerne l'énergie de simplicité, nous prenons simplement la longueur totale de tous les contours, comme dans le modèle de Mumford et Shah [8]. La normalisation

s'effectue par division par le nombre de pixels de l'image.

Nous avons comparé ces critères sur des résultats de segmentation obtenus par différents algorithmes. La figure 1 montre 6 résultats de segmentation obtenus par split and merge (SM), Tominaga (T), apprentissage compétitif (C), croissance de région (G) et classification d'histogramme 2D (H) (cf. [1]) et une méthode floue (F) [9] .

Nous calculons séparément les deux termes d'énergie (attache et simplicité) et nous comparons les différentes expressions de l'énergie d'attache. Le but est de confronter les résultats fournis par les 3 formes de l'énergie d'attache avec notre perception visuelle.

Si l'on représente dans le plan les couples (E_D, E_C) obtenus pour les 6 segmentations (Fig. 2), on constate que Split and merge est nettement discriminé par l'énergie Q, et moins bien par les autres formes d'énergie d'attache. La méthode floue est séparée des autres résultats pour les énergies Q et D, ainsi que par l'énergie contour.

Comme nous l'avons vu, le paramètre k de l'équation (1) est lié à la résolution ou au niveau de détail recherché. Les algorithmes qui donnent peu de régions ou des régions aux frontières régulières sont favorisés par l'énergie de simplicité et défavorisés par l'énergie d'attache aux données.

Dans la figure 2d, nous comparons les résultats en fonction du paramètre d'échelle k : G est toujours meilleur que C, T, SM et H, puisque les courbes E_k pour ces 4 segmentations sont toujours au-dessous de celle pour G. On peut conclure que si l'on recherche une segmentation grossière ($k > 25$) de l'image Maison, F fournit la meilleure segmentation, et que pour une segmentation plus précise, G donne le meilleur résultat.

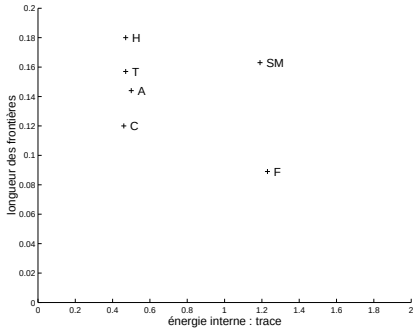
4 Comparaison avec différents critères existants

Nous avons comparé les critères proposés avec 3 critères habituellement utilisés : Levine-Nazif [5], Liu-Yang [6] et Borsotti [2], pour les 6 résultats de segmentation automatique de l'image Maison (cf. table 1) et pour les 5 segmentations manuelles de l'image "Tulipe" issues de la base de Berkeley [7](cf. figure 3 et table 2). Pour tous ces critères, la meilleure segmentation est celle de plus faible score.

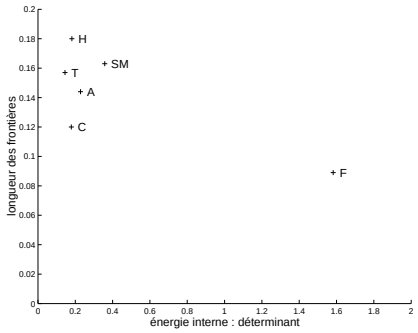
	SM	T	C	G	H	F
nombre de régions	776	1375	1057	654	1384	139
Levine-Nazif	116	78	65	49	70	**31**
Liu-Yang	3.2	0.40	0.37	**0.25**	0.39	0.47
Borsotti	0.4	29	8	1.1	24	**0.1**
Energie $k = 10$	2.82	2.04	1.94	**1.66**	2.28	2.12
Energie $k = 100$	17.5	16.2	14.9	12.5	18.5	**10.2**

Table 1 : Comparaison des critères pour les 6 résultats de segmentation automatique de l'image Maison.

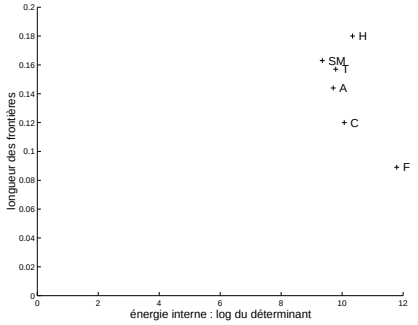
Les résultats sur l'image Maison mettent en évidence que le critère de Borsotti est très sensible aux toutes petites régions (un ou deux pixels). Il classe en premier la segmentation floue de même que Levine et Nazif. Le critère d'énergie classe en premier le résultat G pour un niveau de détail fin ($k = 10$) et le résultat F pour une résolution grossière ($k = 100$), comme montré dans la Fig. 2.



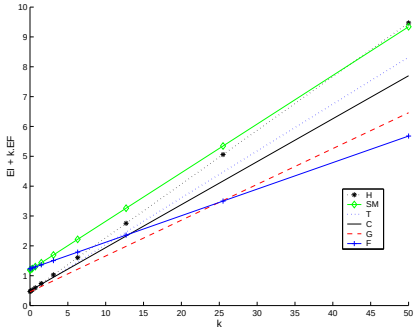
(a) Energie d'attache : Q



(b) Energie d'attache : D



(c) Energie d'attache : G



(d) E_k en fonction de k avec l'énergie d'attache Q

FIG. 2: *Résultats pour les 6 segmentations de l'image Maison.*

Pour l'image Tulipe, les 3 critères Levine-Nazif, Liu-Yang et Borsotti préfèrent la segmentation (a), qui est une caricature en 4 régions de l'image. Ces 3 critères privilégient donc les segmentations grossières. Comme le montre la figure 4, pour notre critère multi-échelles, la segmentation (e) est jugée la meilleure pour les petites échelles; aux échelles moyennes c'est la segmentation (c), et aux grandes échelles c'est la segmentation (a). Les segmentation (b) et (d), elles, ne sont jamais jugées pertinentes. On constate en effet visuellement qu'elles sont

hétérogènes en terme de niveau de détail : globalement grossières mais très détaillées en certains lieux précis.

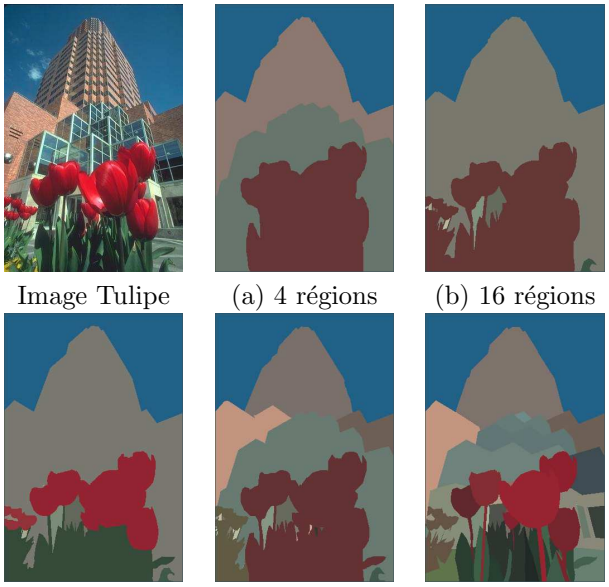


Image Tulipe (a) 4 régions (b) 16 régions

(c) 20 régions (d) 41 régions (e) 79 régions

FIG. 3: *5 segmentations manuelles de l'image "Tulipe" de la base de Berkeley.*

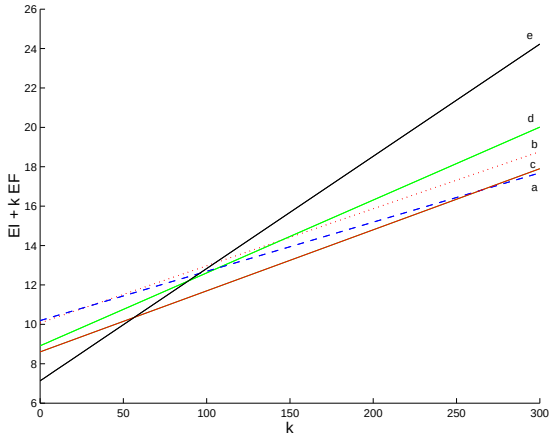


FIG. 4: E_k *en fonction de k (énergie d'attache Q) pour les 5 segmentations manuelles de la figure 3.*

	a	b	c	d	e
nombre de régions	4	16	20	41	79
Levine-Nazif	**2.76**	4.56	6.31	8.78	14.9
Liu-Yang	**0.06**	0.10	0.11	0.20	0.46
Borsotti	**0.1**	0.17	0.17	0.26	0.32
Energie $k = 10$	10.45	10.36	8.91	9.29	**7.70**
Energie $k = 100$	**12.7**	13	11.73	12.65	12.8

Table 2 : Comparaison des critères pour les 5 segmentations manuelles de l'image Tulipe.

5 Conclusion

La plupart des critères pour évaluer des résultats de segmentation prennent en compte d'une part les distances entre les couleurs d'une région et la couleur moyenne de la région et d'autre part une certaine complexité au travers du nombre de régions. Nous pensons que l'évaluation ne peut se faire qu'en fonction d'un but, dont un des éléments est le niveau de détail recherché.

C'est pourquoi nous avons proposé des critères d'évaluation liés au niveau de détail et qui prennent en compte à la fois la complexité de la segmentation et la proximité des régions extraites avec l'image originale. Le premier aspect est mesuré par la longueur des contours, qui représente à la fois le nombre de régions (liée à la résolution) et la régularité des contours. Le second aspect est l'attache aux données, pour laquelle nous avons envisagé trois expressions différentes. Après comparaison, il semble que le plus efficace parmi ces critères soit le critère qui correspond à un modèle constant par morceaux. Pour mesurer la complexité de la segmentation, des critères plus sophistiqués que la simple longueur des contours peuvent être employés.

De plus ces critères sont très simples à calculer, et s'adaptent à tout type d'images, monochromes ou multispectrales, avec ou sans contours entre les régions.

Remerciements : Les auteurs remercient L. Macaire pour les images segmentées de Maison. Si le nombre de régions est différent entre [1] et nos résultats, c'est probablement parce que nous comptons les régions 4-connexes alors qu'elles sont comptées en 8-connexité dans [1].

Références

[1] A. Trémeau, C. Fernandez-Maloigne, and P. Bonton. *Image numérique couleur*. Dunod, Paris, 2004.

[2] M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19:741–747, 1998.

[3] L. Guigues. *Modèles multi-échelles pour la segmentation d'images*. PhD thesis, Université de Cergy-Pontoise, 2003.

[4] Laurent Guigues, Hervé Le Men, and Jean-Pierre Cocquerez. Scale-sets image analysis. In *Proc. of IEEE Int. Conf. on Image Processing (ICIP'03), Barcelona, Spain*, September 2003.

[5] M.D. Levine and A.M. Nazif. Dynamic measurement of computer generated image segmentations. *IEEE Trans. on PAMI*, 7(25):155–164, 1985.

[6] J. Liu and Y.-H. Yang. Multiresolution color image segmentation. *IEEE Trans. on PAMI*, 16(7):689–700, 1994.

[7] D. R. Martin. *An empirical approach to grouping and segmentation*. PhD thesis, University of California, Berkeley, USA, 2002.

[8] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and asociated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[9] S. Philipp-Foliguet, M. B. Vieira, and M. Sanfourche. Fuzzy segmentation of color images and indexing of fuzzy regions. In *First Europ. conf. on Colour in Graphics, Imaging and Vision*, pages 507–512, Poitiers, France, 2002.