

# Etude comparative de critères d'évaluation de la segmentation

Hélène LAURENT, Sébastien CHABRIER, Christophe ROSENBERGER, Bruno EMILE, Pierre MARCHÉ

Laboratoire Vision et Robotique  
ENSI de Bourges - Université d'Orléans  
10 bd Lahitolle, 18020 Bourges Cedex, France

helene.laurent@ensi-bourges.fr, sebastien.chabrier@ensi-bourges.fr  
christophe.rosenberger@ensi-bourges.fr  
Bruno.Emile@univ-orleans.fr, pierre.marche@ensi-bourges.fr

**Résumé** – Nous présentons dans cet article une étude comparative de plusieurs critères d'évaluation de résultats de segmentation d'image dans un contexte non supervisé. Nous avons utilisé une base de cent images synthétiques composées de régions uniformes et texturées. La valeur des sept critères analysés est comparée à la mesure de Vinet (correspondant au taux de classification) sur une base de trois cent résultats de segmentation.

**Abstract** – This paper presents a comparative study of different evaluation criteria for segmentation algorithms with no *a priori* knowledge on the image content. We have worked on a synthetic basis containing one hundred images with uniform and textured regions. The value of the seven tested criteria is compared to the Vinet distance (that corresponds to the classification rate) with three hundred segmentation results.

## 1 Introduction

La segmentation d'une image est une étape essentielle en traitement d'images dans la mesure où elle conditionne son interprétation. Il existe une multitude de méthodes de segmentation dont l'efficacité reste difficile à évaluer. Actuellement, quelques travaux ont été réalisés afin de résoudre le problème plus général de l'évaluation d'un résultat de segmentation [1], [2]. On peut distinguer deux types d'approches.

La première évalue la qualité d'un résultat de segmentation en mesurant sa similarité avec une vérité-terrain. Beaucoup d'applications ont été réalisées notamment dans le milieu médical en utilisant des fantômes [3]. La mesure de Baddeley [4], la distance de Vinet [5] sont des exemples de mesures pouvant être utilisées pour calculer la similarité d'un résultat avec une vérité terrain. Même si cette approche permet de quantifier la qualité de la segmentation en prenant en compte l'interprétation d'une image, elle pose le problème de la confiance dans la vérité terrain qui est souvent subjective. Elle nécessite en outre d'avoir à disposition un expert du domaine applicatif.

Le second se base sur des critères d'évaluation non supervisés permettant d'estimer la qualité d'un résultat de segmentation à partir de statistiques calculées sur chaque région détectée. Dans [6], il est possible de trouver une liste exhaustive de tels critères. L'aptitude de ces méthodes reste assez méconnue surtout dans le cas où l'image à segmenter contient des zones texturées [7], comme dans la plupart des images naturelles.

L'objectif de cet article est de présenter une étude comparative des performances de différents critères d'évaluation dédiés à l'analyse de la segmentation par extraction de régions et ne nécessitant aucune connaissance *a priori* sur les images à segmenter.

## 2 Contexte de l'étude

Nous allons dans un premier temps décrire les critères d'évaluation utilisés dans cette étude ainsi que la base d'images à partir de laquelle nous avons travaillé.

### 2.1 Critères d'évaluation

Les critères présentés dans cet article font, pour la plupart, partie des indicateurs traditionnellement utilisés pour l'évaluation de la segmentation dans un contexte non supervisé. La valeur de chaque critère augmente avec la qualité du résultat de segmentation. Ces valeurs ont été normalisées afin de faciliter leurs comparaisons. Une valeur du critère proche de 1 traduit un très bon résultat de segmentation. Les différents critères sont décrits de façon détaillée dans [2] :

- *Contraste inter-région de Levine et Nazif (Inter-region)*: Ce critère calcule la somme des contrastes des régions ( $R_i$ ) pondérée par leur aire ( $A_i$ ). Le contraste d'une région est défini à partir des contrastes existants avec les régions qui lui sont contiguës.

$$Inter - region = \frac{\sum_{R_i} A_i c_i}{\sum_{R_i} A_i} \quad (1)$$

où

$$c_i = \sum_{R_j} \frac{l_{ij}}{l_i} \frac{|m_i - m_j|}{m_i + m_j} \quad (2)$$

avec,  $m_i$ , moyenne dans la région  $R_i$  ;  $l_{ij}$ , longueur de la frontière commune entre  $R_i$  et  $R_j$  ;  $l_i$ , périmètre de la région  $R_i$ .

- **Critère d'uniformité intra-région de Levine et Nazif (Intra-region) :**

Ce critère calcule la somme des variances normalisées des régions (retranchée à 1).

- **Combinaison disparité intraclasse et interclasse (Intra-inter) :**

L'indice utilisé est une version simplifiée de la méthode proposée dans [9] combinant des versions similaires du contraste inter-région et intra-région de Levine et Nazif.

- **Contraste de Zeboudj (Zeboudj) :**

Cet indice prend en compte le contraste intérieur et le contraste extérieur aux régions mesurés sur un voisinage de chaque pixel. Si on note  $W(s)$  un voisinage du pixel  $s$ ,  $f(s)$  l'intensité de ce pixel et  $L$  le maximum des intensités, les contrastes intérieur et extérieur d'une région  $R_i$  d'aire  $A_i$  et de frontière  $F_i$  se définissent respectivement par :

$$I_i = \frac{1}{A_i} \sum_{s \in R_i} \max\{c(s, t), t \in W(s) \cap R_i\} \quad (3)$$

$$E_i = \frac{1}{l_i} \sum_{s \in F_i} \max\{c(s, t), t \in W(s), t \notin R_i\} \quad (4)$$

où  $l_i$  est la longueur de  $F_i$ .

Le contraste de  $R_i$  est :

$$C(R_i) = \begin{cases} 1 - \frac{I_i}{E_i} & \text{si } 0 < I_i < E_i \\ E_i & \text{si } I_i = 0 \\ 0 & \text{sinon} \end{cases} \quad (5)$$

Le contraste global est enfin :

$$Zeboudj = \frac{1}{A} \sum_i A_i C(R_i) \quad (6)$$

- **Critère de Borsotti (Borsotti) :**

Cet indice est basé sur le nombre, l'aire et la variance des régions.

$$Borsotti = \frac{\sqrt{N}}{1000 \cdot A} \sum_{i=1}^N \left( \frac{\sum_{R_i} (f(s) - m_i)}{1 + \log A_i} + \frac{R(A_i)^2}{A_i^2} \right) \quad (7)$$

avec,  $m_i$ , moyenne dans la région  $R_i$  ;  $R(A_i)$ , nombre de régions ayant une aire égale à  $A_i$ .

- **Critère de dispersion des attributs (Attributs) :**

On calcule sur chaque pixel d'une région, les moments d'ordre 1 à 4 sur un voisinage 3x3 et l'on évalue la mesure de dispersion des attributs à l'intérieur d'une région. Le critère total correspond à la moyenne de ces mesures de dispersion pour chaque région de l'image segmentée [9] retranchée à 1.

- **Evaluation texturée (Texture) :**

Cette méthode développée dans [9] permet d'évaluer un résultat de segmentation sans connaissance *a priori*.

L'originalité de cette méthode réside dans le calcul du critère de qualité en fonction du type de région (uniforme ou texturée). Dans le cas texturé, la dispersion de paramètres de textures est utilisée, et dans le cas uniforme, des paramètres basés sur les niveaux gris sont exploités.

- **Mesure de Vinet (Vinet) :**

Il s'agit d'une évaluation faisant appel à une segmentation de référence et qui repose sur la recherche des couples de régions ayant un recouvrement maximal. Dans la mesure où nous travaillons ici sur une base d'images synthétiques et que nous disposons par conséquent de la segmentation idéale, la mesure de Vinet servira de point de comparaison pour notre étude.

## 2.2 Images tests

La banque d'images sur laquelle nous avons travaillé est constituée d'images synthétiques texturées. Elle comprend cent images composées de 5 régions. Chaque image contient deux régions texturées de l'album de Brodatz [8], deux régions fortement bruitées (avec les mêmes statistiques d'ordre 1 et 2 que les textures) et une région faiblement bruitée (voir Fig. 1).

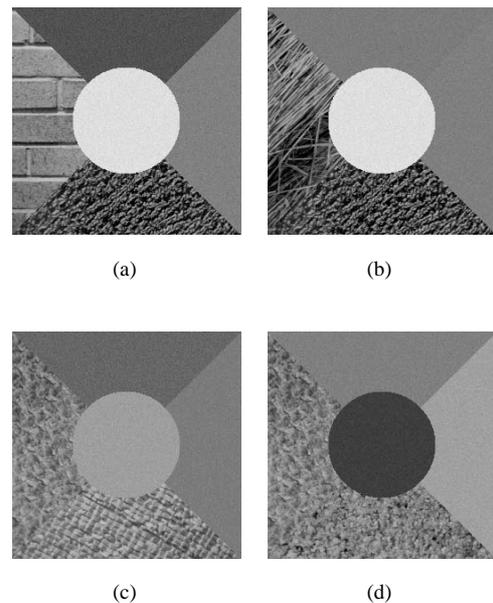


Figure 1: Exemples d'images tests.

Chaque image est segmentée par une méthode de classification (nuées dynamiques floues) avec un nombre de classes égal à 5 et avec les trois types de paramétrage suivants (voir Fig. 2(a),(b),(c)):

- segmentation adaptée aux images uniformes : une taille de fenêtre d'analyse  $5 \times 5$  pixels et les moments de niveaux de gris d'ordre 1 à 4,
- segmentation adaptée aux images faiblement texturées : une taille de fenêtre d'analyse  $9 \times 9$  pixels, les moments de niveaux de gris d'ordre 1 à 4 et les attributs de la matrice de cooccurrence,

- segmentation adaptée aux images fortement texturées : une taille de fenêtre d'analyse  $15 \times 15$  pixels, les moments de niveaux de gris d'ordre 1 à 4, les attributs de la matrice de cooccurrence et l'autocorrélation normalisée.

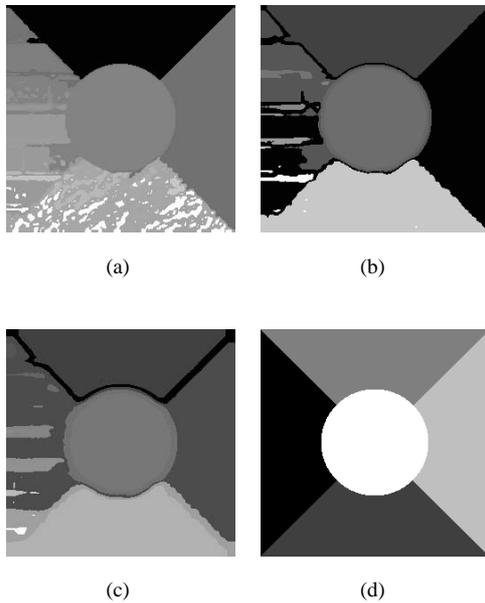


Figure 2: Résultats de segmentation (a),(b),(c) de l'image de la figure Fig. 1 (a) et vérité terrain (d) ayant servi pour la constitution des images synthétiques.

### 3 Résultats expérimentaux

Nous avons étudié dans un premier temps la corrélation des critères comme indicateur de leurs ressemblances (voir TAB. 1) calculée à partir de 300 résultats de segmentation. La dernière colonne montre la corrélation des différents critères avec la mesure de Vinet permettant ainsi de quantifier la ressemblance de l'évolution de chaque critère avec la référence.

Borsotti	1	0.31	0.16	-0.05	0.05	0.5	0.11
Zeboudj		1	0.60	0	0.42	0.11	-0.02
Inter-region			1	0	0.34	0.14	-0.03
Intra-region				1	0	-0.06	-0.23
Intra-inter					1	0.16	-0.08
Texture						1	0.07
Vinet							1

Table 1: Matrice des facteurs de corrélation des critères d'évaluation.

Le taux de corrélation maximal avec de la mesure de Vinet est atteint par le critère *Intra - region* mais reste faible. Il est difficile en ne considérant que cette information de déterminer un critère ayant un comportement analogue à la mesure de Vinet. Aussi, de façon à mieux comprendre l'évolution des différents critères, nous les avons testés sur trois types d'images (allant d'une image uniforme à une image texturée).

A partir des images originales présentées dans les figures 3 (a),(c),(e), les valeurs des sept critères pour les résultats de segmentation de la figure Fig. 3 (g) (résultat très mauvais), des

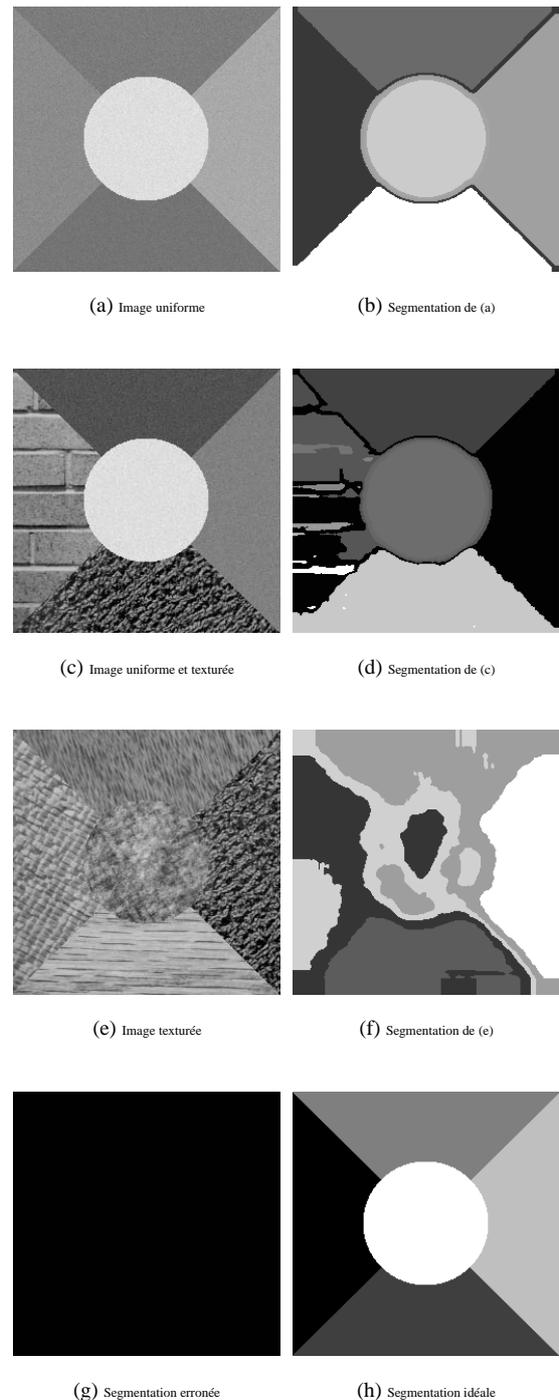


Figure 3: Résultats de segmentation de trois images différentes; un mauvais résultat de segmentation et un résultat de segmentation idéal.

figures Fig. 3(b), (d), (f) (résultats visuellement passables) et de la figure Fig. 3(h) (segmentation idéale) ont été calculées. Les résultats donnés dans les tableaux TAB. 2, 3, 4 devraient mettre en évidence une évolution croissante des critères.

Le critère *Attributs* donne des résultats incohérents particulièrement lorsque l'image originale est non-texturée. On peut remarquer que les critères *Borsotti* et *Intra* ont un comportement satisfaisant lorsque l'image n'est pas trop texturée. Cependant, leurs amplitudes de variation sont faibles

par rapport à la mesure de Vinet.

Le critère *Intra – Inter* respecte l'évolution attendue dans les trois cas de figure mais ne différencie pas bien deux résultats de segmentation proches. Le critère *Texture* donne quant à lui des résultats cohérents mais s'avère plus discriminant en présence de textures dans l'image originale.

On constate que seuls deux critères (*Zeboudj* et *Inter*) ont une évolution conforme aux résultats attendus quelque soit le type d'image considéré. Ces deux critères ayant un facteur de corrélation relativement élevé (0.6), il est normal qu'ils aient des variations similaires.

Critère	Résultat 3(g)	Résultat 3(b)	Résultat 3(h)
Attributs	0.9944	0.9956	0
Borsotti	0.988	0.9995	0.9995
Zeboudj	0	0.044	0.682
Inter	0	0.073	0.168
Intra	0.916	0.958	0.993
Intra/inter	0.457	0.55	0.549
Texture	0	0.647	0.641
Vinet	0.207	0.746	1

Table 2: Comparaison des critères d'évaluation pour trois résultats de segmentation d'une image (Fig. 3 (a)) contenant uniquement des régions uniformes.

Critère	Résultat 3(g)	Résultat 3(d)	Résultat 3(h)
Attributs	0.9978	0.9984	0.9961
Borsotti	0.974	0.987	0.987
Zeboudj	0	0.110	0.531
Inter	0	0.161	0.299
Intra	0.804	0.808	0.816
Intra/inter	0.401	0.553	0.549
Texture	0.2583	0.6767	0.8592
Vinet	0.207	0.74	1

Table 3: Comparaison des critères d'évaluation pour trois résultats de segmentation d'une image (Fig. 3(c)) contenant des régions uniformes et texturées.

Critère	Résultat 3(g)	Résultat 3(f)	Résultat 3(h)
Attributs	0.9953	0.9972	0.9965
Borsotti	0.988	0.966	0.983
Zeboudj	0	0.0001	0.108
Inter	0	0.086	0.142
Intra	0.096	0.510	0.25
Intra/inter	0.451	0.503	0.510
Texture	0	0.733	0.790
Vinet	0.207	0.587	1

Table 4: Comparaison des critères d'évaluation pour trois résultats de segmentation d'une image (Fig. 3(e)) contenant uniquement des régions texturées.

## 4 Conclusion et Perspectives

Cette étude a permis de mettre en évidence les performances de différents critères d'évaluation en fonction du type d'image

considéré. Les critères les plus pertinents parmi ceux testés sont *Zeboudj* et *Inter* puisqu'ils fonctionnent aussi bien lors de l'analyse d'images uniformes ou texturées. On peut noter que nos conclusions diffèrent de l'étude menée par Zhang [6], dans lequel *Zeboudj* s'était avéré inefficace dans un contexte texturé. Cela est probablement dû au fait que nous avons pondéré ce critère par les aires des régions et que nous avons travaillé sur des images multi-régions (ce qui correspond plus au cas des images réelles). Ces critères peuvent être aussi exploités en segmentation ou pour la fusion de différents résultats de segmentation (images multi-composantes).

## Remerciements

*Cette étude a été menée grâce au soutien financier du Conseil Régional du Centre et de l'Union Européenne (FSE).*

## References

- [1] V. Letournel. *Contribution à l'évaluation d'algorithmes de traitement d'images*. Thèse de Doctorat de l'ENST, 2002.
- [2] S. Philipp-Foliguet. *Evaluation de la segmentation*. Rapport Technique, 2001.
- [3] J. Brunette, R. Mongrain et G. Cloutier. *A novel realistic three-layer phantom for intravascular ultrasound imaging*. International Journal Cardiac Imaging, 17(5), pp 371-381 (2001).
- [4] A.J. Baddeley. *An error metric for binary images* Robust Computer Vision, pp. 59-78, 1992.
- [5] L. Vinet. *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques* Thèse de Doctorat de l'université de Paris IX Dauphine, 1991.
- [6] Y.J. Zhang. *A survey on evaluation methods for image segmentation* Computer Vision and Pattern Recognition, (29), pp. 1335-1346, 1996.
- [7] K.A. Bartels et J.L. Fisher. *Multifrequency eddy current image processing techniques for non-destructive evaluation* International Conference on Image Processing, pp. 486-489, 1995.
- [8] P. Brodatz. *Textures, a photographic album for artists and designers* Dover, New York, (1966).
- [9] C. Rosenberger. *Mise en oeuvre d'un système adaptatif de segmentation d'images* Thèse de Doctorat de l'université de Rennes I, 1999.