

# Fusion de critères pour l'évaluation de résultats de segmentation d'images

Sébastien CHABRIER<sup>1</sup>, Hélène LAURENT<sup>1</sup>, Christophe ROSENBERGER<sup>1</sup>, Alain RAKOTOMAMONJY<sup>2</sup>

<sup>1</sup>Laboratoire Vision et Robotique - UPRES EA 2078  
ENSI de Bourges - Université d'Orléans  
10, Bd Lahitolle, 18020 Bourges Cedex, France  
Tél : 02-48-48-40-64, Fax : 02-48-48-40-40

<sup>2</sup>Perception, Systemes, Information - FRE CNRS 2645  
INSA de Rouen, Avenue de l'Université  
76801 Saint Etienne du Rouvray, France  
Tél : 02-32-95-97-03 Fax: 02-32-95-97-08  
sebastien.chabrier@ensi-bourges.fr

**Résumé** – L'évaluation de résultats de segmentation est une tâche très difficile, et ce même pour un expert. L'objet de cet article est de proposer une mesure de qualité d'un résultat de segmentation, sans aucune connaissance *a priori*, en prenant en compte différents critères d'évaluation de la littérature. Nous proposons dans ce papier de fusionner les meilleurs d'entre eux par deux méthodes : une combinaison linéaire optimisée par un algorithme génétique et une méthode d'apprentissage reposant sur un algorithme de type Support Vector Machine. La mesure de Vinet est utilisée comme mesure objective pour comparer le comportement des deux approches envisagées. Nous présentons enfin les résultats que nous avons obtenus.

**Abstract** – Segmentation evaluation is a very difficult task even for an expert. The aim of this article is to propose a quality measure of a segmentation result, without any *a priori* knowledge, by taking into account different evaluation criteria from the literature. We propose in this paper to fusion the best ones by using two methods: a linear combination using a genetic algorithm and a learning method using a support vector machine algorithm. Vinet's measure is used as an objective measure to compare the behavior of the two proposed approaches. Some results are presented in the last section.

## 1 Introduction

La segmentation est une étape essentielle en traitement d'images dans la mesure où elle conditionne l'interprétation d'une image. De nombreux algorithmes ont ainsi été proposés durant les dernières décennies ([3], [5],...). Ils sont basés sur différentes approches : contour, région, texture, ... Devant la multitude des méthodes proposées, l'évaluation de la qualité d'un résultat de segmentation devient primordiale. Habituellement, l'efficacité d'un nouvel algorithme est illustrée par la présentation de quelques résultats de segmentation, ce qui n'autorise que des conclusions subjectives et qualitatives sur les performances de cet algorithme. Un critère d'évaluation de la qualité d'un résultat de segmentation peut alors présenter de nombreux intérêts. Il peut, par exemple, permettre de faciliter le choix du paramétrage d'une méthode de segmentation, ou encore autoriser la comparaison de différentes méthodes de segmentation de même niveau de précision, voire même, de définir de nouvelles méthodes de segmentation (par optimisation du critère).

De nombreux travaux traitent de ce sujet et plusieurs classes de critères d'évaluation ont été isolées ([9], [14]). Nous allons nous intéresser aux critères d'évaluation non supervisée. Ces critères ne nécessitent aucune connaissance *a priori* sur les résultats de segmentation à évaluer. Leur principe consiste à estimer la qualité d'un résultat de

segmentation à partir de statistiques calculées sur chaque région, contour ou texture détectés. Nous avons orienté notre choix vers cette classe de critères par souci de généralité de l'évaluation.

Actuellement, aucun critère d'évaluation de segmentation ne se révèle satisfaisant dans tous les cas de figure. Comme le montre Brendan McCane [2], pour évaluer correctement des résultats de segmentation, il est pertinent d'utiliser le maximum de critères d'évaluation à notre disposition et de combiner leurs résultats.

Dans la première partie de cet article, nous présentons une brève définition de ce qu'est un "bon" résultat de segmentation d'une image, puis dans un deuxième temps, le choix des critères que nous allons fusionner. La troisième partie détaille les processus de fusion mis en place. Dans la quatrième partie, nous présentons les résultats obtenus après fusion.

## 2 Définition de la segmentation

Différentes définitions de la segmentation d'une image peuvent être trouvées dans la littérature. Haralick [5], Zhang [15] et Freixenet [4] résument bien les définitions existantes. D'un point de vue général, on retrouve le fait que l'image doit être décomposée en éléments (régions ou classes) et que le calcul de propriétés à l'intérieur de chacun de ces éléments doit offrir une forte stabilité et une

disparité maximale avec celle des éléments voisins.

Ces définitions mettent en avant le fait qu'un résultat de segmentation n'est pas dépendant d'une application donnée et qu'il est défini par des propriétés calculées sur les différentes régions d'une image. Ces remarques montrent qu'il est alors possible de comparer la qualité de résultats de segmentation sans aucune connaissance *a priori* à un même niveau de précision donné. En effet, cela n'aurait aucun sens de comparer des résultats de segmentation avec des niveaux de précision complètement différents.

### 3 Définition de nouveaux critères

Avant de fusionner différents critères d'évaluation, nous avons voulu déterminer les critères les plus pertinents. Pour ce faire, nous avons utilisé une base (*BIS*) de 300 images synthétiques composées de régions uniformes et texturées. Un exemple de création de telles images est présenté figure Fig. 1. Ces images ont été segmentées par 4 méthodes de segmentation et procurent donc 1200 résultats de segmentation. La mesure de Vinet [12] (taux de classification correcte) est le critère d'évaluation supervisée employé comme référence sur ces images synthétiques.

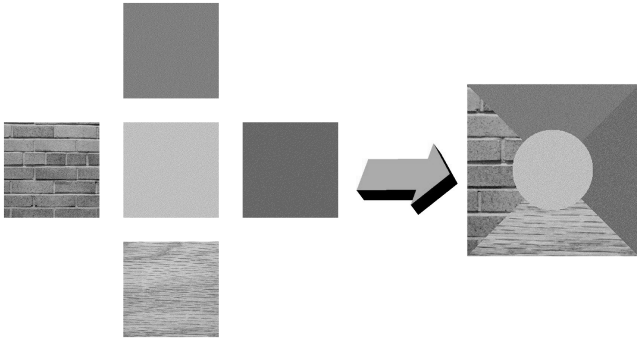


FIG. 1: Exemple de création d'une image mixte comportant 3 régions uniformes bruitées et 2 texturées pour la création de la base *BIS*.

Suite à l'étude comparative de critères d'évaluation non supervisée de résultats de segmentation présentée dans [8], les critères qui suivent le mieux l'évolution de la mesure de Vinet et que nous avons retenus pour la fusion sont : le critère de Borsotti [1], le critère de Zéboudj [13] et le critère de Rosenberger [11].

Les critères de Zéboudj et Borsotti donnent de bons résultats dans le cas d'images uniformes, et des résultats moyens dans le cas mixte (régions texturées et uniformes). Le critère de Rosenberger, quant à lui, donne de bons résultats sur les images texturées, mais se montre un peu plus faible sur les images uniformes que les critères de Zéboudj et de Borsotti.

Notre objectif est alors de fusionner ces trois critères afin d'obtenir un comportement plus proche de la mesure de Vinet. Deux approches sont envisagées : une fusion par combinaison linéaire des critères et une fusion par apprentissage. Ces approches permettront à terme d'évaluer le résultat de segmentation d'une image réelle en fusion-

nant les critères d'évaluation les plus pertinents obtenus à partir de la base d'images synthétiques.

#### 3.1 Principe de la fusion par combinaison linéaire

Cette première approche consiste à effectuer une combinaison linéaire des trois critères que nous avons retenus lors de l'étude comparative. Nous allons alors définir un nouveau critère  $C_{Fus}$  s'écrivant comme la combinaison linéaire des trois précédents :

$$C_{Fus} = a * Zéboudj + b * Borsotti + c * Rosenberger \quad (1)$$

Afin d'identifier la similarité de comportement entre ce nouveau critère et la mesure de Vinet, nous allons utiliser le facteur de corrélation. Notre objectif sera donc de déterminer les paramètres optimaux  $(a, b, c)$  tels que la valeur absolue du facteur de corrélation entre les résultats de  $C_{Fus}$  et ceux de la mesure de Vinet sur la base *BIS* soit maximale. On posera comme contrainte  $(a, b, c) \in [-1, 1]^3$ . Ce choix est arbitraire car n'importe quel espace de recherche peut être utilisé : dans un autre espace de recherche, le vecteur résultat  $(a, b, c)$  serait le même à une constante multiplicative près, et le classement des résultats serait exactement le même. Cette combinaison induit alors un nouveau critère non normalisé s'écrivant comme un barycentre des 3 critères.

Ces critères n'étant pas exprimables simplement en une formule car représentant de nombreux calculs sur toute une image, il est difficile d'utiliser des méthodes d'optimisation telles que le simplexe pour obtenir le résultat optimal. Nous avons opté pour un outil ayant fait ses preuves dans un tel domaine d'application, à savoir les algorithmes génétiques [6]. Les algorithmes génétiques déterminent les solutions d'une fonction en simulant l'évolution d'une population jusqu'à la survie des individus les plus adaptés. Les survivants sont des individus obtenus par croisement, mutation et sélection d'individus de la génération précédente. Un algorithme génétique est défini en considérant cinq données essentielles :

1. la représentation de la solution ou génotype : un vecteur de paramètres  $(a, b, c)$ ,
2. la population initiale : un ensemble de vecteurs de paramètres  $(a, b, c)$ ,
3. la fonction d'aptitude : le facteur de corrélation entre  $C_{Fus}$  de l'équation (1) (calculé sur toute la base d'images avec un vecteur de paramètres donné) et les résultats de la mesure de Vinet sur cette même base,
4. les opérations sur les génotypes : mutation, sélection ou croisement des vecteurs de paramètres,
5. critère d'arrêt de l'évolution de la population : écart type de la fonction d'aptitude.

Une fois définies ces cinq données, l'exécution d'un algorithme génétique s'effectue en cinq étapes :

1. définition de la population initiale et calcul de l'aptitude de chaque individu,

2. sélection et mutation des individus de la population courante,
3. reproduction des individus restant de la population courante,
4. évaluation des individus de la population,
5. retour à l'étape 2 si le critère d'arrêt n'est pas satisfait.

Les coefficients  $(a, b, c)$  optimisant le facteur de corrélation entre  $C_{Fus}$  et la mesure de Vinet sont alors déterminés en utilisant cette méthode [6].

### 3.2 Principe de la fusion par apprentissage

Cette seconde approche consiste en la fusion de plusieurs critères par apprentissage. Soit un ensemble de couples  $\{x_i, y_i\}_{i=1..l}$  avec  $x_i \in \mathbb{R}^d$  un vecteur de  $d$  critères décrivant la qualité d'un résultat de segmentation d'une image donnée et  $y_i$  un indice de qualité du résultat de segmentation (dans notre cas la valeur de la mesure de Vinet discrétisée entre 1 et 10). Notre objectif est alors de construire à partir de la base d'apprentissage  $\{x_i, y_i\}_{i=1..l}$  une fonction  $f$  capable de prédire l'indice de qualité d'un résultat de segmentation d'une nouvelle image  $x$ . Ainsi, notre idée est d'utiliser un apprentissage supervisé pour atteindre cet objectif mais aussi d'utiliser ce contexte pour fusionner différents critères et sélectionner les plus performants.

Ce problème d'apprentissage supervisé est un problème multiclassés avec un nombre de classes dépendant de la quantité d'indices de qualité (discrétisation en 10 classes de la mesure de Vinet pour ce qui concerne cette étude). Un problème multiclassé est résolu par une polychotomie basée sur une approche *un-contre-un* [7]. Notre algorithme d'apprentissage correspond donc à un Séparateur à Vaste Marge pour 2 classes. Ainsi, nous recherchons un hyperplan dans un espace  $\mathcal{H}$  défini comme :  $f(x) = \sum_{i=1}^{\ell} \alpha_i^* y_i K(x_i, x) + b$  qui maximise la marge entre l'hyperplan et les données projetées  $x_i$  dans  $\mathcal{H}$ . Alors, les  $\alpha_i^*$  sont les solutions du problème d'optimisation suivant :

$$\max_{\alpha_i} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (K(x_i, x_j) + \frac{1}{C} \delta_{i,j}) \quad (2)$$

avec  $\sum_i \alpha_i y_i = 0, \quad 0 \leq \alpha_i$

où  $K$  est le noyau associé à  $\mathcal{H}$ ,  $\delta_{i,j}$  est le symbole de kroenecker et  $C$  un paramètre de compromis entre la largeur de la marge et le nombre d'exemples situés au delà de la marge.

Il est à noter que tous les critères d'évaluation de résultats de segmentation sont implicitement fusionnés par le noyau  $K$ . En fait, grâce aux propriétés de  $\mathcal{H}$ , la fonction de décision peut aussi être écrite :  $f(x) = \sum_i \alpha_i y_i K(x_i, x) + b$  ce qui devient  $f(x) = \langle \sum_i \alpha_i y_i x_i, x \rangle + b$ . Ainsi, chaque critère d'évaluation  $u$  d'un résultat de segmentation  $x$  est pondéré par  $\sum_i \alpha_i y_i x_i^{(u)}$  dans la fonction de décision.

Un tel apprentissage combine implicitement plusieurs critères d'évaluation. Or, il est possible que seul un sous-ensemble de ces critères soit pertinent pour identifier l'indice de qualité du résultat de segmentation. Afin d'identifier ce sous-ensemble, nous avons réalisé une sélection de

critères. Cette sélection de critères repose sur l'influence de chaque critère sur la marge dans la classification [10]. Ainsi, chaque critère a été pondéré par un facteur d'échelle  $\sigma$  et la sensibilité de la marge par rapport au critère  $u$  est reliée à  $|\sum_i \sum_j \alpha_i \alpha_j y_i y_j \frac{\partial K(x_i, x_j)}{\partial \sigma_u}|$ . Pour plus de détails sur cette procédure de sélection de variables, le lecteur peut se référer à [10].

## 4 Résultats expérimentaux

Dans le cas de la fusion par combinaison linéaire, nous avons utilisé les résultats de 3 critères sur la base d'images *BIS* et ses 1200 résultats de segmentation présentés précédemment : le critères de Rosenberger avec paramètres automatiques, le critère de Borsotti et celui de Zéboudj.

En ce qui concerne la méthode de fusion par apprentissage présentée, nous avons utilisé les valeurs de 22 critères (le critère de Rosenberger [11] avec 19 jeux de paramètres et une version avec calcul automatique de paramètres, le critère de Borsotti [1] et celui de Zéboudj [13]) sur la même base d'images. Dans ce cas, nous avons effectué une sélection de variables pour identifier les critères les plus performants.

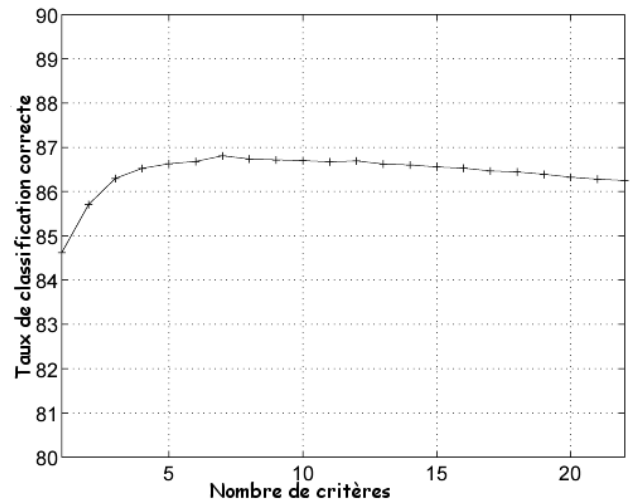


FIG. 2: Performances de l'apprentissage par SVM par rapport au nombre de critères d'évaluation utilisés pour prédire la mesure de Vinet.

Nous pouvons constater sur la figure Fig. 2 que les meilleures performances sont obtenues lorsque nous utilisons 7 critères (dont celui de Rosenberger automatique, celui de Borsotti et celui de Zéboudj). Nous avons alors sélectionné les critères correspondants et nous avons regardé la performance de notre critère fusionné pour différentes tailles de base d'apprentissage. Les résultats sont présentés dans la figure Fig. 3.

On peut remarquer que même avec très peu d'éléments dans la base d'apprentissage (1% de la base d'apprentissage) notre méthode de fusion de critères est capable de prédire avec précision la mesure de Vinet (78% de taux de classification correcte). Ainsi, nous pouvons nous contenter de quelques résultats de segmentation à comparer à une vérité terrain pour les utiliser comme base d'apprentissage.

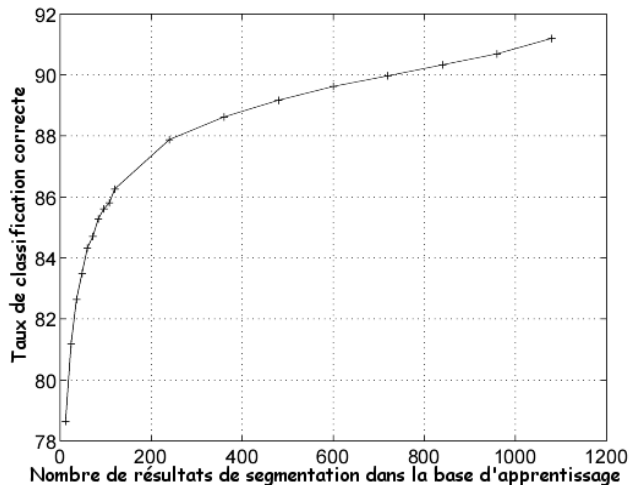


FIG. 3: Taux de classification correcte de la mesure de prédiction de la mesure de Vinet par rapport au nombre d'éléments de la base d'apprentissage.

Le tableau Tab. 1, présente les résultats obtenus par les méthodes présentées précédemment sur toute la base de résultats de segmentation.

TAB. 1: Optimisation du facteur de corrélation avec la mesure de Vinet par les deux procédés de fusion.

	Global	Mixte	Uniforme	Texturé
Meilleur résultat sans optimisation	0,7068	0,2638	0,9009	0,3554
Fusion par combinaison linéaire	0,7711	0,2934	0,9483	0,4195
Fusion par apprentissage	0,8624	0,4651	0,9784	0,5543

Comme nous pouvons le constater, la fusion par combinaison linéaire permet d'obtenir un gain de l'ordre de 10% par rapport au meilleur résultat obtenu avant fusion, tandis que nous obtenons un gain de l'ordre de 20% dans le cas de la fusion par apprentissage.

## 5 Conclusions et perspectives

De nos jours, le traitement d'images prend une importance grandissante. En vue d'interpréter des images acquises par un procédé quelconque, de nombreux algorithmes de segmentation ont été élaborés. L'évaluation de ces méthodes de segmentation est alors indispensable pour quantifier leurs performances.

La première méthode que nous avons présentée, permettant de fusionner par combinaison linéaire plusieurs critères d'évaluation de résultats de segmentation, donne un gain général de fiabilité de l'ordre de 10%, tandis que la deuxième méthode, par apprentissage, obtient un gain général de l'ordre de 20%.

Les perspectives d'une telle étude sont multiples. Elles concernent de simples applications comme l'évaluation de nouvelles méthodes de segmentation ou le paramétrage optimal d'un algorithme de segmentation, mais peut aussi permettre la définition d'un nouvel algorithme de segmen-

tation reposant sur l'optimisation du critère développé.

## 6 Remerciements

Les auteurs tiennent à remercier la région du CENTRE et le FSE pour leurs soutiens financiers.

## References

- [1] M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, 19:741–747, 1998.
- [2] B. Mc Cane. On the evaluation of image segmentation algorithms. *DICTA'97 and IVCN'Z97*, pages 455–459, october 1997.
- [3] J.-P. Cocquerez and S. Philipp. *Analyse d'Images: filtrage et segmentation*. Masson, 1995.
- [4] J. Freixenet, X. Muñoz, D. Raba, J. Marti, and X. Cufi. Yet another survey on image segmentation: Region and boundary information integration. *ECCV*, pages 408–422, 2002.
- [5] R. H. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision Graphics Image Processing*, 29:100–132, 1985.
- [6] C. R. Houck, J. A. Joines, and M. G. Kay. A genetic algorithm for function optimisation: a matlab implementation. Technical report, NCSU-IE, sept 1995.
- [7] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, (13):415–425, 2002.
- [8] H. Laurent, S. Chabrier, C. Rosenberger, B. Emile, and P. Marché. Etude comparative de critères d'évaluation de la segmentation. *Actes du 19ème Colloque GRETSI*, 3:150–153, 2003.
- [9] S. Philipp-Foliguet. Evaluation de la segmentation. Technical report, ETIS, mars 2001.
- [10] A. Rakotomamonjy. Variable selection using svm-based criteria. *Journal of Machine Learning Research*, 3:1357–1370, 2003.
- [11] C. Rosenberger. *Mise en Oeuvre d'un Système Adaptatif de Segmentation d'Images*. PhD thesis, Université de Rennes 1, décembre 1999.
- [12] L. Vinet. *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques*. PhD thesis, Université de Paris IX Dauphine, Juillet 1991.
- [13] R. Zeboudj. *Filtrage, Seuillage Automatique, Contraste et Contours: du Pré-Traitement à l'Analyse d'image*. PhD thesis, Université de Saint Etienne, 1988.
- [14] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.
- [15] Y.J. Zhang. Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters*, 18:963–974, 1997.