

# Segmentation par ligne de partage des eaux basée sur des connaissances texturales

Sébastien DERIVAUX, Sébastien LEFÈVRE, Cédric WEMMERT, Jerzy KORCZAK

Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection  
UMR 7005 CNRS - Université Louis Pasteur, Strasbourg I  
Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France  
{derivaux,lefevre,wemmert,jjk}@lsiit.u-strasbg.fr

**Résumé** – La segmentation d'une image a pour but de créer des régions homogènes selon un critère défini en fonction de l'application considérée. Le critère peut être, par exemple, les valeurs des pixels de l'image ou des indices de texture. Une fois la segmentation obtenue, une classification des régions peut être effectuée afin d'associer une sémantique aux objets présents dans l'image. L'hypothèse d'homogénéité entre des pixels d'un même objet doit permettre la construction de régions correspondant aux objets de l'image. Dans le cas d'images complexes comme les images satellites à très haute résolution spatiale, l'homogénéité classique des pixels d'un même objet n'est pas valide, réduisant la qualité des segmentations obtenues. Dans cet article, nous proposons d'intégrer au processus de segmentation par ligne de partage des eaux des connaissances sous forme d'exemples étiquetés par l'utilisateur. Nous introduisons ainsi un critère d'homogénéité de plus haut niveau basé sur des connaissances texturales.

**Abstract** – The goal of image segmentation is to create homogeneous regions using a criterion depending on the problem to be solved. For instance, this criterion can be pixel values or textural indices. Once the segmentation is obtained, a region-based classification can be computed in order to give a semantic to each object of the image. Assumption of homogeneity between pixels belonging to the same objet should allow to build regions corresponding to image objects. In the case of complex images, like very high spatial resolution remotely sensed images, classical homogeneity is no longer true, thus decreasing the segmentation quality. In this paper, we propose to use prior textural knowledge from labeled samples in a watershed segmentation algorithm. We introduce a higher level homogeneity criterion based on textural knowledge.

## 1 Introduction

L'interprétation automatique d'image devient de plus en plus importante. L'augmentation de la résolution des différents outils d'acquisition génère une masse de données qu'il n'est plus possible de traiter manuellement. Dans ce contexte, des méthodes automatiques doivent permettre d'analyser une image afin d'aider l'expert humain dans son travail d'interprétation. Le problème se définit de la façon suivante. Soit  $I$  une image à interpréter contenant un ensemble  $O = \{O_1, \dots, O_{\bar{O}}\}$  d'objets d'intérêt à détecter. Chaque objet  $O_i$  appartient à la classe  $C_{O_i}$  prise dans l'ensemble des classes d'intérêt  $C = \{C_1, \dots, C_{\bar{C}}\}$ . L'interprétation d'une image  $I$  consiste à retrouver les objets et leur classe correspondante. Comme chaque objet est représenté par plusieurs pixels, il est commun d'utiliser un algorithme de segmentation afin de créer des régions et ensuite d'identifier la classe de celles-ci à l'aide d'un algorithme de classification.

La segmentation est un processus qui regroupe les pixels spatialement voisins selon un critère d'homogénéité. Par exemple, dans une image en niveaux de gris, il est commun de regrouper les pixels voisins qui ont des intensités proches. Il est ensuite possible de caractériser les régions obtenues avec différents attributs (forme, superficie, intensité moyenne, ...) pour effectuer une classification des segments obtenus. Il est souvent plus aisé de discriminer les classes puisque les caractéristiques sont plus nombreuses

et plus variées. Cette méthode nécessite d'avoir une segmentation de bonne qualité. Une segmentation de qualité optimale est une segmentation où chaque région obtenue correspond exactement à un et un seul objet d'intérêt de l'image  $I$ .

Dans la pratique une segmentation est rarement optimale. Elle est sujette aux problèmes de sous-segmentation et/ou de sur-segmentation. La sous-segmentation intervient lorsqu'une région de la segmentation couvre plusieurs objets d'intérêt de classes différentes. Comme la classification se base sur les régions, il est évident que certaines parties de régions seront mal classées. Il convient donc d'éviter autant que possible la sous-segmentation. La sur-segmentation quant à elle intervient quand un objet d'intérêt est divisé en plusieurs régions à l'issue de la segmentation. Les caractéristiques extraites, notamment la superficie et les propriétés topologiques, sont de moins bonne qualité lorsqu'il y a sur-segmentation. Il convient donc de réduire autant que possible la sur-segmentation sans produire de sous-segmentation.

La plupart des méthodes de segmentation existantes créent des régions homogènes selon la valeur des pixels ou d'un indice de texture dérivé. L'hypothèse d'une homogénéité des objets d'intérêt selon ces critères n'est plus valide dans des images complexes. Par exemple, dans le cas des images satellites, un objet tel qu'un toit peut avoir une très grande hétérogénéité spectrale puisqu'ayant une partie à l'ombre et une autre au soleil. De ce fait, les

méthodes classiques de segmentation produisent une sur-segmentation de l'image.

Précédemment, nous avons développé un algorithme de segmentation qui s'appuie sur des connaissances exogènes à l'image sous forme de pixels dont la sémantique est connue [3, 4]. Ces exemples sont utilisés pour induire un nouveau critère d'homogénéité tenant compte des classes d'objets recherchées grâce à une classification supervisée floue. Un algorithme de ligne de partage des eaux [7] est utilisé pour segmenter l'image selon ce nouveau critère. Dans cet article, nous étendons cet algorithme pour utiliser des indices de textures en place des valeurs de chacun des pixels et nous évaluons l'apport des connaissances texturales.

Nous présenterons tout d'abord les différentes étapes de l'algorithme, avant de donner des critères d'évaluation permettant de mesurer l'apport de connaissances texturales dans la segmentation d'images satellites. Enfin une conclusion et des perspectives seront données.

## 2 Présentation de l'algorithme

L'algorithme que nous proposons se décompose en trois parties. Dans un premier temps, des indices de texture sont calculés à partir de l'image fournie en entrée. Ces d'indices sont ensuite utilisés dans une classification supervisée floue pour générer des cartes d'appartenance de chaque pixel à chaque classe. Enfin, une ligne de partage des eaux, appliquée à ces cartes d'appartenance, permet d'obtenir la segmentation finale.

### 2.1 Indices de texture

La première étape de l'algorithme consiste à convertir l'image d'entrée en une image d'indices de texture. L'utilisation d'indices de texture en lieu et place des valeurs des pixels permet d'utiliser l'information spatiale des images. De nombreux indices de texture existent. Nous avons choisi de nous contenter dans un premier temps de la moyenne, de l'écart-type, de l'asymétrie et du kurtosis sur une fenêtre glissante carrée. À chaque pixel de l'image est donc associé 4 indices de texture pour chaque bande spectrale. La fenêtre glissante est tronquée aux bords de l'image. Soit  $(a_x, a_y)$  et  $(b_x, b_y)$  les coordonnées extrêmes de la fenêtre glissante pour un pixel donné  $p$ . Soit  $v_c(d_x, d_y)$  la valeur d'un pixel de coordonnées  $(d_x, d_y)$  dans la bande  $c$  et  $N_p$  le nombre de pixels contenu dans la fenêtre glissante. Les indices texturaux de la moyenne et de l'écart-type pour le pixel  $x$  sont définis comme :

$$\mu_c(p) = \frac{1}{N_p} \sum_{i=a_x}^{b_x} \sum_{j=a_y}^{b_y} v_c(i, j)$$

$$\delta_c(p) = \frac{1}{N_p} \sum_{i=a_x}^{b_x} \sum_{j=a_y}^{b_y} (v_c(i, j) - \mu_c(p))^2$$

L'intérêt de ces indices est leur invariance à l'orientation des objets et leur calculabilité très rapide grâce aux *integral image*[8]. A chaque pixel  $v_c(x, y)$  de l'image d'origine

est associé un pixel dans l'*integral image*  $INT_c^o(x, y)$  tel que :

$$INT_c^o(x, y) = \sum_{i=0}^x \sum_{j=0}^y v_c(i, j)^o$$

ce qui représente pour  $o = 1$  la somme des de tous les pixels  $(x', y')$  tel que  $x' \leq x$  et  $y' \leq y$ . Pour  $o = 2$ , nous obtenons la somme des carrés des valeurs, et ainsi de suite. La fonction  $S_c^o$  permettant de calculer la somme des valeurs des pixels à la puissance  $o$  dans la fenêtre définie par  $(a_x, a_y)$  et  $(b_x, b_y)$  est donc :

$$S_c^o(a_x, a_y, b_x, b_y) = INT_c^o(b_x, b_y) - INT_c^o(a_x, b_y) - INT_c^o(b_x, a_y) + INT_c^o(a_x, a_y)$$

Il devient ainsi rapide de calculer les indices de textures, avec pour la moyenne et l'écart-type :

$$\mu_c(p) = \frac{1}{N_p} S_c^1(a_x, a_y, b_x, b_y)$$

$$\delta_c(p) = \frac{1}{N_p} S_c^2(a_x, a_y, b_x, b_y) - \mu_c(p)^2$$

Le même mécanisme s'applique aux indices d'asymétrie et de kurtosis.

### 2.2 Classification floue supervisée

La classification supervisée est l'une des principales techniques d'interprétation automatique d'image de télédétection. L'objectif d'un tel algorithme est d'induire un modèle de prédiction de classe. Pour ce faire, l'utilisateur fournit en entrée des exemples de pixels pour chaque classe. La classification floue quant à elle assigne à chaque pixel, non pas une classe, mais une appartenance à chaque classe d'intérêt.

Rappelons le processus de classification supervisée floue. La confiance de l'algorithme dans l'appartenance du pixel  $x$  à la classe  $C_l$  est notée  $a_{x,l}$ . Deux relations doivent être satisfaites :

$$0 \leq a_{x,l} \leq 1$$

$$\sum_{l=1}^{\bar{c}} a_{x,l} = 1 \quad \forall x$$

Si un pixel  $x$  est caractérisé par  $a_{x,l} = 1$  et  $a_{x,j} = 0 \quad \forall j \neq l$ , la confiance de l'algorithme dans la classification de  $x$  dans la classe  $C_l$  est maximale. Par contre, un pixel peut avoir une appartenance  $a_{x,l} > 0$  pour plusieurs valeurs de  $l$  : cela indique soit que le pixel appartient à une classe non considérée dans les exemples donnés soit que les attributs du pixels ne sont pas suffisants pour discriminer ces classes dans l'espace des données.

Pour obtenir les valeur d'appartenance, nous proposons d'utiliser l'algorithme des  $K$  plus proches voisins [1]. Pour chaque pixel  $x$ , les  $K$  plus proches voisins dans l'espace des données dont la classe est connue sont sélectionnés. Chacun de ces voisins  $x_k$  augmente l'appartenance du pixel  $x$  pour la classe qu'il représente, cette appartenance étant

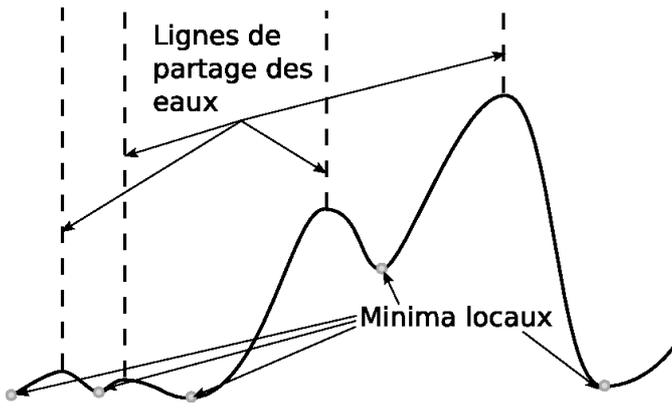


Figure 1: Exemple de coupe d'image d'élévation immergée.

pondérée par l'inverse de la distance  $d(x, x_k)$  dans l'espace des données. L'appartenance  $a_{x,l}$  est donc obtenue par :

$$a_{x,l} = \left( \sum_{k=1}^K \sum_{j=1}^{\bar{C}} w_{k,j} \right)^{-1} \sum_{k=1}^K w_{k,l}$$

$$\text{où } w_{k,j} = \begin{cases} d(x, x_k)^{-1} & \text{si } x_k \text{ est de classe } C_j \\ 0 & \text{sinon} \end{cases}$$

Une carte d'appartenance correspond à la transformation de l'image source depuis son espace initial (intensité, couleur RVB, indices de textures, ...) vers l'espace des appartenances où chaque pixel est décrit par son vecteur d'appartenance aux différentes classes.

La carte d'appartenance peut alors être segmentée par un algorithme de ligne de partage des eaux. Néanmoins, au lieu d'obtenir des segments homogènes au sens des réponses spectrales ou des indices de textures, nous obtenons des régions homogènes au sens des appartenances aux différentes classes d'intérêt.

### 2.3 Ligne de partage des eaux

La ligne de partage des eaux [7] s'applique sur une image vue comme une carte d'altitude. Cette image est généralement une image de gradient permettant d'identifier les zones homogènes et hétérogènes du signal d'entrée. L'élévation du pixel étant directement proportionnelle à l'hétérogénéité de la zone dans laquelle il se situe. Cette carte est inondée à partir de ses minima locaux et lorsque deux bassins de rétention se rencontrent, une ligne de partage des eaux est créée. Un exemple de coupe d'une image d'élévation immergée est donné en figure 1.

Cette image d'élévation peut être générée à l'aide d'un gradient morphologique [6] calculé marginalement sur chaque bande de l'image puis d'une combinaison des élévations des différentes bandes à l'aide d'une norme euclidienne.

## 3 Évaluation

Nous allons maintenant définir des critères d'évaluations et évaluer l'algorithme précédent sur des données réelles.

### 3.1 Critères d'évaluation

L'évaluation de segmentation n'est pas clairement définie. Dans cet article, des méthodes d'évaluation empiriques sont utilisées, adaptées de celles présentées dans Carleer et al. [2] pour utiliser une segmentation de référence partielle. Le premier critère mesure la sur-segmentation ( $SS$ ) pour chaque classe :

$$SS = \frac{1}{\bar{C}} \sum_{l=1}^{\bar{C}} \frac{N_{Sl}}{N_{Rl}}$$

où  $l$  est la classe considérée,  $N_{Sl}$  le nombre de régions segmentées qui contiennent au moins un pixel étiqueté de la classe  $l$  et  $N_{Rl}$  le nombre de régions expertes pour la classe  $l$ . Plus la sur-segmentation est forte, plus les objets sont fractionnés en plusieurs régions lors de la segmentation. Il faut également évaluer la sous-segmentation, qui apparaît quand un segment contient des pixels de régions expertes de classes différentes. Dans ce cas, la précision maximale ( $PM$ ) d'une classification ultérieure par régions est réduite. Par exemple, si la  $PM$  est de 63 %, la précision de la classification par régions de l'image ne pourra en aucun cas dépasser 63 %. Pour ce critère, l'étiquette d'un segment est définie comme la classe qui a le plus grand nombre de pixels au sein de ce segment. Une matrice de confusion  $C$  peut être extraite à partir de cette classification en  $\bar{C}$  classes, où la classe  $C_{ij}$  représente le nombre de pixels assignés à la classe  $i$  mais appartient en réalité à la classe  $j$ . La  $PM$  est définie par :

$$PM = \frac{\sum_{i=0}^{N_{Ref}} C_{ii}}{\sum_{i=0}^{N_{Ref}} \sum_{j=0}^{N_{Ref}} C_{ij}}$$

Le dernier critère d'évaluation est la précision empirique ( $PE$ ). Un classifieur de type K plus proches voisins [1] est utilisé sur les régions produites par la segmentation. Chaque région est représentée par sa réponse spectrale moyenne dans chaque bande. L'ensemble d'apprentissage s'appuie sur les régions où une classe est majoritaire parmi les étiquettes des pixels. Cette procédure de classification est simpliste, de nombreux autres attributs pouvant être extraits des segments. Néanmoins, nous présentons avant tout une nouvelle méthode de segmentation et les résultats sont suffisants pour démontrer l'intérêt de notre approche.

### 3.2 Résultats

La méthode proposée a été évaluée sur une image de Strasbourg prise par le satellite Quickbird. Les capteurs Quickbird renvoie une bande panchromatique d'une résolution de 70cm et 3 bandes couleurs d'une résolution de 2,8 mètres. Les bandes couleurs ont été fusionnées avec la bande panchromatique pour obtenir une image à 4 bandes d'une résolution de 70cm [5]. La taille de l'image est de  $900 \times 900$  pixels et la résolution spectrale est de 8 bits par pixels.

Pour les procédures d'évaluation et d'apprentissage, 3 zones sont étiquetées par un expert. Les pixels dont la

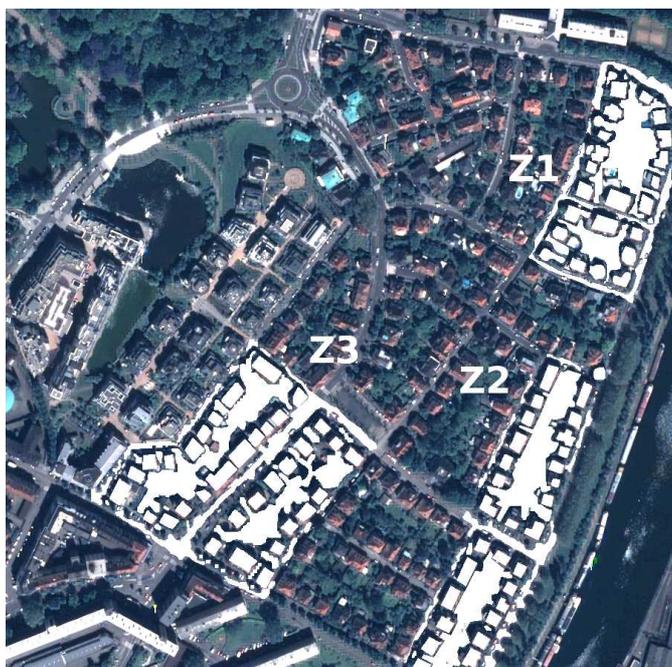


Figure 2: Image de test et ses 3 zones (Z1, Z2 et Z3) d'apprentissage et d'évaluation en surbrillance.

classe n'a pas pu être déterminée (cas des pixels mixtes ou d'objets qui ne sont pas d'intérêt comme les voitures) ne sont pas étiquetés.

Sur cette image, 4 algorithmes ont été appliqués :

- **Segmentation spectrale** : La ligne de partage des eaux est appliquée directement sur l'image.
- **Segmentation spectrale supervisée** : La ligne de partage des eaux est appliquée sur le résultat de la classification supervisée floue avec un classifieur 5 plus proches voisins.
- **Segmentation texturale** : La ligne de partage des eaux est appliquée sur l'image des indices de texture. La taille de la fenêtre utilisée est  $9 \times 9$  (taille obtenant les meilleurs résultats).
- **Segmentation texturale supervisée** : L'algorithme complet tel qu'il a été décrit à la section 2 est appliqué avec  $K = 5$ . La taille de la fenêtre utilisée est  $11 \times 11$  (taille obtenant les meilleurs résultats).

Les résultats sont donnés dans le tableau 3.2. Les méthodes supervisées donnent de meilleurs résultats que leurs équivalents non supervisés pour les 3 critères d'évaluations. L'utilisation des indices de textures en lieu et place des réponses spectrales augmente la précision empirique et la sur-segmentation tout en conservant une sous-segmentation (critère  $PM$ ) marginale.

## 4 Conclusion

Dans cet article, nous avons proposé une méthode de segmentation utilisant des connaissances texturales définies à partir des pixels étiquetés a priori et d'un classifieur flou basé sur des indices de texture. Celui-ci permet d'affecter

Table 1: Résultats des segmentations (écart-type entre parenthèses)

Method	PM	PE	SS
spectrale	98.98% (0.18)	87.10% (3.18)	50.5 (12.0)
spectrale supervisée	99.02% (0.12)	89.84% (3.38)	20.9 (8.0)
texturale	97.81% (0.32)	85.48% (5.08)	32.9 (7.7)
texturale supervisée	98.84% (0.04)	90.67% (2.64)	17.9 (10.0)

à chaque pixel une appartenance aux classes d'intérêt. Une ligne de partage des eaux est appliquée pour produire une segmentation basée sur l'homogénéité des appartenances aux classes texturales.

Néanmoins, l'algorithme des plus proches voisins utilisé se prête mal à une augmentation des attributs. Nous envisageons aussi d'utiliser un algorithme génétique pour trouver les pondérations optimales de plusieurs indices de textures.

## References

- [1] David W. Aha, Dennis F. Kibler, and Marc K. Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.
- [2] A. P. Carleer, Olivier Debeir, and E. Wolff. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, 71(11):1285–1294, 2005.
- [3] S. Derivaux, S. Lefèvre, C. Wemmert, and J. Korczak. Watershed segmentation of remotely sensed images based on a supervised fuzzy pixel classification. In *Proceedings of IEEE International Geosciences and Remote Sensing Symposium (IGARSS)*, 2006.
- [4] S. Derivaux, S. Lefèvre, C. Wemmert, and J. Korczak. On machine learning in watershed segmentation. In *Proceedings of IEEE Workshop on Machine Learning for Signal Processing*, 2007.
- [5] T. Ranchin, B. Aiazzi, L. Alparone, and S. Baronti. Image fusion. The ARSIS concept and some successful implementation schemes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1):4–18, 2003.
- [6] Pierre Soille. *Morphological Image Analysis*. Springer-Verlag, 2nd edition, 2003.
- [7] Lee Vincent and Pierre Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.
- [8] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.