

# Partitions d'Énergie et Segmentation d'Images

## *Energy Partitions and Image Segmentation*

Pablo Andrés Arbeláez

Laurent D. Cohen

CEREMADE, UMR 7534 Université Paris Dauphine  
Place du maréchal de Lattre de Tassigny,  
75775 Paris cedex 16, France

{arbelaez, cohen}@ceremade.dauphine.fr

### Résumé

Nous présentons une méthode de segmentation d'images de bas niveau. Le problème est formulé en termes de la partition induite par un ensemble de sources et une certaine énergie. Nous étudions deux types d'énergies qui traitent l'image à des niveaux d'analyse différents. La première classe est issue de la théorie des chemins minimaux et produit une présegmentation de l'image. Dans la deuxième partie, où l'énergie est définie au moyen d'un processus de fusion de régions, l'image simplifiée est utilisée pour créer des hiérarchies de partitions

### Mots Clés

segmentation, mosaïque des extrema, variation de chemin, ligne de partage des eaux, morphologie mathématique

### Abstract

*We present a method for low-level image segmentation. The problem is formulated in terms of the partition induced by a set of sources and a determined energy. We study two types of energies that process the image at different levels of analysis. The first class is based on the theory of minimal paths and produces a pre-segmentation of the image. In the second part, where the energy is defined through a region merging process, the simplified image is used to create hierarchies of partitions.*

### Keywords

*segmentation, extrema mosaic, path variation, watershed, mathematical morphology.*

## 1 Introduction

La segmentation d'images est un problème fondamental dans le domaine de la vision par ordinateur. Sa grande complexité peut s'expliquer par le fait que créer une partition du domaine de l'image en régions "importantes" nécessite une certaine compréhension du contenu sémantique de la scène représentée. Une étude récente [21] a mis en évidence que même l'évaluation de la segmentation humaine est un problème difficile où la reconnaissance et la subjectivité semblent jouer un rôle prépondérant. En conséquence, l'introduction d'information sémantique s'avère comme une étape cruciale dans l'élaboration de tout système de vision par ordinateur de haut niveau.

Une première tâche est néanmoins l'extraction de l'information fournie par les données de l'image sans connaissance préalable de son contenu. Le présent travail traite un tel problème de segmentation de bas niveau pour des images en niveaux de gris. L'approche proposée se fonde sur une formulation du problème consistant à trouver les régions d'attraction d'un ensemble de points source, déterminées par une énergie donnée. Ainsi, le problème se ramène à la définition de l'énergie et à la sélection des sources.

La conception de la vision comme une tâche de traitement d'information [20] suggère l'emploi d'une stratégie à plusieurs niveaux pour aborder les problèmes de la vision par ordinateur. Suivant cette idée, notre méthode est divisée en deux parties qui traitent l'image à des niveaux d'analyse différents. Bien qu'indépendantes, les deux étapes sont présentées dans un cadre commun, décrit dans la Section 2, et correspondent au choix de deux énergies et ensembles de sources différents. Le premier type d'énergies abordé se fonde sur la théorie des chemins minimaux ; un cas particulier, la *variation de chemin*, est étudié dans la Section 3. Cette énergie est une généralisation de la variation totale en dimension 1 aux fonctions définies sur le plan. Dans la Section 4, le choix des extrema de l'intensité en tant que sources fournit une simplification initiale de l'image, que nous appelons la *mosaïque des extrema*. Dans la deuxième partie, traitée dans la Section 5, l'énergie est définie au moyen d'un processus de fusion de régions. Cette *énergie de fusion* est employée pour la création de familles de partitions emboîtées d'où la segmentation finale est choisie.

Notre méthode est proche de certaines techniques de segmentation issues de la morphologie mathématique. En particulier, la ligne de partage des eaux [2] et les hiérarchies construites par l'inondation du gradient [31, 23] ; une comparaison avec ces approches est aussi établie.

## 2 Distances et Énergies

Cette section présente le cadre général dans lequel nous plaçons ; les définitions de base sont rappelées et les notations établies.

## 2.1 Partitions Induites par une Energie

Soit  $\Omega \subset \mathbb{R}^2$  un domaine connexe compact du plan. Un chemin entre deux points  $x$  et  $y$  est une fonction injective continue  $\gamma : [0, L] \rightarrow \Omega$ , telle que  $\gamma(0) = x$  et  $\gamma(L) = y$ . Le chemin est paramétré par l'abscisse curviligne, de sorte que  $L$  correspond à sa longueur euclidienne. L'image de  $\gamma$  est donc un arc simple rectifiable dans le domaine. L'ensemble des chemins reliant  $x$  à  $y$  est noté par  $\Gamma_{xy}$ . Soit  $d : \Omega \times \Omega \rightarrow \mathbb{R}^+$  une application symétrique

$$d(x, y) = d(y, x), \forall x, y \in \Omega,$$

et qui satisfait l'inégalité triangulaire

$$d(x, y) \leq d(x, z) + d(z, y), \forall x, y, z \in \Omega.$$

Considérons l'espace quotient  $\hat{\Omega} = \{\hat{x} | x \in \Omega\}$ , obtenu en prenant les classes d'équivalence  $\hat{x} = \{y \in \Omega | d(x, y) = 0\}$ .

Alors, l'application  $\hat{d} : \hat{\Omega} \times \hat{\Omega} \rightarrow \mathbb{R}^+$ , donnée par la formule  $\hat{d}(\hat{x}, \hat{y}) = d(x, y)$  est, par définition, une distance sur  $\hat{\Omega}$ . On suppose que chaque paire de points dans  $\hat{\Omega}$  est reliée par un chemin qui minimise la distance  $\hat{d}$ .

L'énergie par rapport à la source  $x_0 \in \Omega$  induite par  $\hat{d}$  est l'application  $E_0 : \Omega \rightarrow \mathbb{R}^+$  définie comme la distance du point à la source :

$$E_0(x) = \hat{d}(x_0, \hat{x}).$$

L'énergie par rapport à un ensemble de sources  $S = \{x_i\}_{i \in J}$  est donnée par l'énergie individuelle minimale :

$$E_S(x) = \inf_{i \in J} E_i(x).$$

En présence de plusieurs sources, une information importante est fournie par l'interaction des sources, laquelle est exprimée par leurs régions d'attraction :

$$R_i = \{x \in \Omega | E_i(x) < E_j(x), \forall j \in J\}, \forall i \in J$$

Ainsi, la région d'attraction, ou simplement la région, d'une source est un sous ensemble connexe du domaine complètement déterminé par l'énergie et les autres sources. L'union des régions est notée par :

$$R(E, S) = \bigcup_{i \in J} R_i.$$

L'ensemble isoénergétique est défini comme le complémentaire de  $R(E, S)$  :

$$I(E, S) = \{x \in \Omega | \exists i, j \in J : E_S(x) = E_i(x) = E_j(x)\}.$$

Ainsi, l'union des régions d'attraction et de l'ensemble isoénergétique constitue une partition d'énergie du domaine :

$$\Pi(E, S) = R(E, S) \cup I(E, S) = \Omega.$$

## 2.2 Chemins Minimaux

Dans ce paragraphe, nous nous intéressons aux énergies qui peuvent s'exprimer en termes d'un potentiel.

L'action minimale du potentiel  $P : \Omega \rightarrow \mathbb{R}^+$  par rapport au point source  $x_0$ , évaluée en  $x$  est définie par

$$E_0(P)(x) = \inf_{\gamma \in \Gamma_{x_0x}} \int_0^L P(\gamma(s)) ds . \quad (1)$$

Dans ce cas,  $\hat{x}$ , la classe du point  $x$ , correspond à l'ensemble connexe maximal à potentiel nul qui contient  $x$ . Si le potentiel est strictement positif, alors l'espace quotient  $\hat{\Omega}$  coïncide avec le domaine. La distance correspondante est appelée une distance pondérée [6].

A titre d'exemple, si  $P \equiv 1$ , alors l'énergie en  $x$ ,

$$E_0(P)(x) = \inf_{\gamma \in \Gamma_{x_0x}} \int_0^L ds$$

devient la distance géodésique à la source, ou la longueur euclidienne du plus court chemin reliant  $x_0$  et  $x$ . Si, en plus, le domaine est convexe, alors  $E_0(P)$  coïncide avec la distance euclidienne. En présence de plusieurs sources, l'ensemble isoénergétique correspond dans ce cas au diagramme de Voronoï, aussi appelé squelette par zones d'influence [28], et les régions d'attraction aux cellules de Voronoï.

Une autre interprétation de ce type d'énergies provient du domaine de l'optique géométrique, où  $P$  représente la valeur de l'indice de réfraction du milieu optique et  $E_0(P)$ , nommée l'eikonale dans ce contexte, fournit le temps de parcours des rayons lumineux. On peut alors montrer que  $E_0(P)$  est une solution de l'équation :

$$\|\nabla E_0\| = P \quad (2)$$

avec condition initiale  $E_0(x_0) = 0$ .

Cette remarque permet le calcul de l'énergie en employant la méthode du *Fast Marching* de Sethian [30, 6]. Cette méthode utilise le fait que l'information se propage vers l'extérieur à partir des sources pour construire une approximation correcte de la solution de viscosité de (2).

Dans le cas général, une action minimale s'écrit sous la forme

$$E_0(x) = \inf_{\gamma \in \Gamma_{x_0x}} \int_0^L P(\gamma(s), \dot{\gamma}(s)) ds . \quad (3)$$

Cependant, lorsque le potentiel dépend de la direction du chemin, la construction de l'énergie au moyen du *Fast Marching* n'est habituellement plus possible.

Parmi les nombreuses applications des chemins minimaux au traitement d'images, on peut citer le minimum global pour les modèles de contours actifs [6], le *shape from shading* [17] et le groupement perceptuel [4].

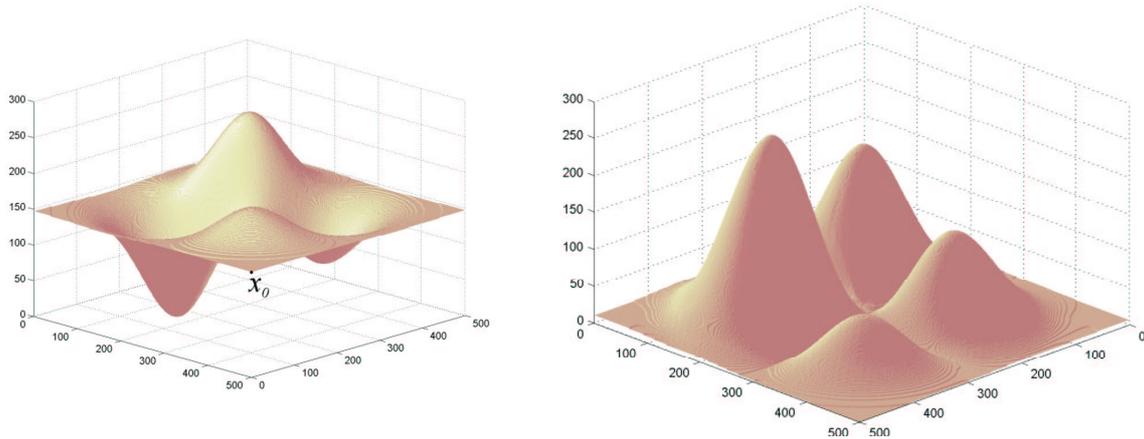


FIG. 1 – Graphes de  $u$  et  $V_0(u)$ .

### 3 La Variation de Chemin

Lors de l'application de la technique des chemins minimaux à l'analyse d'images, une grande partie du problème consiste à construire un potentiel approprié pour une situation et un type d'images spécifiques. Néanmoins, dans cet article, une perspective différente est adoptée car les notions introduites dans la section précédente sont employées pour l'étude d'une énergie particulière, dont la définition dépend uniquement des propriétés géométriques de l'image.

#### 3.1 Domaine Continu

Pour les fonctions d'une variable réelle, la variation est une fonctionnelle qui possède des propriétés connues [12, 26]. Elle a été introduite par Jordan [15] sous la forme suivante : Soit  $f : [0, L] \rightarrow \mathbb{R}$  une fonction,  $\sigma = \{s_0, \dots, s_n\}$  une partition finie de  $[0, L]$  telle que  $0 = s_0 < s_1 < \dots < s_n = L$  et  $\Phi$  l'ensemble de ces partitions.

La *variation totale* de  $f$  est définie par la formule

$$v(f) = \sup_{\sigma \in \Phi} \sum_{i=1}^n |f(s_i) - f(s_{i-1})| .$$

Dans le cas où  $f \in \mathcal{C}^1([0, L])$ , alors la variation peut s'exprimer à partir de sa dérivée :

$$v(f) = \int_0^L |f'(s)| ds . \quad (4)$$

La variation de chemin est une généralisation de la variation totale aux fonctions définies sur le plan, en considérant la variation minimale de la fonction sur tous les chemins entre deux points :

**Définition 1.** La *variation de chemin* d'une fonction  $u : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  par rapport à une source  $x_0$ , évaluée en  $x$ , est définie par

$$V_0(u)(x) = \inf_{\gamma \in \Gamma_{x_0 x}} v(u \circ \gamma) .$$

On dira que  $u$  est une *fonction à variation de chemin bornée* si elle satisfait la condition suivante :  $\forall x_0, x \in \Omega, \exists \hat{\gamma} \in \Gamma_{x_0 x} : V_0(u)(x) = v(u \circ \hat{\gamma}) < \infty$ .

Dans la suite, on suppose que  $u$  est une fonction à variation de chemin bornée.

Si  $u$  est une fonction continuellement différentiable, alors (4) permet d'obtenir l'expression suivante pour la variation de chemin :

$$V_0(u)(x) = \inf_{\gamma \in \Gamma_{x_0 x}} \int_0^L |D_{\dot{\gamma}} u(\gamma(s))| ds , \quad (5)$$

Ainsi,  $V_0$  peut être vue comme l'action minimale du potentiel  $P = |D_{\dot{\gamma}} u|$ , la valeur absolue de la dérivée directionnelle de  $u$  dans la direction tangente au chemin.

La notion intuitive exprimée par la variation de chemin est illustrée dans la Fig. 1 : si on considère une particule se déplaçant sur le graphe de la fonction, montré à gauche, et partant de la source  $x_0$ , alors la valeur de  $V_0(x)$ , à droite, représente la somme minimale des montées et des descentes réalisées pour atteindre  $x$ .

La variation de chemin correspond à la notion de variation linéaire introduite par Kronrod, bien que sous une formulation différente, dans le cadre d'une théorie géométrique pour les fonctions de deux variables [18].

La *composante* de  $u$  qui contient  $x$ , notée  $K_x$ , est définie comme le sous-ensemble connexe maximal de  $\Omega$  tel que  $u(y) = u(x), \forall y \in K_x$ . Les composantes d'une fonction continue sont fermées et disjointes deux à deux. Pour les fonctions  $\mathcal{C}^1$ , les composantes des niveaux non singuliers (*c.a.d.* les niveaux  $t$  tels que  $0 \notin \nabla u(u^{-1}(t))$ ) coïncident avec les lignes de niveau de  $u$  et peuvent être décrites par des courbes de Jordan.

Une conséquence immédiate de la Définition 1 est la propriété suivante :

**Proposition 1.** La *variation de chemin agit sur les composantes de la fonction* :

$$\forall x_0, x, y \in \Omega, K_x = K_y \Rightarrow V_0(u)(x) = V_0(u)(y).$$

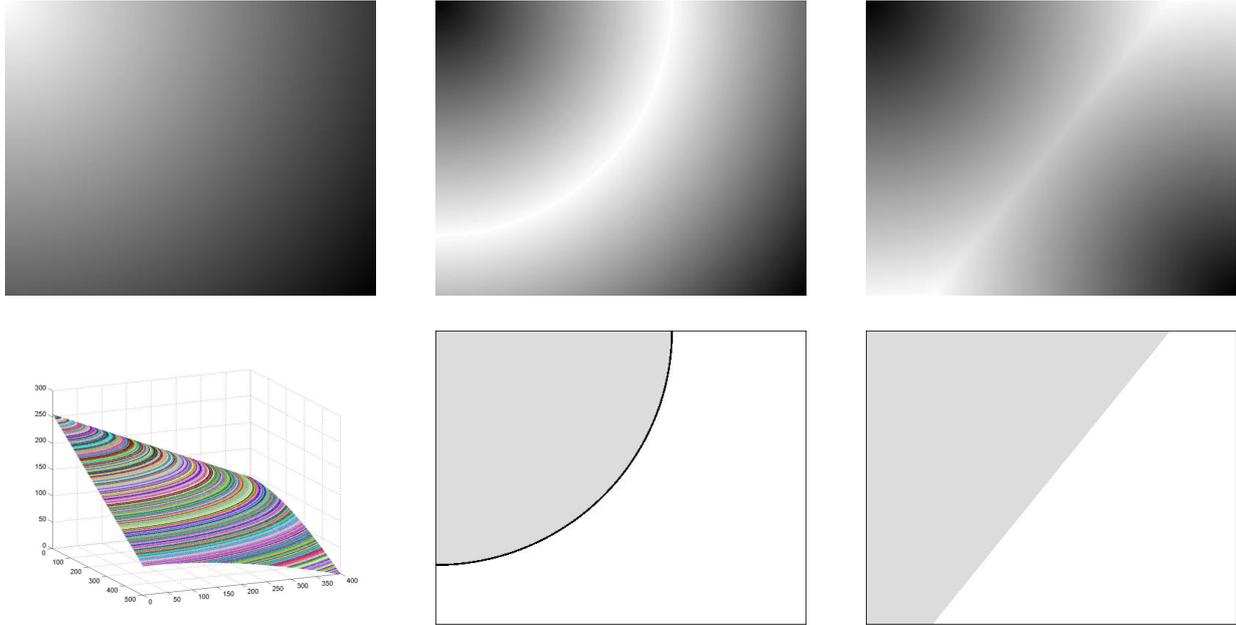


FIG. 2 – En haut :  $u$ ,  $V_S(u)$  et  $W_S(u)$ . En bas : graphe de  $u$ , partitions d'énergie  $\Pi(V(u), S)$  et  $\Pi(W(u), S)$ .

Donc,  $\hat{x} = K_x$  et l'espace quotient  $\hat{\Omega}$  coïncide avec l'ensemble de composantes de  $u$ . Ainsi, chaque élément des partitions d'énergie associées à la variation de chemin est une union de composantes de la fonction. En conséquence, l'opérateur qui associe la partition  $\Pi(V(u), S)$  à un ensemble de sources  $S$  est connexe [29].

### 3.2 Comparaison avec la Distance Image

Dans le contexte de la morphologie mathématique, l'énergie associée au potentiel  $P = \|\nabla u\|$  et donnée par la formule :

$$W_0(u)(x) = \inf_{\gamma \in \Gamma_{x_0 x}} \int_0^L \|\nabla u(\gamma(s))\| ds$$

a été utilisée pour la définition de la ligne de partage des eaux dans le domaine continu [24, 22]. Si, comme dans la classe des fonctions de Morse,  $u$  possède uniquement des points critiques isolés, alors l'espace quotient pour  $W_0$  coïncide avec le domaine et la fonction distance correspondante est appelée la *distance image* [24] ou la *distance topographique* [22].

Dans ce cadre régulier, la relation entre  $W$  et  $V$  est donnée par la propriété suivante :

$$|u(x) - u(x_0)| \leq V_0(x) \leq W_0(x), \quad \forall x \in \Omega.$$

En particulier, si  $x$  et  $x_0$  appartiennent à une ligne maximale du gradient de  $u$ , alors

$$|u(x) - u(x_0)| = V_0(x) = W_0(x), \quad (6)$$

car, par définition,  $|D_{\dot{\gamma}}(u)| = \|\nabla u\|$  lorsque  $\dot{\gamma}$  est parallèle au gradient.

L'effet produit par ces deux énergies est comparé en utilisant l'image test dans la colonne de droite de la Fig. 2. Dans ce cas,  $S = \{x_0, x_1\}$ , où  $x_0$  est le sommet supérieur gauche du domaine et  $x_1$  l'inférieur droit ; la fonction est donnée par la formule  $u(x) = c\|x - x_0\|$ . La colonne centrale montre l'effet de la variation de chemin : comme conséquence de la Prop. 1, dans cet exemple,  $u$  et  $V_S(u)$  ont les mêmes composantes et seul leur niveau est modifié. L'ensemble isoénergétique  $I(V(u), S)$ , montré en noir, est la composante dont le niveau est la moyenne des niveaux des sources. Dans la colonne de droite on peut observer que,  $\|\nabla u\|$  étant constant,  $W_S(u)$  est proportionnelle à la distance euclidienne à la source la plus proche et  $I(W(u), S)$  correspond à la médiatrice entre les sources. Néanmoins, à cause de la quantification, celui-ci est absent de la partition. Une remarque importante est que toute fonction  $u$  pour laquelle  $\|\nabla u\|$  est constant produirait la même partition d'énergie selon  $W$ . Cet exemple illustre comment  $\Pi(V(u), S)$ , la partition induite par  $V$ , préserve mieux le contenu de l'image que  $\Pi(W(u), S)$ .

### 3.3 Domaine Discret

Dans ce paragraphe, nous proposons une définition discrète de la variation de chemin. Bien que celle-ci aurait pu être prise comme définition de départ, l'interprétation continue des paragraphes précédents permettra d'éclairer les liens entre notre approche et celle de la ligne de partage des eaux.

Soit  $u$  une image définie sur une grille d'échantillonnage discrète. Dans ce contexte, le choix d'une connexité discrète définit un graphe d'adjacence  $G$  où les noeuds sont des composantes discrètes et le lien entre deux compo-

santes voisines  $K_p$  et  $K_q$  est pondéré par  $|u(p) - u(q)|$ . Puisque  $V$  agit sur les composantes de la fonction,  $G$  est la contrepartie discrète de l'espace quotient  $\hat{\Omega}$  pour  $V$ . Ainsi, nous proposons de définir la variation de chemin discrète directement sur  $G$ .

Un chemin dans  $G$  reliant les composantes des points  $p$  et  $q$  est un ensemble  $\gamma = \{K_0, \dots, K_n\} \subseteq G$  tel que :

$K_p = K_0, K_n = K_q, K_{i-1}$  et  $K_i$  sont voisines,  $\forall i = 1, \dots, n$ . L'ensemble de ces chemins est noté par  $\Gamma_{pq}^G$ . Chaque chemin sur  $G$  correspond donc à une classe de chemins discrets entre  $p$  et  $q$ .

Ainsi, l'expression discrète de la variation de chemin de  $u$  au point  $q$ , par rapport à la source  $p$ , devient

$$V_p(u)(q) = \min_{\gamma \in \Gamma_{pq}^G} \sum_{i=1}^n |u(p_i) - u(p_{i-1})| ,$$

où  $p_i \in K_i, \forall i = 1, \dots, n$ .

En conséquence, la construction de  $V_p(u)(q)$  équivaut à trouver le chemin de coût minimal sur un graphe. Ce problème classique peut être résolu en employant, par exemple, l'algorithme de Dijkstra [8, 19]. La complexité de cette implémentation de la variation de chemin est donc  $O(N \log(N))$ , où  $N$  est le nombre de composantes de l'image. De plus, si  $u$  prend des valeurs entières, le tri dans l'actualisation de la bande étroite peut être supprimé et la complexité est réduite à  $O(N)$ .

Remarquons que la distance image  $W$  peut s'implémenter au moyen du Fast Marching avec une complexité de  $O(M \log(M))$ , où  $M$  est le nombre total de pixels dans l'image. Néanmoins, puisque le potentiel de la variation de chemin dans (5) dépend non seulement de la position mais aussi de la direction du chemin, sa construction avec cette méthode n'est pas possible.

## 4 La Mosaïque des Extrema

### 4.1 Définition

Dans cette section, la partition d'énergie induite par la variation de chemin est appliquée à la simplification d'images à niveaux de gris. Dans ce but, une fois l'énergie définie, la deuxième étape consiste à choisir un ensemble de sources. La variation de chemin, comme c'est souvent le cas pour les actions minimales, est un outil intéressant pour un niveau local d'analyse de l'image. Ceci est dû au fait que leur définition dans (3) est basée sur une intégration le long des chemins. Ainsi, les régions obtenues avec un faible nombre de sources ne sont pas toujours significatives. En plus, le fait de remplacer une source  $x_i \in S$  par un autre point  $x'_i \in R(x_i)$  modifie en général la partition d'énergie correspondante.

En conséquence, les sources doivent être choisies avec soin. D'une part, elles doivent être physiquement représentatives du contenu de l'image. D'autre part, chaque structure significative doit en contenir au moins une. Puisqu'ils satisfont ces conditions, les extrema de l'intensité s'avèrent être des candidats naturels pour les sources.

**Définition 2.** La *Partition des Extrema (PE)* d'une image  $u$  est définie comme la partition d'énergie  $\Pi(V(u), ext(u))$ , où  $ext(u)$  dénote l'ensemble des composantes extrémales de  $u$ .

Donc, la *PE* est composée de deux types d'éléments. D'une part, les régions d'attraction des extrema, que nous désignerons sous le nom de *régions élémentaires* ou *atomes* de l'image. D'autre part, la Prop. 1 implique que  $I(V(u), ext(u))$ , l'ensemble isoénergétique, est une union de composantes, appelées *composantes frontière* des régions.

La Figure 3 illustre l'approche dans un exemple simple. En haut, la fonction  $u$  est un noyau gaussien où les seules composantes extrémales sont le centre et le bord du domaine carré. Donc, le graphe de  $V_{ext(u)}(u)$  a la forme de volcan montrée en bas à gauche. La partition des extrema, en bas à droite, est composée de deux régions élémentaires séparées par une composante frontière circulaire.

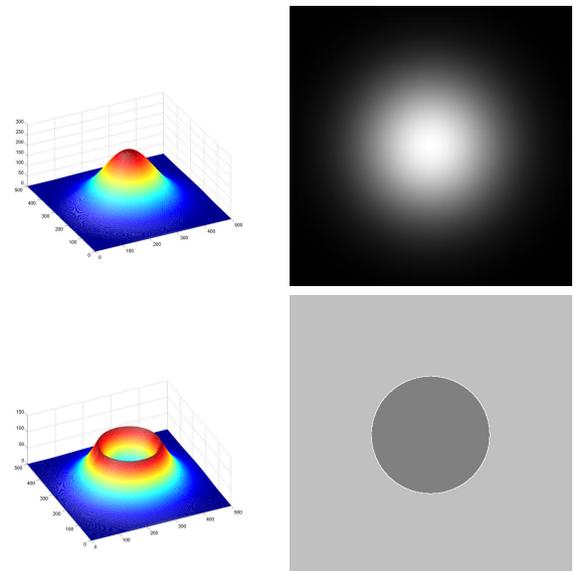


FIG. 3 – En haut : graphe et intensités d'un noyau gaussien. En bas : graphe de l'énergie et partition des extrema.

Etant définie à partir de la variation de chemin, la partition des extrema est invariante par rotations et changements de contraste linéaires. En particulier,  $PE(u) = PE(-u)$  et la *PE* traite les régions sombres et claires de façon symétrique.

En pratique, dans le but de produire une partition composée uniquement de régions élémentaires, les composantes frontière sont assimilées à une de leurs régions d'attraction voisines. Ce choix, bien qu'arbitraire, n'affecte que très peu le résultat.

La dernière étape pour créer une version simplifiée d'une image  $u$  est l'assignation d'un niveau de gris à chaque région élémentaire de  $PE(u)$ . Cette nouvelle image sera nommée la *mosaïque des extrema* de  $u$  et notée  $ME(u)$ . Pour ce point, le niveau de gris moyen ou médian sur l'atome peut être pris. Cependant, en général, la valeur de

$u$  à l'extremum représente bien les niveaux dans l'atome. Dans la suite, ce dernier choix a été retenu car, en plus, la mosaïque correspondante a la propriété suivante :

Si deux extrema de  $u$ ,  $e_0$  et  $e_1$ , sont reliés par un chemin monotone, alors la variation de chemin entre eux est préservée dans  $EM$  :

$$V_0(u)(e_1) = V_0(EM(u))(e_1). \quad (7)$$

Cette propriété peut s'interpréter comme la concentration du contraste dans les bords des atomes, comme le montre la Fig. 4. A gauche, on retrouve l'image originelle de la Fig. 3. Au centre, la mosaïque des extrema est composée par un disque blanc sur un fond noir. La variation de chemin entre les deux extrema est de 255 aussi bien dans  $u$  que dans  $EM$ . Finalement, l'image d'intensités de l'énergie est montrée à droite.

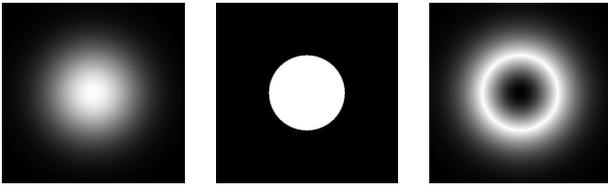


FIG. 4 – Image, mosaïque des extrema et énergie.

Puisque la mosaïque des extrema est un opérateur connexe, son application simplifie l'image tout en préservant l'information des contours.

La première ligne de la Fig. 5 montre le type de simplification obtenue sur une image réelle, en termes de réduction du nombre de composantes. A gauche, les 83742 composantes en 8-connexité d'une image de taille  $470 \times 472$  sont montrées en table de valeurs aléatoire (TVA). A droite, la partition des extrema possède 1231 composantes. La deuxième ligne présente l'image originelle à gauche et la mosaïque des extrema à droite. Trois propriétés importantes de la méthode sont illustrées dans cet exemple : D'abord, les frontières des atomes fournissent un ensemble de contours fermés qui décrivent les caractéristiques sémantiques importantes des bords comme les jonctions en T et les coins. Ensuite, on peut observer dans le fond que le flou de l'image originelle est diminué ; en effet, puisque les composantes qui appartiennent à des contours flous ou à des zones de transition ne sont pas extrémales, elles sont absorbées par un des extrema voisins. Enfin, comme conséquence de (7), le contraste des formes claires sur les ailes obscures du papillon est rehaussé.

Ainsi, la mosaïque des extrema est une décomposition de l'image en régions élémentaires où un compromis entre simplification et conservation du contenu est atteint. Elle sera utilisée comme une présegmentation sans paramètres dans la Section 5.

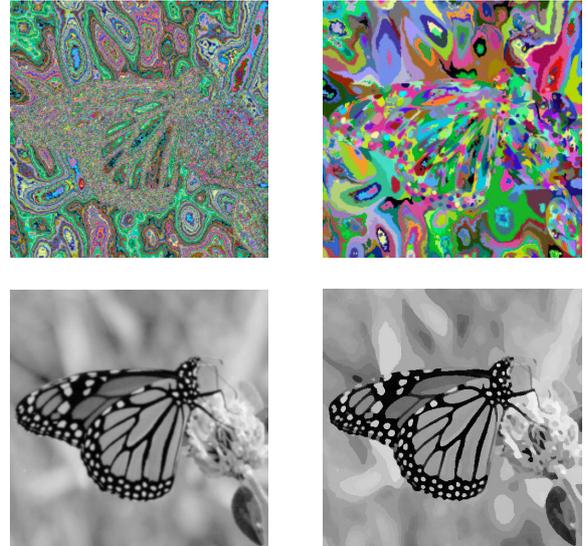


FIG. 5 – En haut : composantes de l'image originelle et PE en TVA. En bas : image originelle et mosaïque des extrema.

## 4.2 Mosaïque des Extrema et Ligne de Partage des Eaux

Un aspect essentiel en morphologie mathématique est la relation qui lie les concepts à l'intuition. Un exemple illustratif est l'analogie géographique qui considère le graphe d'une image comme une surface topographique, en identifiant le niveau de gris à une altitude. Bien plus qu'une simple métaphore, cette perspective a guidé le développement des principales méthodes de segmentation morphologique, en commençant par la ligne de partage des eaux (LPE). A l'origine, cette notion provient de la géographie, où la LPE d'un relief désigne l'ensemble des points à partir desquels une goutte d'eau peut atteindre deux vallées différentes. Une alternative intéressante est l'interprétation de la LPE en termes d'une *inondation* [1] : le relief est percé aux minima et immergé progressivement dans de l'eau. L'eau inonde uniformément les vallées, ou bassins versants des minima, et, aux points où deux lacs se rencontrent, un barrage est construit. Quand le relief est totalement immergé, l'union des barrages forme la LPE. Une motivation pour l'emploi de la LPE en segmentation d'images est la modélisation des contours de l'image  $u$  par les lignes de partage des eaux du module de son gradient  $g = \|\nabla u\|$ .

La formulation par inondation a permis la formalisation de la LPE dans le domaine continu en tant que squelette par zones d'influence de la distance image [24]. Dans notre notation, l'énergie associée s'écrit :

$$\widehat{W}_{min(g)}(g) = \inf_{m_i \in min(g)} \widehat{W}_i(g),$$

où  $\widehat{W}_i(g) = g(m_i) + W_i(g)$  et  $min(g)$  dénote l'ensemble des minima régionaux de  $g$ .

Ainsi,  $\Pi(\widehat{W}(g), min(g))$ , la partition d'énergie associée à

METHODE	MOSAÏQUE DES EXTREMA	LPE
Image	originelle $u$	$g = \ \nabla u\ $
Sources	extrema de $u$	minima de $g$
Energie	$V$	$\widehat{W}$
Régions d'Attraction	régions élémentaires	bassins versants
Ensemble Isoénergétique	composantes frontière	lignes de partage des eaux

TAB. 1 – Comparaison entre ME et LPE.

la LPE a l'interprétation suivante : l'ensemble isoénergétique  $I(\widehat{W}(g), \min(g))$  correspond aux lignes de partage des eaux et représente les contours des objets dans  $u$ . En plus,  $R(\widehat{W}(g), \min(g))$ , les régions des minima, sont interprétées comme les bassins versants, ou les lacs, du relief. Le Tableau 1 résume les différences entre la LPE et la ME.

Au cours de la dernière décennie, plusieurs définitions alternatives et implémentations pour la LPE ont vu le jour [27]. Certaines applications se centrent sur l'ensemble de contours fourni par la LPE, comme par exemple les *valuated contours* [25] ou les *watersnakes* [16]. Néanmoins, pour la segmentation basée sur des régions, une approche très populaire consiste à utiliser la partition induite par la LPE [31, 23]. Cette technique se fonde sur la définition d'une image simplifiée, que nous appellerons la *mosaïque LPE*, où chaque lac est à l'origine d'une région. Sa construction dépend des mêmes facteurs que la mosaïque des extrema : la connexité discrète, le niveau de gris des régions et la règle d'assignation pour l'ensemble isoénergétique. Cependant, la différence fondamentale est que la ME est définie à partir de l'image originelle, alors que la mosaïque LPE est construite sur le module du gradient. Ceci introduit un paramètre supplémentaire pour la mosaïque LPE car différentes approximations discrètes du gradient produisent des résultats différents. En outre, les opérateurs de gradient lissent en général l'image pour bien poser la dérivation ; ce lissage implique une perte de contenu dans l'image et, en conséquence, la mosaïque LPE souffre d'une résolution limitée dans certains cas. Ces problèmes ne peuvent pas être négligés dans des domaines où la précision des caractéristiques extraites joue un rôle essentiel, comme par exemple dans l'analyse d'images médicales.

La Figure 6 compare les méthodes sur un détail de l'image test *cameraman*. La première ligne montre l'image originelle et la mosaïque des extrema. Dans la deuxième ligne on trouve, à gauche, la mosaïque LPE, construite sur le gradient morphologique et, à droite, la mosaïque associée à la partition  $R(W(u), \text{ext}(u))$ . Dans tous les cas, la 8-connexité a été employée, la représentation des atomes est la valeur à la source et les points de l'ensemble isoénergétique ont été assignés à la première source qui les atteint. On peut observer que, à cause du grand nombre et de la distribution spatiale des sources, les méthodes préservent les attributs principaux de la scène, comme la silhouette du photographe. Cependant, la ME rehausse des détails im-

portants pour la perception tels que la bouche ou les parties intérieures de l'appareil qui sont effacés avec la LPE. La perte d'information est due à l'absence de minima à l'intérieur de ces régions et, même si le résultat peut être amélioré en changeant la connexité ou le type de gradient, le problème est intrinsèque à la méthode. D'autre part, la mosaïque montrée en bas à droite illustre les effets décrits pour la Fig. 2 sur une image réelle et souligne les avantages dans le choix de  $V$  comme énergie au lieu de  $W$ .



FIG. 6 – En haut : image originelle et mosaïque des extrema. En bas : autres mosaïques (voir le texte).

## 5 Segmentation Hiérarchique

La mosaïque des extrema fournit une décomposition de l'image en régions élémentaires. Cependant, en général, le niveau d'analyse où cette simplification de l'image agit est trop local pour faire correspondre les atomes aux objets physiques présents dans la scène. Dans cette section, la ME est employée comme le point de départ pour la création de familles de partitions emboîtées.

### 5.1 Fusion de Régions

L'idée de fusionner progressivement les régions d'une partition initiale a été exploitée depuis longtemps pour aborder les problèmes de segmentation d'images [3, 14, 5, 11]. Dans ces approches, les objets sont généralement définis

au moyen d'une notion d'homogénéité désirée, laquelle est exprimée comme une mesure de ressemblance entre les régions.

Dans cette partie, un processus général de fusion de régions basé sur un graphe a été employé, comme dans [13] et [9]. Un tel algorithme agit sur un Graphe d'Adjacence de Régions (RAG) où les noeuds correspondent aux régions, les liens traduisent la relation de voisinage entre régions et sont pondérés par une mesure de similarité. La similarité  $\mathcal{S}$  est une fonction définie pour chaque couple de régions voisines et dont les valeurs appartiennent à un ensemble totalement ordonné  $\Lambda$ , nommé ensemble des *indices* ou *échelles*. La partition initiale, la similarité et le modèle des régions déterminent un ordre de fusion ; la stratégie consiste à enlever les liens du RAG pour des valeurs croissantes de la similarité et fusionner les régions correspondantes.

Pour chaque seuil  $\lambda \in \Lambda$ , l'ensemble des noeuds restants dans le RAG représente une partition du domaine  $P_\lambda$ . Lorsque l'ordre de fusion est connu a priori, la construction de  $P_\lambda$  équivaut à trouver les régions connexes maximales dans le RAG initial pour lesquelles la similarité entre noeuds adjacents est inférieure à  $\lambda$ . La complexité de l'algorithme dans ce cas dépend linéairement du nombre de noeuds initial.

Pour des applications particulières, une mesure de qualité pour les partitions tel que le PSNR peut être définie comme critère de terminaison du processus. Cependant, la continuation de la fusion jusqu'à ce qu'il ne reste plus qu'une seule région engendre une famille de partitions emboîtées, ou hiérarchie,  $\{P_\lambda\}_{\lambda \in \Lambda}$  où chaque région de  $P_\mu$  est une union disjointe de régions de  $P_\lambda$ , pour  $\mu \geq \lambda$ .

Un exemple classique de segmentation hiérarchique reprend l'inondation du gradient : le relief est immergé à partir des minima mais, au lieu de construire un barrage, les lacs fusionnent. Les niveaux croissants d'eau produisent des partitions de plus en plus grossières et la hiérarchie résultante est connue sous le nom de *dynamique* [10]. En termes de fusion de régions, la partition initiale est la mosaïque LPE et la similarité est définie comme l'altitude du col entre lacs adjacents, *c.a.d.* la valeur minimale du gradient sur le bord commun des régions. Cet exemple sera appelé la *fusion LPE*.

Dans le but de lier étroitement le processus de fusion à l'information fournie par la mosaïque des extrema et pour rendre indépendant l'ordre de fusion du modèle des régions, la similarité a été mesurée sur la partition initiale. Un exemple typique de similarité est la différence de niveau moyen sur le bord des régions, noté  $\mathcal{S}_a$ . Cette similarité nécessite une actualisation de l'ordre après chaque fusion ; cependant, la prise en compte de l'information sur toute la frontière contribue à la robustesse du résultat.

La Figure 7 illustre l'approche sur la ME de l'image *cameraman*. La segmentation montrée correspond à un seuil de  $\lambda = 54$  dans la similarité  $\mathcal{S}_a$ . On peut observer que les régions homogènes sont bien détectées. Néanmoins, étant donné que cette similarité exprime une notion de contraste,

les petites régions contrastées du fond sont aussi présentes dans la partition.



FIG. 7 – Mosaïque des extrema et segmentation ( $\lambda = 54$ ).

## 5.2 Energie de Fusion

Une stratégie de fusion est un outil puissant pour la segmentation car, souvent, les objets de la scène peuvent s'identifier avec une région à une certaine échelle du processus. Cependant, le fait de définir les régions en se basant uniquement sur la similarité peut s'avérer restrictif. La Figure 8 montre un moment critique dans le processus de fusion de la ME de la Fig. 5 avec similarité  $\mathcal{S}_a$  : à l'échelle  $\lambda = 129$  les ailes du papillon fusionnent avec le fond. Puisque le contraste entre les motifs clairs et les ailes est supérieur au contraste entre les ailes et le fond, aucune région définie par le processus ne correspond au papillon complet (ailes et motifs).



FIG. 8 – Partitions pour les seuils  $\lambda = 128$  et  $\lambda = 129$ .

L'approche proposée dans cette partie consiste à utiliser la stratégie de fusion pour définir une énergie :

**Définition 3.** Soit  $\mathcal{S}$  une mesure de similarité entre régions et  $x_0 \in \Omega$  une source. L'énergie de fusion induite par  $\mathcal{S}$  par rapport à  $x_0$ , évaluée en un point  $x$ , est définie par

$$U_0(x) = U_0(\mathcal{S})(x) = \inf_{\lambda \in \Lambda} \{R \in P_\lambda \mid x_0, x \in R\}.$$

De même que pour les actions minimales, cette énergie détermine des régions d'attraction pour un ensemble de sources. Dans la Fig. 9, la méthode est appliquée à l'exemple de la Fig. 5. Au centre, on peut observer  $U_0(\mathcal{S}_a)$ , l'énergie de fusion induite par  $\mathcal{S}_a$  pour la source  $x_0$ . À droite, on voit les régions d'attraction lorsque la source  $x_1$  est ajoutée.



FIG. 9 – Sources, énergie de fusion pour  $x_0$  et régions d’attraction.

L’utilisation d’une énergie de fusion pour un niveau global d’analyse de l’image est motivé par le fait que la distance correspondante satisfait l’inégalité ultramétrique :

$$\hat{d}(\hat{x}, \hat{y}) \leq \max(\hat{d}(\hat{x}, \hat{z}), \hat{d}(\hat{z}, \hat{y})), \forall \hat{x}, \hat{y}, \hat{z} \in \hat{\Omega}.$$

En conséquence, toute autre deuxième source choisie à l’intérieur des ailes de la Fig. 8 produirait les mêmes régions d’attraction que dans la Fig. 9.

Ainsi, le problème se ramène à la sélection des sources, qui peut être vue comme le moment où l’information sémantique est introduite dans le processus de segmentation. Cette idée est employée dans la méthode des *marqueurs* [2] pour supprimer la sursegmentation produite par la LPE. Exprimée dans notre terminologie, cette méthode correspond à utiliser les marqueurs comme sources pour l’énergie associée à la fusion LPE.

Une approche similaire a aussi été adoptée dans [7]; comme nous, ces auteurs utilisent une méthode de filtrage suivie par une fusion de régions. Cependant, le filtrage est réalisé avec une série d’opérateurs connexes classiques, tels que les ouvertures et fermetures d’aire. Les sources, appelées dans ce contexte *flat zone masks*, sont choisies parmi les extrema de l’image filtrée qui satisfont certains critères, par exemple sur leur taille ou leur contraste. Donc, la segmentation finale avec cette méthode dépend du choix de l’opérateur de filtrage ainsi que de ses paramètres, qui doivent être déterminés pour chaque image.

### 5.3 Classement des Sources

La méthode proposée pour la sélection des sources se fonde sur la création d’un ordre des noeuds du RAG. Dans ce but, les attributs internes des régions sont utilisés en complément de l’information de voisinage fournie par la similarité. Par attribut, nous entendons une fonction  $\mathcal{A}$  définie pour les noeuds du RAG et à valeurs dans  $\mathbb{R}^+$ . La seule hypothèse supplémentaire pour  $\mathcal{A}$  est la croissance par rapport à l’ordre d’inclusion.

L’exemple le plus simple d’un attribut croissant est la taille de la région. En plus, le produit entre un attribut et la similarité fournit un nouvel attribut qui prend en considération aussi bien l’information interne que celle de voisinage.

Le classement est obtenu par un processus de fusion avec similarité  $\mathcal{S}$  de la façon décrite à continuation. Initialement, tous les noeuds du RAG ont des labels différents. A chaque pas de fusion, si les régions  $R_1$  et  $R_2$  fusionnent et  $\mathcal{A}(R_1) < \mathcal{A}(R_2)$ , alors on assigne à leur union le label

de  $R_2$ . D’autre part,  $\mathcal{A}(R_1)$  est nommé la *valeur de fusion* de  $R_1$ .

Puisque  $\mathcal{A}$  est croissant, à la fin du processus, l’ordre associé aux valeurs de fusion produit une hiérarchie de partitions, où l’indice s’identifie à l’attribut. Un niveau  $P_\lambda$  de la hiérarchie correspond aux régions d’attraction des régions initiales qui possèdent un indice supérieur à  $\lambda$  selon l’énergie de fusion associée à  $\mathcal{S}$ .

Cette approche peut être vue comme une généralisation des hiérarchies des minima du gradient [31, 23]. Dans cette méthode, lors de l’inondation LPE, lorsque deux lacs se rencontrent, celui qui possède l’aire ou le volume le plus grand absorbe l’autre. Les valeurs de fusion sont appelées dans ce contexte *valeurs d’extinction* et fournissent une hiérarchie des minima.

La Figure 10 montre deux exemples où l’attribut est défini comme le produit entre l’aire de la région et  $S_a$ . On peut observer que, malgré le faible nombre de régions et la simplicité de l’attribut, les caractéristiques principales sont récupérées dans les segmentations.



FIG. 10 – Image originelle et segmentations avec 30 (haut) et 35 (bas) régions.

## 6 Conclusion

Nous avons présenté une méthode pour la segmentation d’images sans connaissance préalable de leur contenu. L’approche définit la segmentation à partir des régions d’attraction d’un ensemble de sources selon une énergie. L’étude de deux types d’énergies et ensembles de sources a permis de traiter l’image à des niveaux d’analyse différents. Dans la première partie, une décomposition de l’image en régions élémentaires a été obtenue en utilisant la variation de chemin comme énergie et les extrema de l’image comme sources. Dans la deuxième partie, une stratégie de fusion de régions a été employée pour définir des énergies de fusion. Un classement des composantes de

l'image presegmentée a permis la création de familles de partitions emboîtées, d'où la segmentation finale a pu être choisie.

## Remerciements

Les auteurs voudraient remercier les relecteurs, dont les commentaires ont contribué à améliorer ce travail.

## Références

- [1] S. Beucher. *Segmentation d'images et morphologie mathématique*. PhD thesis, Ecole des Mines de Paris, 1990.
- [2] S. Beucher et F. Meyer. The morphological approach to segmentation : The watershed transformation. *Mathematical Morphology in Image Processing*, pages 433–481, 1992.
- [3] C.R. Brice et C.L. Fenema. Scene analysis using regions. *Artificial Intelligence*, 1 :205–226, 1970.
- [4] L. D. Cohen. Multiple contour finding and perceptual grouping using minimal paths. *Journal of Mathematical Imaging and Vision*, 14(3) :225–236, 2001.
- [5] L. D. Cohen, L. Vinet, P. Sander, et A. Gagalowicz. Hierarchical region based stereo matching. In *Proc. 6th Scandinavian Conference on Image Analysis*, 1989.
- [6] L.D. Cohen et R. Kimmel. Global minimum for active contour models : A minimal path approach. *International Journal of Computer Vision*, 24(1) :57–78, August 1997.
- [7] J. Crespo, R.W. Schafer, J. Serra, C. Gratin, et F. Meyer. The flat zone approach : A general low-level region merging segmentation method. *Signal Processing*, 62 :37–60, 1997.
- [8] E.W. Dijkstra. A note on two problems in connection with graphs. *Numerische Mathematic*, 1 :269–271, 1959.
- [9] L. Garrido, P. Salembier, et D. Garcia. Extensive operators in partition lattices for image sequence analysis. *Signal Processing*, 66(2) :157–180, April 1998. Special Issue on Video Sequence Segmentation.
- [10] M. Grimaud. New measure of contrast : Dynamics. In *Image Algebra and Morphological Processing III*, SPIE, San Diego, CA., 1992.
- [11] R. Haralick et L. Shapiro. *Computer and Robot Vision*, volume I. Adison Wesley, 1992.
- [12] E. Hewitt et K. Stromberg. *Real and Abstract Analysis*. Springer Verlag, 1969.
- [13] R. Horaud et O. Monga. *Vision par Ordinateur*. Hermes, Paris, 1995.
- [14] S.L. Horowitz et T. Pavlidis. Picture segmentation by a directed split-and-merge procedure. In *Proceedings of the Second International Joint Conference on Pattern Recognition*, pages 424–433, 1974.
- [15] C. Jordan. Sur la série de fourier. *C. R. Acad. Sci. Paris Sér. I Math.*, 92(5) :228–230, 1881.
- [16] J.Park et J. M. Keller. Snakes on the watershed. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10) :1201–1205, October 2001.
- [17] R. Kimmel, N. Kiryati, et A. M. Bruckstein. Distance maps and weighted distance transforms. *Journal of Mathematical Imaging and Vision*, 6 :223–233, May 1996. Special Issue on Topology and Geometry in Computer Vision.
- [18] A.S. Kronrod. On functions of two variables. *Uspehi Mathematical Sciences*, 5(35), 1950.
- [19] R. Kruse et A. Ryba. *Data structures and program design in C++*. Prentice Hall, New York, 1999.
- [20] D. Marr. *Vision*. Freeman, San Francisco, 1982.
- [21] D. Martin, C. Fowlkes, D. Tal, et J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. ICCV'01*, volume II, pages 416–423, Vancouver, Canada, 2001.
- [22] F. Meyer. Topographic distances and watershed lines. *Signal Processing*, 38 :113–125, 1994.
- [23] F. Meyer. Hierarchies of partitions and morphological segmentation. In Michael Kerckhove, editor, *Scale Space and Morphology in Computer Vision*, pages 161–182, Vancouver, Canada, 2001.
- [24] L. Najman. *Morphologie Mathématique : de la Segmentation d'Images à l'Analyse Multivoque*. PhD thesis, Université Paris Dauphine, 1994.
- [25] L. Najman et M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12) :1163–1173, 1996.
- [26] I.P. Natanson. *Theory of Functions of a Real Variable*. Frederick Ungar Publishing, New York, 1964.
- [27] J. Roerdink et A. Meijster. The watershed transform : Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, (41) :187–228, 2000.
- [28] M. Schmitt et J. Mattioli. *Morphologie Mathématique*. Masson, 1994.
- [29] J. Serra et P. Salembier. Connected operators and pyramids. In SPIE, editor, *Image Algebra and Mathematical Morphology*, volume 2030, pages 65–76, San Diego CA., July 1993.
- [30] J.A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge, UK, 2 edition, 1999.
- [31] C. Vachier. *Extraction de Caractéristiques, Segmentation d'Image et Morphologie Mathématique*. PhD thesis, Ecole des Mines de Paris, 1995.