

Segmentation d'images par utilisation de pyramides à bases locales

Image Segmentation Using Local-Base Pyramids

par H. Konik (*), B. Laget (*), M. Calonnier (**)

(*) Laboratoire Traitement du Signal et Instrumentation (TSI), URA 842,
Université Jean Monnet, 23 rue du Dr Michelon
42023 Saint-Etienne Cedex 2

(**) Institut Textile de France, Lyon,
Avenue Guy-de-Collongue, BP 60, 69132 Ecully

Résumé

Dans cet article, nous abordons un problème de segmentation d'images, caractérisées par des irrégularités topographiques qui n'émergent que localement d'un fond bruité.

Dans ce cas, la méthode de segmentation ne peut être que localisée et adaptative. Pour cela, la description d'un objet se fait par l'utilisation d'une pyramide à base locale – appelée « pyramide locale » – simulant la focalisation de l'œil sur les zones sensibles de l'image. Ainsi, chaque objet est analysé indépendamment et en fonction de la texture intrinsèque de l'image.

Mots clés : Segmentation d'image, Traitement multi-résolution, Pyramides, Méthodes locales, Texture

Abstract

This paper describes an approach to image segmentation. The methods intended for images where local topographic irregularities contrast with their noisy backgrounds.

Since the method has to be localized and adaptive, the description of each object uses a local-base pyramid, called "local pyramid". It simulates the human vision in its attention focusing. Using such a tool, each object is described independently and with respect to the image texture.

Key words : Image segmentation, Multiresolution processing, Pyramids, Local methods, Texture

1. Introduction

Extraire d'une scène quelconque les objets caractéristiques est une opération que le système visuel humain est apte à réaliser en très peu de temps. Cette étape, la **segmentation** dans le domaine de l'analyse et du traitement d'image, doit donc essayer de simuler au mieux la vision humaine.

Le but spécifique de notre recherche est d'extraire d'une image fortement bruitée des irrégularités topographiques homogènes émergentes localement par rapport à un fond. Or nous disposons pour ces images à segmenter de leur classification en micro/macro perturbation [1]. Cette étape est intégrée dans un processus de co-tation visuelle assistée par ordinateur des différences d'état de surface [2]. En fait, une image est micro-perturbée lorsque la sensation de petits grains saillants émerge. Par contre, si des sensations d'amas plus marqués dominent, l'image est macro-perturbée. Cette étape utilise notamment la fonction d'autocorrélation mais il n'est pas dans nos considérations de la développer plus. Nous savons cependant que nous possédons des mesures significatives

de la perturbation de l'image, comme par exemple une approximation de la taille a priori des amas caractéristiques de l'image [1].

Ainsi, cette classification en micro/macro perturbation nous assure une connaissance *a priori* des notions d'homogénéité et d'émergences locales caractéristiques de l'image. On a donc la possibilité de mettre en œuvre une **segmentation adaptative**, dont le rôle est d'adapter la segmentation au type de l'image traitée.

Etant données les contraintes imposées par nos applications, on s'oriente plutôt, parmi les nombreuses méthodes de segmentation existantes, vers l'approche **multi-résolution**. Elle permet en effet de mettre en œuvre le principe de **focus d'attention** : « pour chaque entité de l'image, il existe une résolution donnée dans laquelle la recherche (*i.e.* détection) et l'extraction (*i.e.* localisation) sont facilitées » [3]. En fait, cette démarche se justifie en admettant que des objets se reconnaissent plus facilement dans une image de faible résolution. Des détails peu significatifs dans une image de haute résolution disparaissent pour une résolution réduite, facilitant la détection. Cependant, une description détaillée de

l'objet nécessite la prise en compte de toutes les informations présentes à de hautes résolutions.

Plus précisément notre outil sera le *modèle pyramidal*, introduit par Tanimoto et Pavlidis [4] pour l'analyse d'image. Une pyramide peut se définir comme la suite des résolutions successives d'une image, chacune permettant de mettre plus ou moins en évidence des phénomènes locaux caractéristiques. Cette approche de la pyramide n'est pas unique puisque ses utilisations restent pluridisciplinaires comme en témoignent plusieurs ouvrages et synthèses [3][5][6]. Son avantage dans ce cas est d'autant plus probant que les algorithmes pyramidaux ont démontré leur robustesse et sont peu influencés par différentes perturbations structurales [7].

Tout d'abord, nous proposerons une utilisation originale du filtrage par le *Laplacien d'une gaussienne*, développé par Marr [8] et qui calque le système visuel humain. Cet outil sera insuffisant mais engendrera une image de contraintes, vue comme une segmentation grossière et permettant d'exclure dès à présent les émergences topographiques non caractéristiques. Ensuite, après des rappels sur la pyramide classique utilisée dans la segmentation d'images de niveaux de gris, nous mettrons à jour certains problèmes dans les méthodes existantes qui nous ont conduit à développer une pyramide à base locale, appelée *pyramide locale* (§ 2). Cette méthode couple la focalisation de l'œil sur les parties émergentes avec l'approche pyramidale. Ses avantages sont notamment de ne supposer aucune forme *a priori* sur les objets à isoler, d'être suffisamment flexible pour gérer leur diversité et de permettre la mise en œuvre d'une segmentation adaptative. A chaque point prédominant de l'image est associée une pyramide. Le ou les niveaux de résolution les plus représentatifs de l'objet ainsi marqué sont détectés afin de calculer un intervalle d'existence. Au cours d'une étape d'agrégation, les problèmes de connexité et de recouvrement entre les différents objets sont gérés (§ 3). Finalement, des corrections à base de *morphologie mathématique* permettront d'affiner la segmentation. Chaque segment sera repris en fonction de sa valeur moyenne intérieure et de celle de sa couche externe. Précisons que cette étape peut se généraliser comme « post-analyse » de toute segmentation (§ 4). Pour finir, la méthode sera jugée sur une image test de nos applications choisie pour sa complexité, au sens où elle possède des objets à extraire de taille et de contraste très différents. Son évolution au cours de la méthode permettra d'imager les différents points critiques de celle-ci (§ 5).

2. Choix et définition de la pyramide locale

Commençons par fixer quelques notations :

- une image I sera un ensemble de pixels $p = (x, y)$ tel que $I = [0, 2^n - 1] \times [0, 2^n - 1]$.
- à chaque pixel sera associé un *niveau de gris* $f(p) = f(x, y)$ tel que $\forall p \in I, 0 \leq f(p) < M$.
- M est le nombre de niveaux de gris dans l'image.
- le *voisinage* de taille N d'un pixel $p = (x_p, y_p)$ sera : $V_{N(p)} = \{q = (x, y) \in I; |x - x_p| \leq N \text{ et } |y - y_p| \leq N\}$.

Dans cette partie, nous présenterons le filtrage par le Laplacien d'une gaussienne. Ses résultats sont insuffisants pour nos besoins. Cependant, ils restent utiles pour la méthode à base de pyramides locales. Celles-ci sont introduites pour analyser chaque objet *individuellement*, la pyramide classique étant de structure trop rigide.

2.1. FILTRAGE PAR LE LAPLACIEN D'UNE GAUSSIENNE

Le but est de pré-isoler les parties fortement émergentes de l'image. Pour cela, calquant le système visuel humain, le filtrage par le Laplacien d'une gaussienne développé par Marr [8] est utilisé. Présentons rapidement cet outil dont tous les fondements théoriques peuvent être trouvés dans la littérature.

Le filtrage par la gaussienne de variance σ^2

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

lisse fortement le bruit de l'image, supprimant toute structure de taille inférieure à σ . Ensuite, l'opérateur Laplacien

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

permet la détection des contours comme l'illustre la figure 1 sur une ligne.

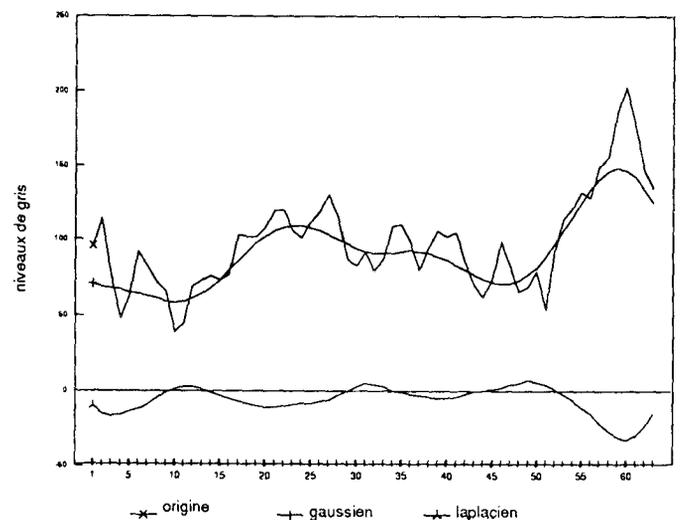


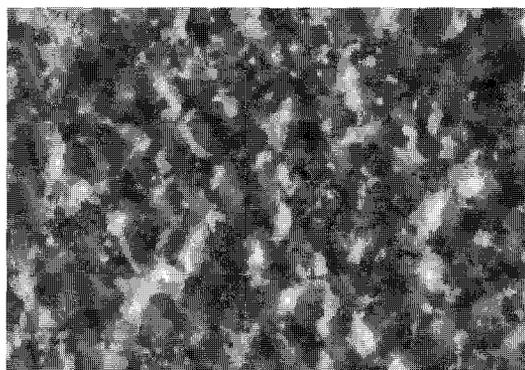
Figure 1. - Effet du filtrage par le Laplacien d'une gaussienne.

En définitive, profitant de la topographie caractéristique de nos types d'images, un seuillage à zéro de l'image obtenue permet la création d'une image de contraintes binaire :

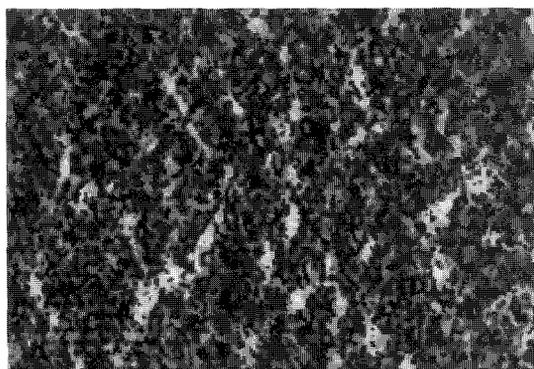
$$I_c = \{(x, y) \in I; \nabla^2(G_\sigma(x, y) * I(x, y)) < 0\}.$$

Cette image de contraintes permet alors de restreindre les régions à analyser, en excluant également les irrégularités non caractéristiques. Or ce traitement reste relatif comme l'illustre la figure 2

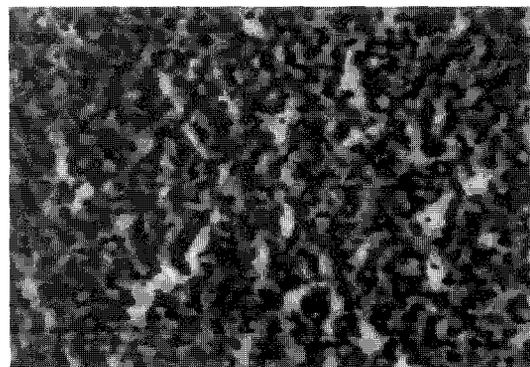
et il n'existe pas d'écart-type σ de la gaussienne significatif pour une image donnée. Cependant, la notion d'amas est suffisamment sensible (visuellement et par calcul [1]) dans l'image test (Fig. 2) et le paramètre σ choisi aura tendance à être dans ce cas proche de 3.



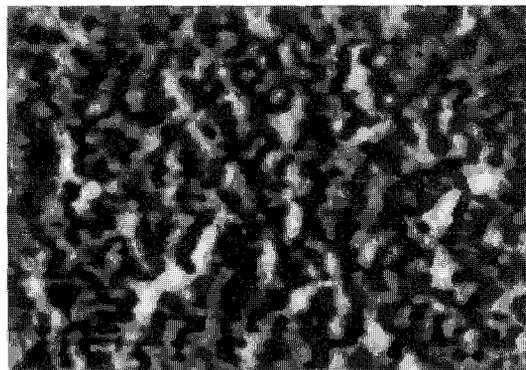
(a) image test.



(b) images de contraintes avec $\sigma = 1$.



(c) images de contraintes avec $\sigma = 2$.



(d) images de contraintes avec $\sigma = 3$.

Figure 2. – Images (a) à (d) : Les pixels non retenus après seuillage apparaissent en noir.

2.2. AVANTAGES DE LA PYRAMIDE LOCALE

2.2.1. Pyramide classique

Une *pyramide* P est une suite d'images I_h , chacune correspondant à une version de l'image initiale I_0 pour une résolution réduite. Chaque niveau h est construit par une convolution du niveau inférieur avec un noyau de taille k suivi d'un sous-échantillonnage. Généralement, la taille est divisée par 2 à chaque niveau si bien que :

$$I_h = [0, 2^{n-h} - 1] \times [0, 2^{n-h} - 1], \forall h \in [0, T - 1].$$

I_0 et T sont respectivement la base et la hauteur de la pyramide.

2.2.2. Masque de pondération

Dorénavant, $f_h(x, y)$ désigne le niveau de gris de l'élément (x, y) de l'image I_h constituant la pyramide. f_0 représente les niveaux de gris de l'image initiale.

Chaque élément du niveau h de la pyramide est construit par l'intervention de k^2 éléments du niveau inférieur, ses *fil*s. En fait,

$$f_h(x, y) = \sum_{i=1}^{i=k} \sum_{j=1}^{j=k} f_{h-1}(2x + i - z, 2y + j - z) \cdot w(i, j)$$

où z est une constante valant $\lfloor (k + 1)/2 \rfloor$ ($\lfloor \cdot \rfloor$ représente la fonction partie entière). Chaque élément du niveau h engendre des *pères* au niveau supérieur, en intervenant dans leur construction.

Cette construction est ainsi définie par son *masque de pondération* $W = [w(i, j)], i \in [1, k], j \in [1, k]$, dont les différentes formes ont été étudiées par Burt [9]. C'est généralement un filtre passe-bas, adoucissant considérablement les contours afin d'éviter les problèmes de repliement du spectre des fréquences. Ainsi, priorité sera donc donnée aux caractéristiques non géométriques, intrinsèques d'une région, plutôt qu'à la recherche des discontinuités locales.

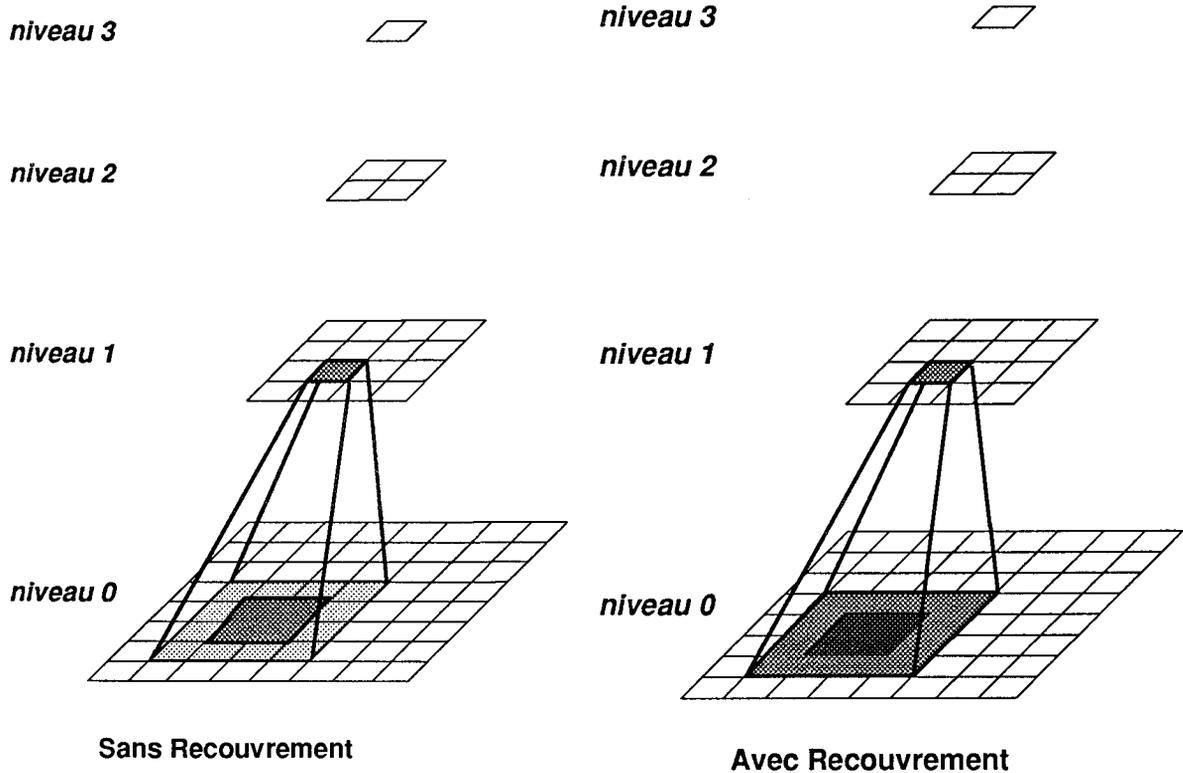


Figure 3. – Structures pyramidales quaternaires de taille 4.

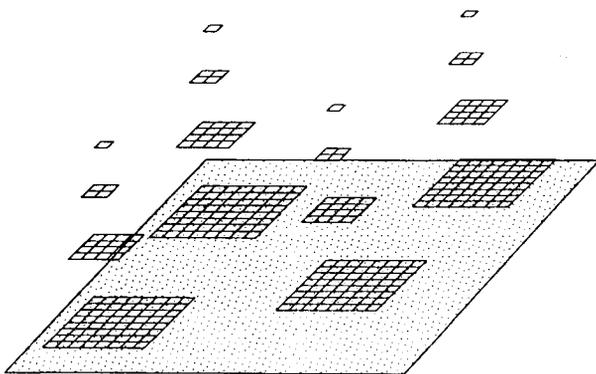


Figure 4. – Pyramides locales dédiées aux points sensibles.

Deux philosophies semblent en fait émerger : les structures *avec* ou *sans recouvrement*. La structure pyramidale sans recouvrement correspond au cas où aucun des k^2 fils d'un élément n'intervient pour la construction d'un autre. Par contre, le recouvrement permet à certains éléments d'intervenir dans la génération de plusieurs pères. Le bruit de l'image est dans ce cas plus lissé d'un niveau à l'autre. L'exemple des structures pyramidales quaternaires sans recouvrement ($k = 2$) et avec recouvrement ($k = 4$) est donné (Fig. 3).

2.2.3. Introduction de la pyramide locale

Les inconvénients de ces constructions dans le contexte de l'extraction d'objets sont résumés par Jolion et Montanvert [10]. De plus, une pyramide classique est globalement associée à l'image. C'est-à-dire que cette même structure reste figée et doit pourtant répondre à n'importe quelle image. Cette construction globale est trop rigide pour s'adapter à des objets extrêmement diversifiés à l'intérieur même de l'image. C'est le cas dans notre problème dont l'optique est l'*extraction d'objets compacts qui émergent d'un fond homogène localement*. Le processus de segmentation doit en effet extraire des «amas» plus ou moins compacts, plus ou moins fibreux, plus ou moins émergents similaires à ceux de l'image test (Fig. 2).

Le but est alors de simuler au maximum le principe de focus d'attention par l'introduction de la notion de *pyramide locale*. La base de chacune d'elle n'est plus dans ce cas l'image entière mais seulement une zone sensible de celle-ci (Fig. 4). Une analyse multirésolution locale est alors faite dans chaque pyramide. Elle repose notamment sur l'hypothèse de l'existence d'un seul objet. Les bases locales sont liées au taux de micro/macro perturbation et plus précisément à la taille *a priori* des objets (partie 1). En effet, elles doivent être suffisantes pour englober les plus grands objets. De toute façon, les résultats pour les plus petits objets resteront les mêmes que ceux que l'on obtiendrait avec des pyramides de base plus restreinte (paragraphe 3.2.2).

Notre approche consiste à construire autant de pyramides que d'objets, chacune localisée dans le support image. Signalons une

autre démarche qui consiste à décomposer chaque forme en sous-ensembles à l'intérieur d'une représentation multi-résolution. Mais dans ce cas, une pyramide entière est cependant associée à l'image [15][16][17].

3. Segmentation adaptative basée sur les pyramides locales

La première étape de la méthode de segmentation consiste à extraire de l'image les points sensibles. Ensuite, la segmentation peut se décomposer en deux parties distinctes. Tout d'abord, une pyramide est dédiée à chacun de ces points caractéristiques (Fig. 4). Par une analyse multirésolution locale et directionnelle, chaque objet est alors personnalisé indépendamment des autres. Finalement, une croissance de région contrainte et hiérarchisée règle les conflits de voisinage éventuels entre objets.

3.1. EXTRACTION DES POINTS SENSIBLES

Pour extraire les points les plus caractéristiques de l'image ou **germes**, on utilise la notion de maximum local sur un voisinage de taille quelconque.

Soit \mathcal{G} l'ensemble des germes, on a par définition :

$$g_i = (x_i, y_i) \in \mathcal{G} \iff \{f(g_i) \geq f(p) \forall p \in V_N(g_i)\}.$$

Signalons que cette définition est adaptée à nos applications, où il faut extraire des émergences topographiques locales dans des images texturées. Cependant, la recherche des zones «noires» d'une image se ferait avec la même démarche, après une simple inversion d'image. La taille N du voisinage est obtenue grâce au taux de micro/macro perturbation [1]. Elle est d'autant plus élevée que l'image est macro-perturbée. Il serait également possible d'envisager une variation de N pour une image donnée. Dans cette optique, les résultats devraient être comparés au cours d'une étape d'analyse.

Pour supprimer les germes accidentels, on utilise l'image de contraintes obtenue au paragraphe 2.1. Ainsi, l'ensemble des germes devient :

$$\mathcal{G} = \mathcal{G} \cap \mathbf{I}_c.$$

3.2. PERSONNALISATION DE CHAQUE OBJET

L'objet est de caractériser au mieux chaque objet marqué par un germe g_i . On lui associe une pyramide P_i . Le processus étant le même pour tous les germes et indépendant des autres jusqu'au paragraphe 3.3, on se limitera jusque-là à un germe g quelconque et à sa pyramide locale associée P . Celle-ci est utilisée au cours d'une analyse multirésolution directionnelle pour générer une base englobante optimale au sens du contraste. De celle-ci est alors extrait un intervalle d'existence de l'objet en terme de niveaux de gris. Un second intervalle en terme de contraste L.I.P. (défini au paragraphe 3.2.4) peut également être utilisé en complément.

3.2.1. Construction des pyramides locales

La base \mathbf{I}_0 de chaque pyramide locale de hauteur T est la partie de l'image initiale «centrée» en g , de taille $[0, 2^{T+1} - 1] \times [0, 2^{T+1} - 1]$. Centrée au sens où le germe est positionné en $(2^T, 2^T)$ dans \mathbf{I}_0 . Rappelons que T , adapté à l'image, est d'autant plus grand que l'image est macro-perturbée [1]. Une pyramide est alors construite à partir de \mathbf{I}_0 suivant une structure quaternaire avec recouvrement (Fig. 3).

Avec cette structure, chaque élément possède 16 fils et 4 pères. Parmi les éléments de la pyramide, on s'intéresse **aux ascendants du germe** (Fig. 5). Ce sont ceux dont le germe est intervenu dans la construction directement (niveau 1) ou indirectement (niveaux supérieurs). On s'aperçoit qu'à chaque niveau, il en existe toujours deux sur une ligne de la pyramide, excepté à son sommet. On introduit ainsi par construction la notion de 4 **orientations** (Fig. 6) dans le parcours d'un niveau à l'autre.

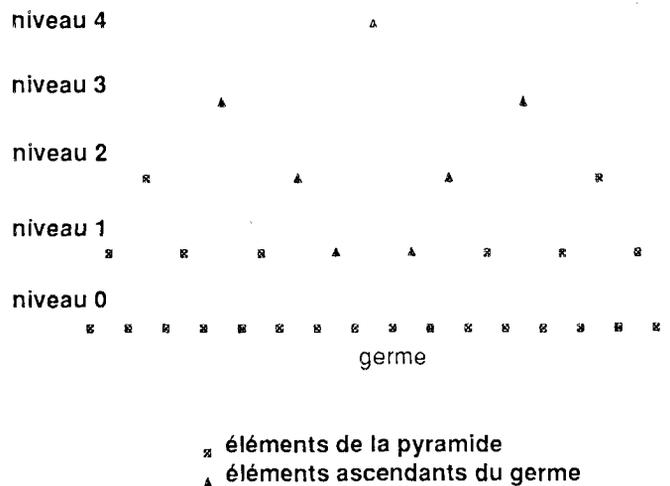


Figure 5. – Eléments ascendants du germe dans une pyramide locale sur une ligne.

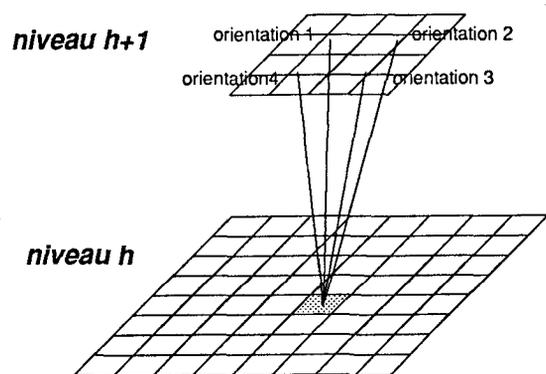


Figure 6. – Numérotation des 4 orientations dans la pyramide.

3.2.2. Base englobante de chaque objet

Le but est de restreindre la base I_0 à une base englobante optimale de l'objet. Pour cela, une analyse multirésolution directionnelle est utilisée. Il faut donc trouver maintenant un critère dont l'évolution dans la pyramide est caractéristique. Gross et Rosenfeld proposent une mesure d'intérêt [11] : le rapport σ_s^2/σ_F^2 . σ représente la variance soit du père F soit du fils S , variance calculée pour un élément sur sa descendance. Rosenfeld et Sher introduisent une notion de contraste en multirésolution [12], défini pour l'élément (m, n) au niveau h de la pyramide par :

$$C_h(m, n) = f_h(m, n) - \frac{1}{8} \sum_{\substack{k=m-1 \\ (k,l) \neq (m,n)}}^{k=m+1} \sum_{l=n-1}^{l=n+1} f_h(k, l).$$

Il représente l'émergence d'un élément par rapport à son voisinage. L'idée est qu'un élément aura un fort contraste absolu si chacun de ses voisins appartient plutôt au fond. Or l'utilisation faite de ces mesures est telle que les objets vont être caractérisés par des racines **uniques**. Une **racine** est par définition l'élément le plus représentatif d'un objet à l'intérieur de la pyramide. Elles sont de plus beaucoup trop sensibles au bruit et justifiées pour des objets de forme plutôt carrée. Elles ne peuvent donc pas répondre correctement à notre problème.

On introduit alors une analyse multirésolution du contraste [12] suivant les 4 orientations introduites au paragraphe précédent. En fait, un contraste C_h^α est calculé pour tous les ascendants du germe (Fig. 5) à chaque niveau h dans chaque orientation α . Signalons que celui-ci n'est plus obtenu sur un voisinage de 8 points [12] mais uniquement sur les **voisins critiques** (Fig. 7). On ne considère pas la valeur du contraste en elle-même mais son évolution à travers les différents niveaux de résolution.

On retient alors l'ascendant du germe au niveau h_α tel que :

$$\begin{cases} h_\alpha \in [1, T - 3] \\ C_{h_\alpha}^\alpha \geq C_h^\alpha & \forall h \in [1, h_\alpha - 1] \text{ pour } \alpha \in [1, 4]. \\ C_{h_\alpha+1}^\alpha < C_{h_\alpha}^\alpha \end{cases}$$

voisins critiques dans l'orientation 1

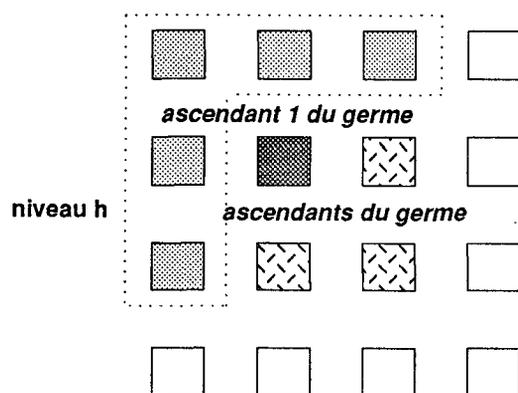


Figure 7. – Voisins critiques dans le cas particulier de l'orientation 1 à un niveau h .

Cette recherche se fait de façon croissante à partir du premier niveau. h_α correspond à la résolution optimale de l'objet dans l'orientation α . La base englobante B est en définitive la projection de chaque ascendant du germe au niveau h_α sur la base I_0 . Par définition, elle est constituée des feuilles terminales de l'arborescence (base) en prenant à chaque niveau l'ensemble des fils.

Cette procédure ne suppose la donnée d'aucune information *a priori* sur la forme des objets à isoler dans l'image. Néanmoins, l'outil pyramidal permet la mise en œuvre d'une mesure extérieure (partie 1). Cet outil est également extrêmement commode car la détection des objets est automatique et suivant la même construction pour chaque germe. De plus, l'information nécessaire dans l'image initiale est définitivement connue au début de la procédure. Toutefois, on pourrait analyser le contraste dans chaque orientation jusqu'à ce que celui-ci commence à décroître. Il faut choisir la base de chaque pyramide locale avec une taille toujours plutôt grande. Avec cette précaution, on ne risque pas de fausser l'analyse multirésolution par un nombre trop limité de niveaux dans la pyramide. En effet, si les bases pyramidales étaient trop restreintes, les bases englobantes risqueraient d'être insuffisantes et donc non significatives. Un exemple de base englobante sera donné dans la partie résultats et discussion (Fig. 13).

3.2.3. Intervalle d'existence de l'objet en niveau de gris

La moyenne m^{ng} et l'écart-type σ^{ng} en niveaux de gris, calculés sur B , sont choisis comme caractéristiques de la région. Son intervalle d'existence en **niveau de gris** est par définition :

$$I^{ng} = [m^{ng} - \sigma^{ng}, f(g)].$$

L'intervalle d'existence revient à conserver environ 80% des pixels à l'intérieur des bases englobantes. Un tableau des résultats sera fourni au paragraphe 5 (Tab. 2).

Ce pourcentage peut être interprété de deux manières :

- dans le cas de l'hypothèse gaussienne sur l'histogramme associé à B tout d'abord, ces bornes reviennent à conserver environ 84% des points.

- dans une optique plus géométrique ensuite, prenons l'exemple des bases englobantes optimales associées au disque et à l'ellipse (Fig. 8). Dans ce cas, la région à conserver occupe environ 79% de la surface de la base englobante. En effet, le disque occupe $\pi R^2/4R^2$ de la surface du carré englobant. Ce rapport reste bien sûr le même dans le cas de l'ellipse.

3.2.4 Intervalle d'existence complémentaire en contraste L.I.P (Logarithmic Image Processing)

L'intervalle I^{ng} suffit généralement pour la description de chaque objet. Cependant, il est nécessaire, pour certaines applications, d'assurer l'homogénéité intrinsèque de l'objet quant au voisinage en tenant compte des trop fortes variations en niveaux de gris. Ceci permet aussi d'augmenter la confiance donnée aux niveaux de gris d'une image fortement bruitée.

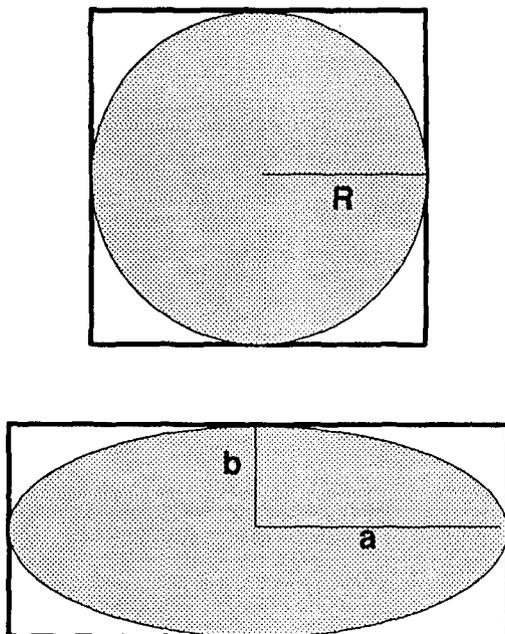


Figure 8. – Exemples de bases englobantes optimales.

Par définition, soit f_1 et f_2 deux niveaux de gris, on appelle contraste L.I.P. [13] la quantité :

$$CL(f_1, f_2) = \frac{|f_1 - f_2|}{1 - \frac{\min(f_1, f_2)}{M}}$$

où M est le nombre de niveaux de gris dans l'image.

Ce modèle est intéressant car il amplifie la dénivellation entre deux niveaux de gris proches d'autant plus que l'on se trouve en haut de l'échelle de luminance.

Par exemple, avec $M = 256$: $C_L(10, 12) = 2.08$, $C_L(128, 130) = 4$, $C_L(246, 248) = 51.2$.

En comparaison, le contraste linéaire classique $(f_1 - f_2)/(f_1 + f_2)$ donne respectivement 0.091, 0.008 et 0.004.

Par définition, en un point p , on pose :

$$C_{L.I.P.}(p) = \text{Max}_{q \in V_1(p)} C_L(f(p), f(q)).$$

On exécute alors les mêmes calculs que pour I^{ng} en terme de contraste L.I.P. sur B , d'où I^c défini par :

$$I^c = [m^c - \sigma^c, \text{Max}_{q \in B} C_{L.I.P.}(q)].$$

L'influence de ce paramètre est justifié sur un exemple réel (Fig. 9). Le contraste L.I.P. a tendance à exclure des pixels trop éloignés en terme de niveaux de gris. Il sera utilisé au cours de nos applications et donc intégré dans la suite de l'exposé de la méthode.

L'objet est en définitive caractérisé par deux intervalles d'existence. Pour l'instant, les traitements effectués pour un germe le

111	95	82	60	81
94	71	79	94	100
99	92	79	95	103
109	105	120	94	111
117	91	90	116	115

Moyenne : 96.12 Ecart-type : 14.92
Borne Inférieure : 81.20

55	33	28	44	38
31	16	23	44	38
38	43	59	39	33
33	37	59	41	46
41	44	46	40	33

Moyenne : 39.28 Ecart-type : 9.74
Borne Inférieure : 29.54

111	95	82	60	81
94	71	79	94	100
99	92	79	95	103
109	105	120	94	111
117	91	90	116	115

- En gris clair sont marqués les pixels conservés. - En gris foncé est marqué le pixel supprimé non pas à cause de son niveau de gris mais car son contraste n'est pas caractéristique. - En blanc apparaissent les pixels hors des intervalles d'existence.

Figure 9. – Effets des doubles intervalles d'existence.

sont indépendamment de ceux effectués pour les autres germes. L'approche locale est donc respectée.

3.3. CROISSANCE DE RÉGION CONTRAINTE ET HIÉRARCHISÉE

Au cours de cette étape, il sera fait appel à certains outils de **morphologie mathématique** (Serra [14]). Rappelons donc quelques définitions utilisées par la suite :

- dilatation \oplus : $A \oplus B = \{x; \exists a \in A, b \in B, x = a + b\}$
- soustraction ensembliste $/$: $A/B = \{x; x \in A, x \notin B\}$
- 8-connexité : x et y sont 8-connectés $\Leftrightarrow y \in V_1(x)$.

Une croissance de région sur l'image initiale va utiliser les intervalles d'existence de chaque germe. Chaque germe est alors de nouveau réindexé; on notera $I_i^{n,g}$ et I_i^c les intervalles d'existence du germe g_i . Ceux-ci sont triés par ordre de moyenne en niveaux de gris décroissante. Avec cette notion d'**ordre**, on donne de l'influence aux objets les plus émergents de l'image. La croissance de région est **contrainte** par le fait que les segments obtenus ne peuvent croître qu'à l'intérieur de l'image de contraintes I_c .

Un objet R_i est donc construit en agréant tous les pixels p 8-connectés entre eux à partir de g_i et vérifiant les conditions :

$$p \in I_c \text{ et } \{f(p) \in I_i^{n,g} \text{ et } C_{L.I.P}(p) \in I_i^c\}.$$

Cependant, au cours de ces agrégations, certains objets peuvent s'intersecter donnant lieu à des zones de conflit. Une fois R_i construit, les seuls cas litigieux pouvant se présenter dans la génération de R_j , tel que $m_i^{n,g} > m_j^{n,g}$ vu le tri effectué, sont présentés figure 10. La démarche adoptée est alors la suivante :

- Dans le premier cas ($g_j \in R_i$), on ne conserve aucun des pixels de R_j dans la segmentation. R_j est trop voisin de R_i et par la notion d'ordre, on privilégie l'objet le plus émergent.
- Dans le second cas ($g_j \notin R_i$), seule est conservée la partie 8-connexe de R_j contenant g_j telle que $R_i \cap R_j = \emptyset$. Ainsi, priorité est donc toujours laissée aux zones les plus émergentes de l'image.

En définitive, tous les objets R_i ne se retrouvent pas dans la segmentation et d'autres sont 8-connectés entre eux. La méthode reste donc flexible par rapport au choix des germes. Après labélisation, une première segmentation est alors obtenue avec autant ou moins de segments que de germes initiaux :

$$S = \bigcup_{i=1}^{nbs} S_i \text{ avec } nbs \leq \text{Card}(G) \text{ et } S_i \cap S_j = \emptyset \text{ pour } j \neq i.$$

4. Corrections à base de morphologie mathématique

Au cours de l'étape d'agrégation, les deux cas critiques (partie 3) génèrent néanmoins des zones de flou. La démarche suivie implique que certains objets S_i ne correspondent pas exactement

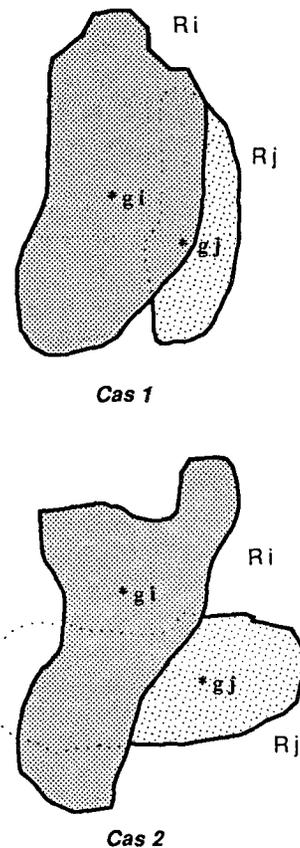


Figure 10. - Cas litigieux au cours de l'agrégation.

aux R_i générés par l'utilisation des pyramides locales. En effet, certains pixels qui n'ont pas été agrégés à un objet doivent peut-être le devenir (cas 1). D'un autre côté, plusieurs objets voisins possédant une frontière commune (cas 2) sont peut-être suffisamment distincts pour apparaître déconnectés. Pour gérer ces cas critiques, des corrections à base de morphologie mathématique ont été développées. Ces étapes sont **générales**. Elles sont en effet adaptées pour confirmer toute segmentation. Certains critères liés au type d'application peuvent être introduits sans problème à ce niveau.

L'optique choisie est de caractériser un objet S_i localement par sa moyenne en niveaux de gris m_i^g et celle de sa couche externe $(S_i \oplus kB)/S_i$, que l'on note m_i^e . L'élément B est la **boule unité** et k un scalaire positif. $(S_i \oplus kB)$ représente l'objet et ses premiers voisins, c'est-à-dire ceux qui sont à une distance inférieure ou égale à k . Par soustraction ensembliste avec S_i est alors obtenue la couche externe, soit les premiers voisins de chaque objet (Fig. 11). Notre choix est $k = 1$ car nos images sont telles que les objets à extraire peuvent être très proches. Pourtant, il faut les différencier, s'ils sont déconnectables en terme de niveaux de gris, pour avoir une cotation efficace.

Premièrement, une **décroissance de région** est effectuée. Elle a pour but d'extraire d'un objet les pixels trop éloignés de celui-ci en terme de niveaux de gris. Ainsi, cette analyse ne peut que diminuer le nombre de pixels retenus dans la segmentation d'où

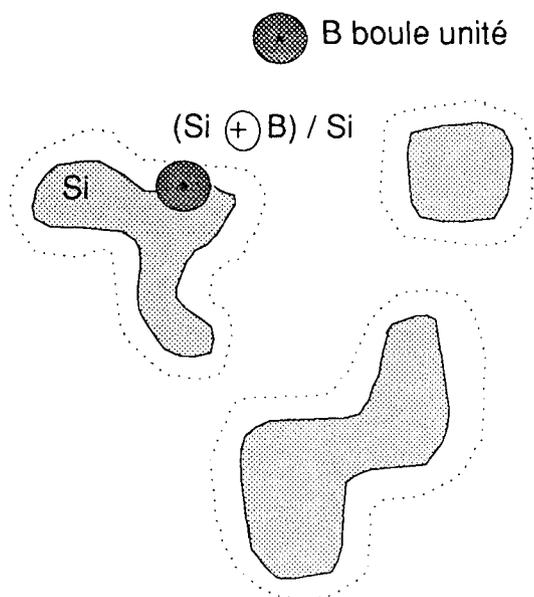


Figure 11. – Objets et leurs couches externes associées.

le terme «décroissance». L'appartenance d'un pixel p à S_i se juge par la proximité avec m_i^s :

pour $p \in S_i$, si $(|f(p) - m_i^c| < |f(p) - m_i^s|)$,

alors p est affecté au fond.

De la même façon, une **croissance de région** va augmenter si nécessaire le nombre de pixels retenus pour chaque objet, d'où le terme «croissance». Elle permet de s'affranchir des problèmes éventuels de croissance insuffisante pour un objet :

pour $p \notin S_i$ mais 8-connecté à S_i ,

si $(|f(p) - m_i^s| < |f(p) - m_i^c|)$,

alors p est affecté à S_i .

Cette analyse se poursuit jusqu'à ce que plus aucun nouveau pixel ne soit affecté à un des objets. Signalons que les valeurs moyennes ne sont pas redéfinies après chaque itération.

Ces analyses restent, par l'utilisation de la morphologie mathématique et des premiers voisins, toujours locales. Chaque objet est en effet jugé dans son contexte propre.

De plus, le facteur k peut évoluer suivant le type d'application. Certains critères, comme des contraintes de forme par exemple, peuvent également être introduits en terme de caractérisation de chaque objet.

5. Résultats et discussion

La méthode a été jugée sur une série d'images avec des résultats satisfaisants. L'image test présentée figure 2 a été choisie car elle est représentative du type des images étudiées. Sa complexité vient du fait que les événements à isoler dans l'image sont très diversifiés et le fond très inhomogène globalement.

Tout d'abord, précisons les caractéristiques extérieures de cette image obtenues par la classification en micro/macro perturbation (Tab. 1). Celle-ci se range dans la catégorie macro-perturbée, d'où une valeur σ de la gaussienne, une taille N de détection des maxima locaux et une hauteur T des pyramides locales plutôt grandes.

Tableau 1. – Caractéristiques associées à l'image test.

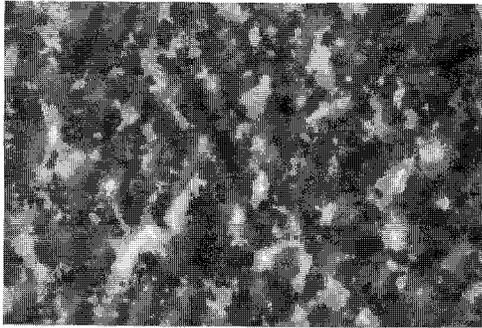
σ de la gaussienne utilisée par le filtrage	taille N du voisinage de détection des germes	hauteur T des pyramides locales
3	10	5

On obtient alors l'image des germes \mathcal{G} (Fig. 12). Afin de ne pas trop saturer les images et de ce fait noyer l'information, les points critiques de la méthode vont être exhibés sur deux zones particulières caractérisées sur la figure 12 par deux rectangles lumineux. La première zone (zone 1) va servir de test pour l'obtention des bases englobantes alors que la seconde zone (zone 2) va montrer l'influence des étapes de décroissance et croissance de région.

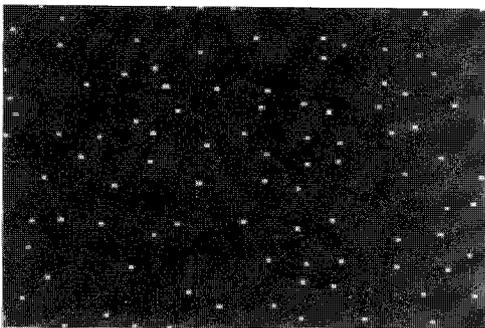
A présent, il paraît opportun de s'intéresser à la borne de 80% qui est apparue dans le choix des intervalles d'existence. Pour cela, les pourcentages des pixels conservés à l'intérieur des bases englobantes sont donnés (Tab. 2). En fait, pour chaque base englobante est calculé un pourcentage de pixels conservés sur un critère de niveaux de gris, puis de contraste et enfin de cumul des deux. Les nombres fournis correspondent aux valeurs moyennes des pourcentages obtenus sur la totalité de l'image. Les écarts-type associés sont également donnés. On retrouve sur cet exemple la borne associée et on remarque l'effet de l'utilisation du contraste L.I.P. La hauteur T des pyramides locales est élevée pour ne pas risquer des bases englobantes insuffisantes (paragraphe 3.2.2).

Tableau 2. – Pourcentage des pixels conservés dans les bases englobantes. Les nombres donnés correspondent aux moyennes et écarts-type calculés sur l'ensemble des bases englobantes de l'image.

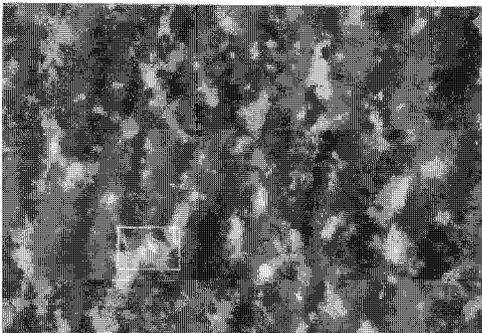
hauteur T des pyramides locales	5
moyenne en niveaux de gris	81.97
écart-type en niveaux de gris	2.80
moyenne en contraste	84.13
écart-type en contraste	3.15
moyenne pour le cumul des deux critères	72.49
écart-type pour le cumul des deux critères	4.03



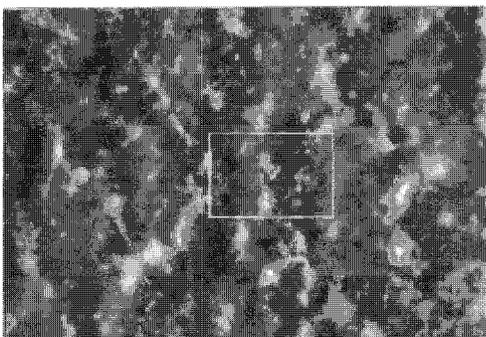
(a) image test.



(b) image des germes sur un voisinage de taille 10.



(c) zone 1 d'étude des bases englobantes.



(d) zone 2 d'étude des effets des analyses.

Figure 12. – Images (a) à (d) : L'image (b) a subi une dilatation de taille 1.

Considérons à présent l'exemple de la zone 1 (Fig. 13). Celui-ci permet de juger la base englobante associée à un objet. Dans l'image 13.(c) apparaissent les pixels conservés par application de l'intervalle d'existence en niveaux de gris. Les pixels non conservés y apparaissent en noir. 83% des pixels sont dans ce cas gardés. 85% le sont dans le cas du contraste; ce qui finalement impose une conservation cumulée de 77% des pixels dans la base englobante. L'image 13.(d) permet de visionner les pixels gardés à la fin de la segmentation, c'est-à-dire après les étapes de décroissance et croissance de région. Les zones blanches émergentes ne sont pas assez éloignées pour être déconnectées et vont donc apparaître sous la forme d'un seul objet dans la segmentation finale.

Pour juger de l'influence des étapes de décroissance et croissance de région, il est préférable de s'intéresser à la seconde zone d'étude présentée figure 12 (Fig. 14).

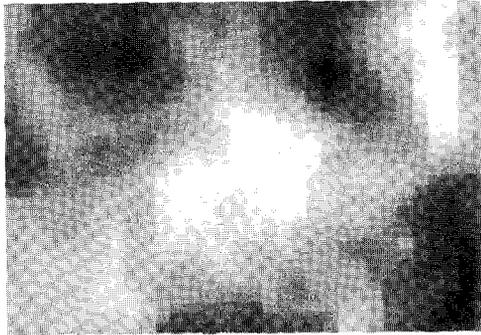
L'image 14.(b) montre l'exemple de deux objets connectés «à tort» en sortie de la segmentation par les pyramides locales. La décroissance de région les déconnecte alors suivant l'image 14.(c). Finalement, la croissance de région se charge de reprendre le deuxième segment pour aboutir à la segmentation définitive de l'image 14.(d).

Enfin, après avoir étudié les différents points critiques de la méthode, intéressons-nous aux résultats sur l'image totale (Fig. 15).

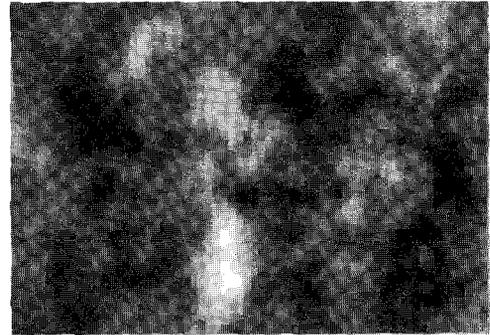
Les résultats ont été jugés *perceptuellement* satisfaisants. En effet, la méthode paraît extraire de l'image les objets qui à l'œil seraient retenus. De plus, l'image est telle que seule une approche véritablement *locale* permet de résoudre convenablement le problème. Les méthodes classiques de segmentation ne donnent logiquement pas des résultats satisfaisants. Maintenant, la qualité de la segmentation se juge généralement par le système visuel humain en fonction de critères très précis et difficilement implémentables automatiquement. En tout cas, les résultats restent *stables* et *reproductibles* et donc susceptibles d'être utilisés dans une étape d'analyse d'un processus de cotation visuelle assistée par ordinateur. L'étape d'*analyse* va consister à extraire les traits caractéristiques des différents segments de l'image, permettant ainsi de côter une surface par une statistique des irrégularités topographiques. Pour finir, la méthode basée sur les pyramides locales permet d'intégrer les paramètres de micro/macro perturbation (partie 1) indispensables pour permettre une analyse objective et cohérente. Les images classifiées en micro-perturbation auront des variables σ , N et T plus faibles que dans le cas typique de l'image test macro-perturbée.

6. Conclusion

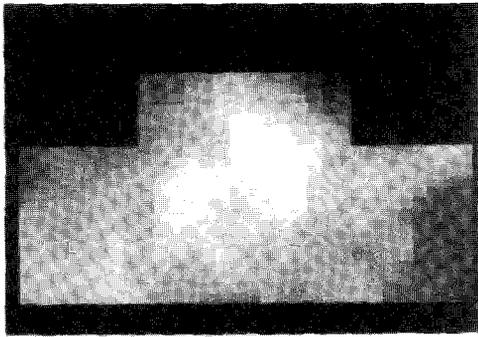
Nous avons développé une méthode de segmentation en accord avec le principe de focus d'attention, totalement locale et adaptative. Testée sur des séries d'images, diversifiées et inhomogènes, elle conduit à des résultats perceptuellement satisfaisants. Stables et reproductibles, ceux-ci sont intégrés dans un processus de cotation visuelle assistée par ordinateur [2].



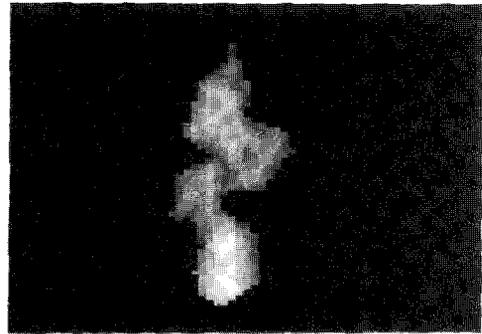
(a) zone 1 d'étude.



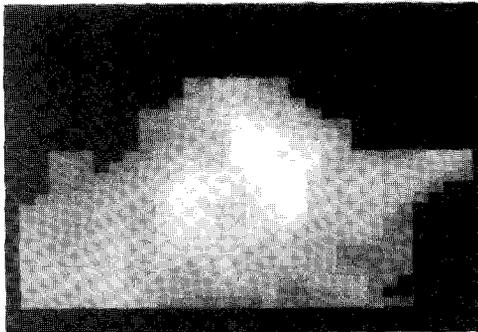
(a) zone 2 d'étude.



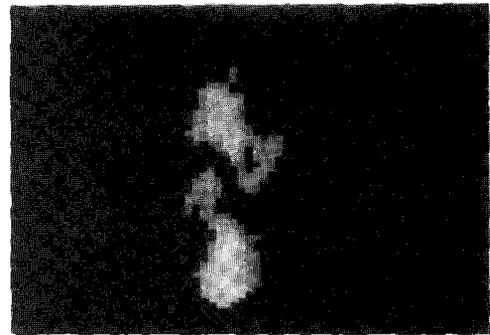
(b) base englobante associée.



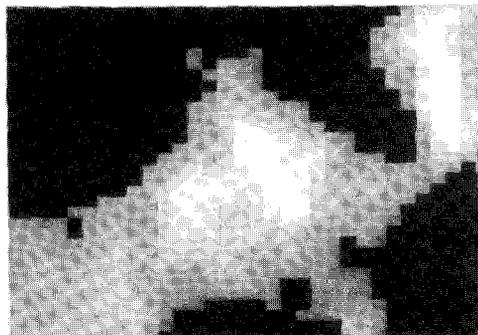
(b) segment sortie des pyramides.



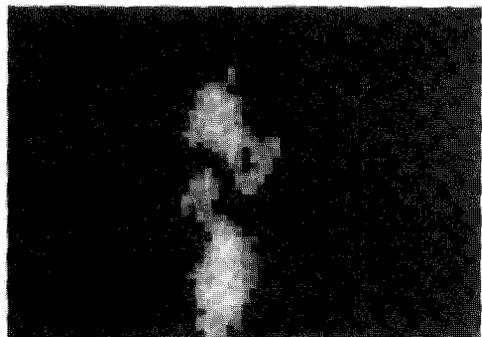
(c) pixels conservés en niveaux de gris.



(c) segment sorti de la décroissance de région.



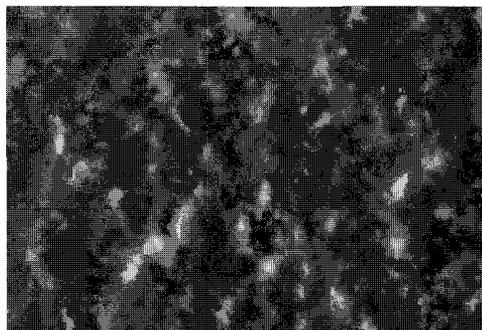
(d) pixels conservés dans la segmentation finale.



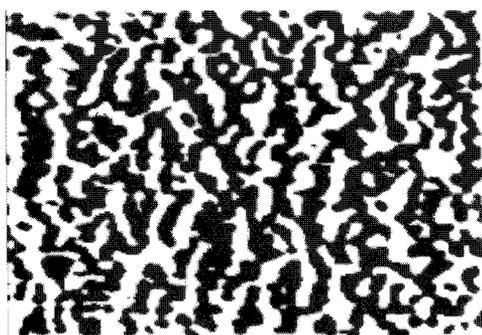
(d) segment final après la croissance de région.

Figure 13. – Images (a) à (d) : Les pixels non retenus apparaissent en noir.

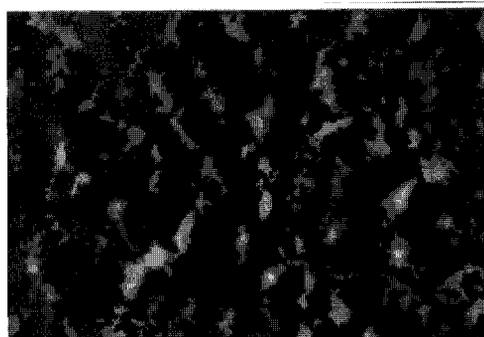
Figure 14. – Images (a) à (d).



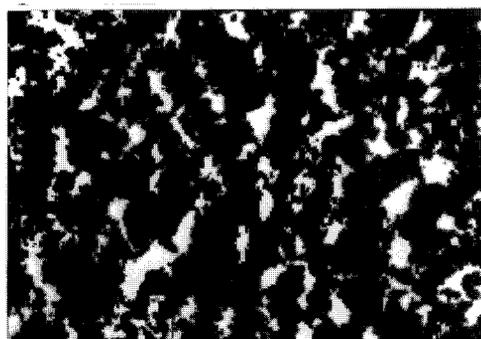
(a) image test.



(b) image de contraintes binaire.



(c) image des objets.



(d) image des objets binaire.

Figure 15. – Images (a) à (d).

L'approche suivie permet de segmenter automatiquement des images de texture complexes. Les objets à extraire doivent être détectables par des germes, qui sont dans notre cas des maxima locaux. En effet, la pyramide locale permet de focaliser l'attention sur l'objet critique en faisant abstraction de l'aspect global de l'image. Ainsi, aucun autre *a priori* n'est nécessaire sur les objets à extraire et le fait de rester toujours local permet de s'affranchir de toute dérive topographique.

Cependant, il est clair que certains problèmes pourraient apparaître pour des images où les objets sont trop homogènes. La détection des germes serait dans ce cas problématique, engendrant des objets mal détectés et donc insuffisamment décrits. Des efforts sur ce point pourraient permettre d'améliorer encore la méthode, par addition si nécessaire de contraintes de forme. Il faut préciser que les dernières corrections à base de morphologie mathématique (partie 4), postérieures à la méthode pyramidale proprement dite, sont suffisamment maléables pour intégrer des contraintes liées au type d'application.

BIBLIOGRAPHIE

- [1] B. REDORTIER. Contribution de l'analyse de Fourier à la segmentation et à la quantification des surfaces multi-texturées. Application à l'analyse d'aspect des surfaces textiles, thèse de doctorat, Université de Saint-Etienne, année 1992.
- [2] M. CALONNIER. Contribution de l'imagerie numérique à l'étude des matériaux textiles, thèse de doctorat, Université de Saint-Etienne, année 1990.
- [3] J.M. JOLION. Analyse d'images : le modèle pyramidal, *Traitement du Signal*, vol. 7, n° 1, 1990, p. 5-17.
- [4] S. TANIMOTO, T. PAVLIDIS. A hierarchical data structure for picture processing, *Computer Graphics and Image Processing*, Vol. 4, n° 2, juin 1975, p. 104-119.
- [5] A. ROSENFELD. *Multiresolution Image Processing and Analysis*, Springer, Berlin, 1984.
- [6] V. CANTONI, S. LEVIALDI. *Pyramids Systems for Computer Vision*, Springer, Berlin, 1986.
- [7] P. MEER, S.N. JIANG, E.S. BAUGHER, A. ROSENFELD. Robustness of image pyramids under structural perturbations, *Computer Vision, Graphics, and Image Processing*, Vol. 44, 1988, p. 307-331.
- [8] D. MARR. *Vision*, W.H. Freeman and Company, San Francisco, 1982.
- [9] P.J. BURT. Fast filter transforms for image processing, *Computer Graphics and Image Processing*, Vol. 16, 1981, p. 20-51.
- [10] J.M. JOLION, A. MONTANVERT. La pyramide adaptative : construction et utilisation pour l'analyse de scènes 2D, *7^e Congrès RFIA, AFCET-INRIA*, Paris, novembre 1989, p. 197-206.
- [11] A.D. GROSS, A. ROSENFELD. Multiresolution object detection and delineation, *Computer Vision, Graphics and Image Processing*, Vol. 39, 1987, p. 102-115.
- [12] A. ROSENFELD, A.C. SHER. Detection and delineation of compact objects using intensity pyramids, *Pattern Recognition*, Vol. 21, n° 2, 1988, p. 147-151.
- [13] M. JOURLIN, J.C. PINOLI, R. ZEBODJ. Contrast definition and contour detection for logarithmic images, *Journal of Microscopy*, Vol. 156, pt 1, octobre 1988, p. 33-40.
- [14] J. SERRA. *Image Analysis and Mathematical Morphology*, Academic Press, Londres, 1982.
- [15] J.L. CROWLEY, A.C. PARKER. A representation for shape based on peaks and ridges in the difference of low-pass transform, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, n° 2, March 1984, p. 156-170.
- [16] J.L. CROWLEY, A.C. SANDERSON. Multiple resolution representation and probabilistic matching of 2-D gray-scale shape, *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, Vol. PAMI-9, n° 1, janvier 1987, p. 113-121.

- [17] P.J. BURT. Attention mechanisms for vision in a dynamic world, *Proc. of the 9th Int. Conf. on Pattern Recognition*, Rome, 14-17 november 1988, p. 977-987.

CURRICULUM VITAE

Hubert KONIK

DEA "Informatique, Image" de l'Université Jean-Monnet de Saint-Etienne en 1991. Prépare actuellement une thèse, intitulée "Contribution des méthodes multi-échelles à la segmentation des images textiles". Travaux financés par la Région Rhône-Alpes avec une bourse doctorale de recherche,

dans le cadre d'une collaboration entre le Laboratoire TSI, URA CNRS n° 842, et ITF-Lyon.

Maurice CALONNIER

Docteur en Informatique de l'Université Jean-Monnet de Saint-Etienne. Responsable de la plate-forme "Image et Textile" d'ITF-Lyon.

Bernard LAGET

Docteur ès Sciences Mathématique, Professeur à l'Université Jean-Monnet. S'intéresse aux applications industrielles des mathématiques.

Manuscrit reçu le 21 octobre 1992.

Dernière version le 16 mars 1993.