

Chapitre 3

La segmentation par régions

Chapitre rédigé par Henri MAÎTRE

L'approche duale de la détection des contours pour la décomposition d'une image en ses formes élémentaires est l'approche par régions. Elle repose sur la recherche de zones possédant des attributs communs, soit de luminosité, soit, plus rarement, de textures (nous n'aborderons pratiquement pas ce second point dans ces lignes, mais au chapitre 4).

Nous allons voir ici, tout d'abord, les méthodes utilisant l'histogramme, qui sont souvent très proches des méthodes de classification conventionnelles, mais dans lesquelles serait renforcé l'aspect de cohérence de zone. Nous verrons ensuite les méthodes pyramidales, prototypes des méthodes de subdivision hiérarchique descendante, puis les méthodes de croissance de régions, qui s'inspirent des méthodes de conquête de l'optimisation combinatoire, et les méthodes de partage et réunion qui tirent profit simultanément des avantages des deux méthodes précédentes.

Enfin, nous donnerons très brièvement un aperçu de méthodes à base de règles qui se proposent de combiner approches par contours et approches par régions dans un seul vaste programme de segmentation.

3.1 Les méthodes sur histogramme

Ces méthodes sont de mise en œuvre assez simple et de performances souvent réduites car elles ne tirent pas profit de l'aspect spatial de l'information d'image. Elles sont recommandées dans les cas suivants :

1. lorsque les images présentent des classes évidentes : documents écrits ou schémas en noir et blanc ou en couleur, objets très contrastés (par exemple cellules d'une biopsie ou avion sur un ciel), etc.
2. lorsque les images sont définies sur de nombreux canaux (images multi- ou hyper-spectrales), ce qui enrichit l'information portée par l'histogramme.

L'idée générale de ces méthodes consiste à isoler des pics de l'histogramme. À une dimension on procède donc à des seuillages ou des multi-seuillages. À n-dimensions on procède à des classifications [Dubuisson, 1990].

3.1.1 Avec apprentissage

Seuillage bayésien

Si l'on dispose d'une connaissance sur les classes que l'on recherche (en particulier la probabilité d'occurrence d'un niveau de gris pour chaque classe, et la probabilité à priori des classes), alors on peut aisément se replacer

dans les conditions de la théorie bayésienne de la décision [Dubuisson, 1990, Duda et Hart, 1973] et choisir les seuils qui minimisent le coût des fausses erreurs.

Pour des histogrammes à 1 dimension et pour 2 populations X et Y , en dénotant par :

- $P(X)$ et $P(Y)$ les probabilités à priori des classes X et Y (sous la contrainte $P(X) + P(Y) = 1$),
- $P(n/X)$ et $P(n/Y)$ les probabilités conditionnelles qu'un pixel de la classe X ait un niveau de gris n (et idem pour Y),
- C_X et C_Y les coûts des mauvaises classifications de X et Y ,

en supposant ces grandeurs toutes connues, et sous l'hypothèse de stationnarité de l'image (c'est-à-dire que les propriétés statistiques sont invariantes en tout point de l'image), le seuil optimal est défini comme celui minimisant le coût global :

$$\Gamma(s) = C_x \int_s^N P(X) \cdot P(n/X) dn + C_y \int_0^s P(Y) \cdot P(n/Y) dn \quad (3.1)$$

Ce seuil est obtenu en testant pour tout n le rapport de vraisemblance $\Lambda(n)$ face au seuil s :

$$\Lambda(n) = \frac{P(n/Y)}{P(n/X)} \quad \text{et} \quad s = \frac{C_y P(Y)}{C_x P(X)}$$

Seuillage de Neyman-Pearson

Une autre décision peut se faire selon le critère de Neyman-Pearson. Dans ce cas on s'intéresse en particulier à une classe (ici on choisit X). On définit la probabilité de fausse alarme pour cette classe :

$$P_f = \int_0^s P(Y) P(n/Y) dn$$

On maximise alors la probabilité de détection pour une probabilité de fausse alarme donnée. Dans ce cas on considère que la fausse alarme est l'erreur la plus grave. On fixe sa probabilité à une valeur acceptable : $P_f = \alpha$. La probabilité de détection de X est donnée par :

$$P_d = \int_0^s P(X) P(n/X) dn$$

Le seuil de Neyman-Pearson est donné par la méthode du lagrangien :

$$L = P_d - \lambda(P_f - \alpha) \quad (3.2)$$

$$= \lambda\alpha - \int_0^s [P(n/Y)P(Y) - P(n/X)P(X)] dn \quad (3.3)$$

et la décision se fait en comparant le rapport de vraisemblance $\Lambda(n)$ à α .

Cette décision est moins sensible aux probabilités à priori et conduit en particulier à des décisions plus proches de nos choix intuitifs que la décision bayésienne dans le cas où un événement est très rare.

D'autres critères existent bien sûr qui s'appuient sur d'autres choix statistiques (par exemple minimum d'entropie).

3.1.2 Seuillage sans apprentissage

Lorsque l'on dispose du seul histogramme pour extraire des classes, on recherche généralement des modes de cet histogramme qu'on isole par des seuillages au creux des vallées. Souvent les histogrammes sont trop irréguliers et il convient alors de les filtrer, soit par des fenêtres glissantes, soit par des équations de diffusion (éventuellement isotropes). De nombreuses règles empiriques ont été proposées pour choisir les seuils automatiquement qui ne sont guère généralisables [Kittler et al., 1985, Weszka, 1978].

3.1.3 Méthodes de classification

Disposant d'un histogramme, éventuellement multidimensionnel, la plupart des techniques de classification s'appliquent à sa segmentation. Les plus utilisées sont :

- les techniques de nuées dynamiques (*k-means*) qui procèdent alternativement en classifiant au plus proche voisin le nuage des points, selon une distance à des noyaux donnés, puis en estimant la position des meilleurs noyaux de ces classes ainsi obtenues. Il est important pour cette méthode de disposer du nombre de classes recherchées. Si l'on ne le connaît pas, on choisit souvent des critères entropiques (comme le critère d'Akaike ou le critère de Hannan et Quinn [Olivier et al., 1997]) qui permettent d'évaluer des classifications obtenues avec des nombres de classes différents¹.
- les réseaux neuromimétiques et en particulier les cartes de Hopfield.
- et, si l'on dispose d'un certain nombre d'échantillons déjà classés, les plans, ou les courbes, séparateurs, ainsi que les k-plus-proches voisins.

Dans les espaces de grande dimension (imagerie hyperspectrale par exemple), on peut avoir intérêt à réduire tout d'abord la dimension des espaces pour éviter d'avoir à estimer des probabilités très faibles. On peut le faire par ACP (analyse en composantes principales) ou par analyse de Karhunen-Loeve. Ces deux transformations sont identiques à une normalisation près des axes. Elles procèdent en projetant le nuage de points sur le sous-espace construit à partir d'un nombre limité des vecteurs propres de la matrice de covariance des données.

Dans les espaces de dimension p , les distances utilisées entre deux nuages de points caractérisés par leur moyenne (vectorielle) m_i et leur matrice de covariance Γ_i sont, par ordre de complexité croissante [Fukunaga, 1990, Zhang et Modestino, 1990] :

- la distance euclidienne qui pondère également toutes les variables :

$$(m_1 - m_2)^t (m_1 - m_2)$$

- la distance de Mahalanobis (par exemple du nuage 2 par rapport au nuage 1), qui rend compte de l'orientation des inerties des nuages dans l'espace :

$$(m_1 - m_2)^t \Gamma_2^{-1} (m_1 - m_2)$$

- la distance de Bhattacharyya, qui permet de distinguer des lois de mêmes moyennes, mais de variances différentes. :

$$1/4(m_1 - m_2)^t (\Gamma_1 + \Gamma_2)^{-1} (m_1 - m_2) + \frac{1}{2} \ln \frac{\frac{1}{2}|\Gamma_1 + \Gamma_2|}{\sqrt{|\Gamma_1||\Gamma_2|}}$$

3.1.4 Sélection sur l'histogramme et dans l'image

Le défaut des approches par classification est de complètement négliger les relations spatiales entre pixels, pour ne s'attacher qu'à leurs propriétés de radiométrie. Pour pallier ce défaut, dans une approche proposée dans [Ohlander et al., 1978] et souvent reprise, on procède de façon itérative dans l'espace image et sur l'histogramme :

1. sur l'histogramme on sélectionne un mode isolé,
2. parmi les zones de l'images contribuant à ce mode on sélectionne la zone connexe la plus grande.

On itère ce processus, soit en subdivisant à nouveau l'histogramme de la zone retenue, soit en s'occupant d'un autre mode de l'histogramme original.

3.1.5 Sélection sur histogramme et régularisation

Mais si l'on souhaite améliorer les propriétés spatiales des zones obtenues par sélection de modes sur l'histogramme ou par classification, l'une des méthodes les plus appropriées est de modéliser le problème par un champ

¹On note cependant avec [Olivier et al., 1997] que ces modèles ont tendance à sur-évaluer le nombre de classes trouvées.

markovien (voir chapitre ??). On considère que les régions forment une partition sur l'image (cf. section 3.2). Chaque région est représentée par une fonction caractéristique et identifiée par une étiquette $\lambda_i \in E$. Le champ d'étiquettes est un champ caché à l'utilisateur qui n'observe que la réalisation bruitée de l'image $f(x, y)$ en chaque pixel, et qui va essayer d'estimer la meilleure distribution des étiquettes λ_i connaissant les n (généralement selon le critère du Maximum A Posteriori de Λ (MAP)).

Dans la formalisation markovienne on passe d'une représentation probabiliste à une représentation en énergie (généralement au moyen d'une formalisation bayésienne de la probabilité à postériori). Très souvent l'énergie (représentant la probabilité à postériori de la classe sachant l'observée) est constituée de deux termes : l'un traduit la similarité de la classification aux données que l'on observe, c'est le terme d'attache aux données (probabilité des observées conditionnellement aux classes), le second exprime les a priori que nous avons sur les classes (par exemple classes compactes, aux contours réguliers ou de forme prédefinie, dans des relations de voisinage particulières, probabilité à priori de classe, etc.). Des hypothèses peuvent également être faites sur les distributions des niveaux de gris (par exemple lois gaussiennes sur les régions), des étiquettes (dépendance ou indépendance spatiale), ainsi que des distributions conditionnelles des niveaux de gris sachant les étiquettes.

Dans la démarche la plus simple on utilise alors souvent un terme d'attache aux données de la forme :

$$U_0 = (f(x, y) - \mu_i)^2$$

où $f(x, y)$ est le niveau de gris du pixel, μ_i le niveau de gris de la classe λ_i dans laquelle on a placé le pixel (x, y) . Un tel terme est équivalent à une probabilité gaussienne d'appartenance à la classe λ_i :

$$p(f(x, y) | \lambda_i) \propto \exp(-U_0) = \exp\left[-\frac{(f(x, y) - \mu_i)^2}{2\sigma_i^2}\right]$$

Le terme de régularisation est calculé sur les cliques $\{c\}$ constituant le voisinage que l'on souhaite donner à (x, y) (cf. figure ??). Il prend en compte la compatibilité $\Phi(\lambda_i, \lambda_k(x', y'))$ entre l'étiquette λ_i attribuée à (x, y) et celle $\lambda_k(x', y')$ attribuée à (x', y') :

$$E_1 = \sum_{(x', y') \in \{c\}} \Phi(\lambda_i, \lambda_k(x', y'))$$

Par exemple, dans le cas le plus simple, on peut attribuer une énergie négative si $\lambda_i = \lambda_k$ et positive dans le cas contraire (modèles de Potts et d'Ising, voir chapitre ??). Des modèles plus complexes peuvent prendre en compte des compatibilités plus subtiles (classes plus ou moins proches). On recherche alors un minimum de $U = U_0 + \theta U_1$ par toutes les techniques usuelles d'optimisation itératives (ICM, recuit simulé, etc.). Le coefficient θ permet de pondérer les effets de la régularisation par rapport à ceux de la classification. Pour un θ fort, on obtiendra des classes très compactes et grandes au détriment d'une bonne séparation des classes sur l'histogramme [Derin et Elliott, 1987, Lakshmanan et Derin, 1989, Heitz et al., 1994, Won et Derin, 1992].

Dans d'autres approches, on peut se dispenser d'introduire des étiquettes de classes en introduisant explicitement un processus de bord, (cf. section ?? et [Geman et Geman, 1984]), processus b discret binaire (0 ou 1) ou continu (entre 0 et 1) qui prend, généralement, ses valeurs entre les pixels. Dans le cas binaire, lorsqu'il vaut 1 le voisin concerné n'a plus d'effet sur le site, selon la formule :

$$U_1 = \sum_{(x', y') \in \{c\}} (1 - b) \cdot \Phi(g_i, g_k(x', y'))$$

C'est donc l'équivalent d'un contour introduit dans l'image. Mais on pénalise usuellement l'introduction d'un contour par un terme de la forme :

$$U'_1 = \sum_{(x', y') \in \{c\}} b$$

ou par un coût de type MDL (Minimum Description Length) (cf. section 3.4).

Enfin, on peut également supprimer le processus de bord en le « cachant » à l'intérieur du potentiel U_1 . Pour cela on choisit une fonction Φ qui sature pour des valeurs fortes des différences entre niveaux de gris de sites voisins. On perd alors, avec la convexité de la fonctionnelle d'énergie, des bonnes propriétés de convergence du champ markovien et l'optimisation devient plus complexe (cf. figure ?? et section ??) [Blake et Zisserman, 1987, Black et Rangarajan, 1996].

3.2 Les méthodes par transformation de régions

Les méthodes que l'on va examiner maintenant s'appuient toutes sur la notion de **prédictat** et sur celle de **partition**.

Un prédictat \mathcal{P} est une proposition logique dont la valeur dépend de son argument. Nous nous intéresserons aux prédictats sur les régions \mathcal{R}_i de l'image.

Le prédictat de base de la segmentation est : *la région \mathcal{R}_i est homogène*. Pour vérifier le prédictat, parmi les arguments les plus utilisés en segmentation d'images, nous retiendrons les suivants :

- le contraste sur la région : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff \max_{\mathcal{R}_i} [f(x, y)] - \min_{\mathcal{R}_i} [f(x, y)] < \sigma$
- l'écart-type de la région : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff \sqrt{\frac{1}{N} \sum_{\mathcal{R}_i} (f(x, y) - m)^2} < \sigma$ (avec $N = \text{Card}(\mathcal{R}_i)$ et $m = \frac{1}{N} \sum_{\mathcal{R}_i} n(x, y)$),
- la distance interquartile sur la région, c'est-à-dire la distance séparant les 25 % inférieurs des 25 % supérieurs de l'histogramme,
- les différences limitées : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff \forall (k, j) \text{ voisins}, |f(k) - f(j)| < \sigma$,
- l'entropie : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff -\sum_{\mathcal{R}_i} p(f) \log(p(f)) < \sigma$.
- etc.

Un autre prédictat également utilisé est : *la région \mathcal{R}_i est distincte de ses voisines*. On utilise des arguments qui mettent en jeu les différences de valeurs moyennes, les distances inter-médianes, les contrastes moyens ou minimums aux frontières, etc.

Une partition Π est un ensemble de sous-ensembles \mathcal{R}_i (appelées régions) de l'image, vérifiant :

$$\Pi \iff \begin{cases} \forall i, j \in \mathcal{R}_i \cap \mathcal{R}_j = \emptyset \\ \bigcup_i \mathcal{R}_i = \text{support}(\text{image}) \\ \forall i \in \mathcal{R}_i \neq \emptyset \end{cases}$$

Lorsque toutes les régions d'une partition vérifient le prédictat, on dit que la partition le vérifie. La partition triviale où tous les pixels sont dans des régions différentes de cardinal 1 vérifie tout prédictat qui peut se vérifier sur une image. Tout prédictat qui ne peut être vérifié par la partition triviale est impossible. On connaît donc au moins une partition qui vérifie tout prédictat non impossible.

Il existe bien sûr un très grand nombre de partitions d'une image et il existe également généralement un très grand nombre de partitions qui vérifient le prédictat (il suffit, partant d'une telle partition, de subdiviser n'importe quelle région pour obtenir une nouvelle partition vérifiant le prédictat). On ne sait pas trouver toutes les partitions vérifiant le prédictat (la combinatoire est généralement inaccessible) mais on ne sait non plus généralement pas choisir entre des partitions vérifiant un prédictat. Des critères empiriques peuvent être utiles :

- le cardinal de la partition (à minimiser),
- la taille de la plus petite région (à maximiser),
- une « distance entre régions » (par exemple somme des distances entre zones adjacentes) (à maximiser).

Dans de nombreux cas par exemple, on ne recherche que des partitions maximales, c'est-à-dire telles que deux régions adjacentes ne vérifient pas le prédictat :

$$\forall (i, j : i \text{ et } j \text{ adjacents}) \Rightarrow \mathcal{P}(\mathcal{R}_i \cup \mathcal{R}_j) = \text{faux}$$

En l'absence de stratégie pour trouver la meilleure partition vérifiant un prédicat donné, nous allons décrire des stratégies de base qui sont très souvent utilisées.

3.2.1 La croissance de région

Les méthodes de croissance de région (*region growing*) n'aboutissent pas à des partitions car la propriété $\bigcup_i \mathcal{R}_i = \text{image}$ n'est généralement pas satisfaite. Partant de **germes**² on applique successivement à l'image des prédicats plus sévères que le prédicat \mathcal{P} . Ainsi on commence à associer aux germes les seuls pixels qui sont en très bon accord avec le prédicat. On réduit cette sévérité progressivement, et on se rapproche ainsi du prédicat objectif. La décision d'associer un pixel à une région se fait alors le plus souvent sans ambiguïté à moins que ses « distances » à deux régions soient égales (et en ce cas un choix quelconque est peu décisif). Plus importante est la décision de regrouper 2 régions qui sont adjacentes et vérifient le prédicat. Il a été montré qu'il convient alors d'être assez sévère pour fusionner des régions et qu'il est souvent préférable de traiter ce point lors d'une étape ultérieure en acceptant donc une **sur-segmentation** de l'image, plutôt qu'une fusion abusive qui ne serait plus récupérable.

Les méthodes de croissance de région cessent souvent de créer de nouveaux germes bien avant que le prédicat ne vaille \mathcal{P} et rejettent les régions de cardinal 1. Ainsi tous les pixels ne se retrouvent pas dans des régions (cf. figure 3.1) [Adams et Bishof, 1994, Beaulieu et Goldberg, 1989, Chang et Li, 1994].

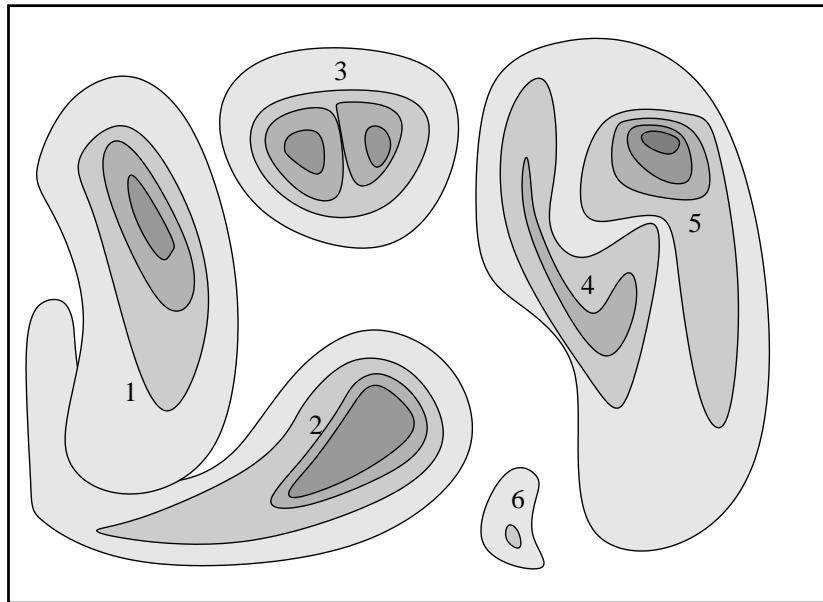


FIG. 3.1 – Croissance de région : 6 régions ont été obtenues par 4 prédicats successifs. La zone 6 n'a été créée qu'au second prédicat. La zone 3 regroupe 2 sous-régions qui ont été fusionnées car leur union vérifie le postulat et des critères annexes sur la forme résultante sont vérifiés. Ce n'est pas le cas de 1 et 2.

3.2.2 Le partage de région

Dans une démarche par partage de région (*region splitting*), on part de l'image entière (à laquelle généralement le prédicat \mathcal{P} ne s'applique pas). On appelle \mathcal{R} cette région. On lui applique plusieurs divisions δ produisant de nouvelles régions \mathcal{R}_i^δ . On teste \mathcal{P} sur chaque \mathcal{R}_i^δ , et on retient la meilleure subdivision δ , c'est-à-dire :

²les germes sont souvent des régions où le prédicat est trivialement vérifié, par exemple des zones où l'image est constante

- celle qui conduit à des sous-régions vérifiant toutes \mathcal{P} ,
 - ou celle qui donne le plus de sous-régions vérifiant toutes \mathcal{P} ,
 - ou, si aucune \mathcal{R}_i^δ ne vérifie \mathcal{P} , celle qui fournit la meilleure valeur à un critère \mathcal{E} dit critère d'échec.
- Pour δ choisi, chaque sous-région \mathcal{R}_i^δ ne vérifiant pas \mathcal{P} devient alors une région \mathcal{R} possible du traitement ci-dessus.

Les stratégies de subdivision sont peu nombreuses, on utilise généralement les 2 bipartitions régulières horizontale ou verticale (cf. figure 3.2), parfois des tripartitions régulières.

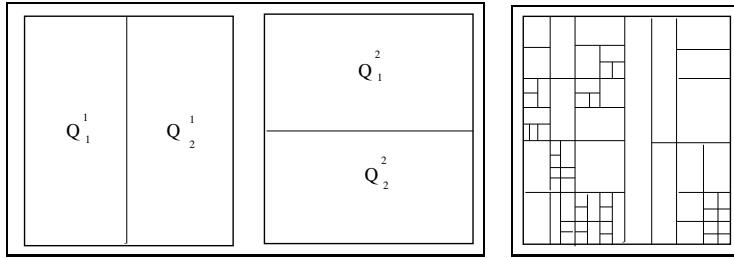


FIG. 3.2 – A gauche : partition d'une zone : on choisit entre la partition verticale ou la partition horizontale. A droite, à la fin du partage, l'image partitionnée.

Les critères d'échec \mathcal{E} sont souvent les mesures de variance ou de dynamique qui sont testées dans les prédictats \mathcal{P} . Lorsque toutes les régions ont été récursivement testées, il peut se trouver que des zones adjacentes le long d'une frontière d'ordre supérieur vérifient le prédictat. On peut alors les réunir avec les mêmes précautions que celles signalées au paragraphe 3.2.1.

Les algorithmes de partage sont mal adaptés à une mise en œuvre informatique sur machine séquentielle car les calculs effectués sur tous les pixels de la zone Q ne sont généralement pas réutilisables au niveau inférieur Q_i .

3.2.3 La réunion de région

Les techniques de réunion (*region merging*) prennent l'exact contre-pied des précédentes. Ce sont des méthodes ascendantes (*bottom-up*) lorsque les autres étaient descendantes (*top-down*). Les pixels sont visités systématiquement. Pour chaque carré de 2×2 pixels le postulat \mathcal{P} est testé, et s'il est accepté les pixels sont regroupés en une région. Après le parcours de toute l'image, les groupes de 2×2 régions se voient appliquer le même test et, éventuellement, les mêmes conséquences (réunion en une région de niveau 2) [Zhu et Yuille, 1996].

Les tests de réunion de région sont fréquemment faits sur des tests statistiques [Saporta, 1978]. On se place souvent dans l'hypothèse de bruits gaussiens sur des fonctions à valeur moyenne constante. Les tests les plus utilisés sont :

- le Chi2 (χ^2),
- le test de Wilcoxon, il travaille sur les pixels triés par ordre croissant de niveaux de gris des deux régions. Pour chaque pixel de la liste issue de la région 1 on compte combien de pixels de 2 sont plus sombres. Ces nombres sont additionnés pour donner une variable U de distribution asymptotiquement gaussienne de moyenne $n_1 n_2 / 2$ et de variance $n_1 n_2 (n_1 + n_2 + 1) / 12$. On teste donc cette valeur U face à sa distribution asymptotique.
- le test de Student d'égalité des espérances, on teste la variable de Student :

$$T(n_1 + n_2 - 2) = \frac{(m_1 - m_2) \sqrt{n_1 + n_2 - 2}}{\sqrt{(n_1 \sigma_1^2 + n_2 \sigma_2^2)(1/n_1 + 1/n_2)}}$$

- le test de Fisher-Snedecor d'égalité des moyennes et des variances, on teste la variable de Fisher :

$$F(n_1 - 1; n_2 - 1) = \frac{n_1 \sigma_1^2 (n_2 - 1)}{n_2 \sigma_2^2 (n_1 - 1)}$$

3.2.4 Les pyramides

Afin de bénéficier des avantages des deux méthodes précédentes, Horowitz a proposé une approche par pyramide [Horowitz et Pavlidis, 1976]. L'image est représentée sur une pyramide appelée *quad-tree*, constituée de N niveaux, l'image originale étant au niveau 0. Chaque pixel au niveau ν a 4 fils au niveau $\nu - 1$. Le pixel au niveau ν est la moyenne de ses 4 fils. Un schéma plus complet a été proposé par Burt [Burt, 1981] où les passages entre niveaux se font par filtrage gaussien. Une abondante littérature a été proposée depuis pour améliorer ces filtrages par une approche en ondelettes.

La segmentation procède directement à un niveau intermédiaire (par exemple $\nu = 2$), et tous les pixels fils sont testés avec le prédictat \mathcal{P} . Si le prédictat n'est pas vérifié, le pixel considéré est étiqueté comme candidat à un partage. Sinon il est candidat à une réunion avec ses voisins. C'est donc une technique de partage et réunion, où l'on profite du passage au niveau supérieur (ici $\nu = 3$) pour accélérer la procédure (cf. figure 3.3).

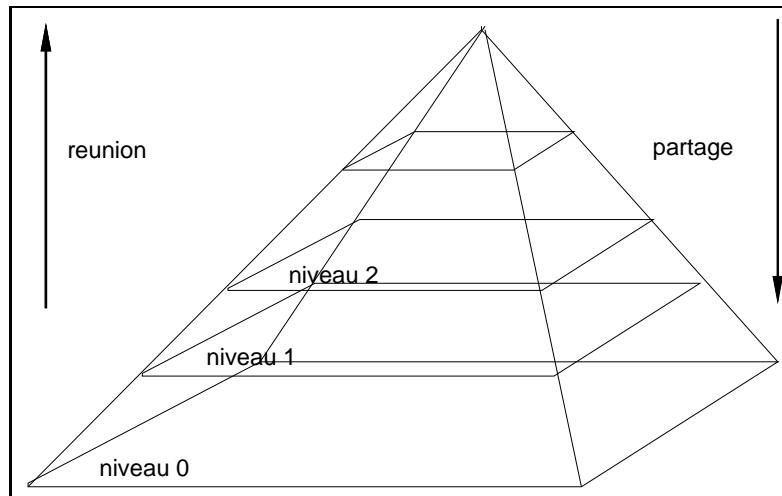


FIG. 3.3 – Méthode de pyramidage par quad-tree.

Mais la stratégie du quad-tree est trop rigide et conduit à des partitions trop régulières que les techniques de croissance ne permettent plus de rattraper. Des techniques inspirées de la pyramide ont été proposées qui donnent de bien meilleurs résultats. Ainsi, la méthode proposée dans [Suk et Chung, 1993]), autorise beaucoup plus de fusion que les simples réunions de pixels 2×2 . Douze fenêtres différentes sont autorisées, avec une priorité aux plus grandes (cf figure 3.4) (voir aussi [Chen et al., 1991]).

3.3 Les graphes d'adjacence

Les techniques par graphes d'adjacence sont beaucoup utilisées à partir de sur-segmentations (c'est-à-dire de segmentations où les zones sont subdivisées trop finement). Ces sur-segmentations sont par exemple le résultat d'un prédictat trop sévère dans la phase de segmentation où d'un algorithme très sensible aux variations locales comme la technique des lignes de partage des eaux (cf. chapitre ?? et [Schmitt et Mattioli, 1994]).

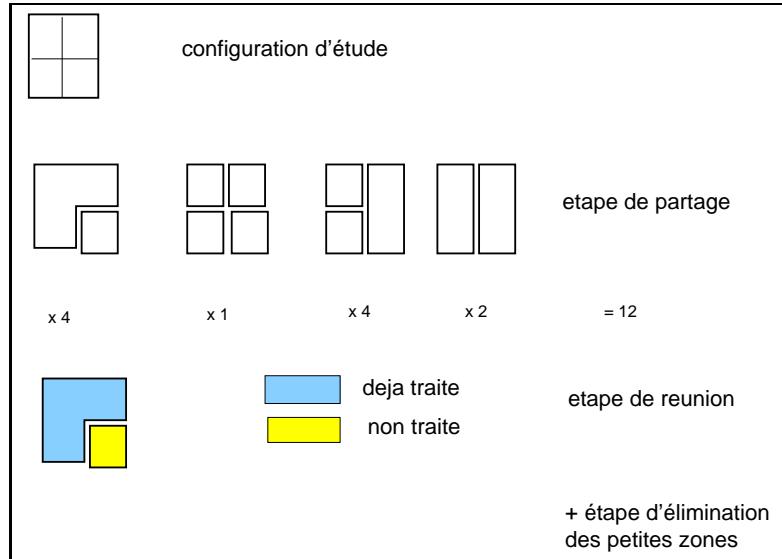


FIG. 3.4 – Les 3 étapes de la méthode de partage et réunion de Suk. Pour chaque étape il y a un prédicat différent. Celui du partage et de la réunion est généralement un prédicat de contraste, celui sur l'élimination des petites zones est un prédicat de taille. Les petites zones sont associées à la zone la plus proche en niveaux de gris.

L'idée de base consiste à plonger les régions obtenues dans une structure de graphe où une région est un noeud et un arc une relation d'adjacence. Puis on définit une fonction de similarité entre 2 noeuds. On trie tous les couples de noeuds adjacents dans une liste ordonnée. On regroupe les 2 meilleurs candidats. On remet à jour la liste et on itère.

La méthode proposée par Beaulieu et Goldberg [Beaulieu et Goldberg, 1989] est un bon exemple d'une telle technique. Le critère d'homogénéité $H(\mathcal{R}_i)$ d'une zone \mathcal{R}_i est le résidu de son approximation par un polynôme de degré d fixé :

$$H(\mathcal{R}_i) = \sum_{(x,y) \in \mathcal{R}_i} \left[f(x,y) - \sum_{p+q \leq d} a_{p,q}^i x^p y^q \right]^2$$

et plus particulièrement d'une approximation par une constante. En partant d'une partition très fine en régions très homogènes, et en acceptant progressivement des réunions de régions de moins en moins semblables, on s'approche de l'optimisation du critère global de minimum de variance sur les partitions.

D'autres schémas plus complexes sont mis en œuvre, par exemple dans [Wang, 1998] pour obtenir des segmentations meilleures.

3.4 La méthode MDL = Minimum Description Length

C'est une technique issue de la théorie stochastique de l'information [Rissanen, 1984, Rissanen, 1987, Rissanen, 1989] pour optimiser la représentation de données. Elle a été reprise de façon simplifiée dans divers domaines du traitement des images (en particulier en codage). L'idée du MDL consiste à exploiter l'analogie, au sens de la théorie de l'information, entre longueur minimale de description et quantité d'information. On cherche donc à optimiser le choix d'un modèle pour décrire les données conduisant à la plus courte description :

- d'une part par le choix d'un modèle qui s'adapte bien aux données,
- d'autre part par le choix d'un modèle simple qui nécessite peu de paramètres.

L'optimisation au sens du MDL procure un compromis entre ces deux termes. Dans le cas de la description d'une image, le modèle doit décrire les contours des régions et approcher au mieux l'image entre les contours.

On se place donc dans le domaine de la représentation paramétrique des données. On dispose d'un modèle dépendant de paramètres et on recherche le meilleur jeu de paramètres pour approcher les données. Soit $\phi = (\phi_1, \dots, \phi_k)$ le vecteur à k composantes des paramètres. Soit x l'observée (l'image). Dans le cadre de l'estimation au maximum de vraisemblance, on cherche le meilleur $\phi \in \mathbb{R}^k$ qui maximise $p(x|\phi)$. Cela revient à maximiser la log-vraisemblance :

$$\mathcal{L}(x|\phi) = -\log p(x|\phi) \quad (3.4)$$

Si l'on dispose de plusieurs modèles, chacun ayant des jeux différents de ϕ_i , la transmission de la représentation de x nécessite la transmission d'une part des paramètres décrivant la forme dans ce modèle, d'autre part les éléments permettant de reconstruire le modèle puisqu'il existe de nombreux modèles potentiellement disponibles. On sait que selon la théorie de l'information, il existe des codes optimaux pour transmettre le premier terme en utilisant un nombre de bits égal au logarithme de la probabilité $p(x|\phi)$, c'est à dire le terme de l'équation 3.4. On devra ensuite transmettre le second terme. Pour rechercher la meilleure description au sens de la théorie de l'information, on minimisera donc une expression plus complète que celle utilisée pour la représentation au maximum de vraisemblance (équation 3.4) :

$$\mathcal{L}'(x|\phi) = -\log p(x|\phi) + \mathcal{L}(\phi) = \mathcal{L}(x|\phi) + \mathcal{L}(\phi) \quad (3.5)$$

où $\mathcal{L}(\phi)$ représente la quantité d'information nécessaire pour transmettre le modèle. Dans le cadre général de la classification ou du traitement du signal, il est parfois possible de mener une optimisation complète et explicite du MDL.

Utilisé dans le cadre de la segmentation d'image, le MDL introduit dans la segmentation un terme $\mathcal{L}'(x|\phi)$ de pénalité lors d'une représentation en régions. Il vient limiter d'une part le nombre de contours (le premier terme $\mathcal{L}(x|\phi)$), d'autre part la complexité des contours par le choix si possible d'un petit nombre de paramètres très simples à coder (le second terme : $\mathcal{L}(\phi)$).

Dans la pratique, il n'est généralement pas possible en traitement des images d'obtenir une solution explicite de l'équation 3.5 et l'on recherche très souvent des solutions itératives (par modification des lignes de contour) sous le contrôle du terme 3.5. Le MDL est utilisé en segmentation d'images de différentes façons. On exprime généralement :

1. le coût de la représentation d'une région par une fonction (par exemple constante). Ce coût s'exprime par le nombre de bits nécessaires à coder la constante plus le nombre de bits nécessaires à coder l'erreur résiduelle,
2. le coût de la représentation des contours (coût de codage de la chaîne des contours). Une chaîne de contours peut être codée de façons très diverses : Freeman (cf. chapitre 5.4), approximations polygonales, splines, processus de bords, etc. et conduira alors à des codages très différents.

Le MDL tend à procurer un compromis entre ces deux termes, c'est-à-dire à donner peu de contours réguliers mais au bon endroit. Les techniques mises en œuvre en MDL sont à base de champs de Markov (optimisation par recuit simulé) [Lee, 1998, Leclerc, 1989, Zhu et Yuille, 1996], ou d'optimisation de courbes [Darell et al., 1990, Keren et al., 1990].

3.5 L'approche de Mumford et Shah

3.5.1 Approche formelle

C'est une façon formelle de poser le problème de segmentation [Mumford et Shah, 1989] qui n'a malheureusement pas de solution exacte. Soit $f(x, y)$ l'image³, de support \mathcal{I} , on la remplace sur des domaines \mathcal{R}_i par des fonctions régulières $g_i(x, y)$ qui se rejoignent le long de contours Γ_j (cf. figure 3.5). La fonctionnelle à minimiser se compose :

$$U(\Gamma, g, f) = \mu^2 \iint_{\mathcal{I}} (f(x, y) - g_i(x, y))^2 dx dy + \iint_{\mathcal{R}_i} \|\nabla g_i(x, y)\|^2 dx dy + \nu \int_{\Gamma_i} dl \quad (3.6)$$

On montre aisément que les 3 termes sont simultanément nécessaires sous peine de convertir l'expression en une formule triviale. Mumford et Shah ont déduit des propriétés particulières des solutions Γ_j et g , que nous examinons ci-dessous.

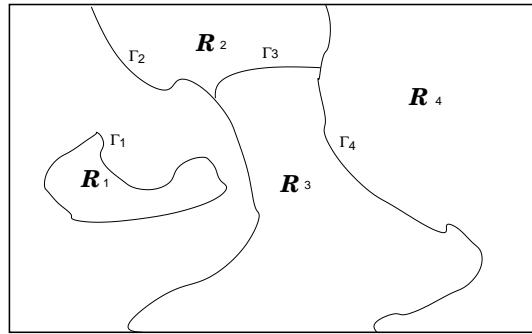


FIG. 3.5 – Approche de Mumford et Shah : on recouvre chaque domaine \mathcal{D}_i d'une fonction continue, on cherche à minimiser la distance entre la fonction de représentation et l'image, ainsi que la longueur des côtés.

On sait résoudre ce problème dans quelques cas simples (par exemple si g_i est une constante qui ne prend que deux valeurs (généralement choisies comme +1 et -1), on retrouve le problème d'Ising qui a une solution - en théorie exacte - par recuit simulé (cf. chapitre ??)⁴.

Mumford et Shah ont montré par ailleurs que l'équation générale 3.6 peut évoluer vers deux formes limites.

– Tout d'abord la minimisation de l'énergie U_o :

$$U_o(\Gamma, f) = \frac{1}{\mu^2} U(\Gamma, f, g) = \sum_i \iint_{\mathcal{R}_i} (f - g_i)^2 dx dy + \nu \cdot \int_{\Gamma_j} dl$$

³Attention dans le texte original de Mumford et Shah, [Mumford et Shah, 1989] les rôles de f et g sont interchangés, afin de garder la cohérence de nos notations nous avons préféré désigner par f l'image et par g son approximation.

⁴On se souvient cependant que les conditions de convergence vers le maximum absolu par recuit simulé sont liées à une descente en température infiniment lente.

que l'on obtient en imposant à g d'être constante par morceaux sur les \mathcal{R}_i : $g(x, y) = g_i$ ssi $(x, y) \in \mathcal{R}_i$. Il est alors aisément de montrer que $g_i = \frac{1}{s_i} \iint_{\mathcal{R}_i} f(x, y) dx dy$ où s_i représente la surface de \mathcal{R}_i (c'est donc la valeur moyenne sur la surface).

U est la limite de U_μ quand μ tend vers 0. La minimisation de U_μ est un problème bien posé. Le problème d'Ising évoqué plus haut en est un sous-problème réduit au cas à 2 classes, le problème de Potts l'étend à un plus grand nombre de classes.

- Puis la minimisation de l'énergie :

$$U_\infty(\Gamma) = \int_{\Gamma} \left[\nu_\infty - \left(\frac{\partial f}{\partial n} \right)^2 \right] dl$$

où ν_∞ est une constante et $\frac{\partial f}{\partial n}$ mesure la composante normale à Γ du gradient de f .

Ce problème peut se ramener à un problème de géodésique : on cherche simultanément à déterminer les trajectoires Γ_i de longueur minimale (terme en ν_∞) et à maximiser le gradient transverse de f le long de Γ_i ⁵. Ce problème est généralement mal posé si les Γ sont quelconques ; sur des contours de forme restreinte, le problème peut être bien posé.

On peut montrer que U_∞ est la limite de U quand μ tend vers l'infini.

On montre que si les Γ sont fixés (c'est-à-dire si l'on a fait par ailleurs une détection de contours), alors la minimisation de U est la recherche du minimum d'une fonction quadratique définie positive possédant un minimum unique. Elle est solution de :

$$\Delta g = \mu^2(g - f) \quad (3.7)$$

avec pour conditions aux limites :

$$\frac{\partial g}{\partial n}|_{\Gamma_j} = 0$$

Si les contours Γ_j sont des courbes régulières (généralement on recherche des solutions parmi les courbes de classe \mathcal{C}^2)⁶, on déduit alors les propriétés suivantes :

- les contours Γ_j ne peuvent posséder que des points singuliers de deux types :
- soit des points triples où 3 courbes se joignent à 120° ,
- soit des points extrémités d'où démarre une courbe unique (démarrage d'un pli) ;
- les contours se rattachent aux bords de l'image selon des angles droits ;
- par ailleurs, les fonctions g ont une dérivée normale aux Γ_j horizontale le long des Γ_j .

3.5.2 Les variantes autour de la formulation de Mumford-Shah

Des variantes du problème de Mumford et Shah ont été proposées qui abordent le problème de nombreuses façons. Par exemple des méthodes déterministes issues de la mécanique et de la résistance des matériaux ont été proposées pour apporter des solutions en termes de *membranes* ou de *plaques minces* [Blake et Zisserman, 1987]. On cherche alors à résoudre l'équation 3.7. Ces solutions conduisent à des équations données par :

- dans le cas d'une membrane :

$$\begin{aligned} U(g, f) = & \alpha_1 \cdot \sum_i \iint_{\mathcal{D}_i} (f(x, y) - g_i(x, y))^2 dx dy \\ & + \alpha_2 \cdot \sum_i \iint_{\mathcal{D}_i} (||\nabla g_i(x, y)||)^2 dx dy \end{aligned} \quad (3.8)$$

⁵On a également abordé ce problème lorsqu'on a évoqué les modèles géodésiques actifs (cf. paragraphe 2.4.4)

⁶Une courbe de classe \mathcal{C}^n possède n dérivées continues.

– dans le cas d'une plaque mince :

$$\begin{aligned} U(g, f) = & \alpha_1 \cdot \sum_i \iint_{\mathcal{D}_i} (f(x, y) - g_i(x, y))^2 dx dy \\ & + \alpha_2 \cdot \sum_i \iint_{\mathcal{D}_i} (\Delta g_i(x, y))^2 dx dy \end{aligned} \quad (3.9)$$

Il est également possible de prendre en compte dans ce second cas un terme de torsion (dérivées croisées) dans le terme de régularisation.

Dans les 2 cas les discontinuités Γ sont fixées (par exemple par un détecteur de contour préalable).

La résolution de ces problèmes (connaissant la position des contours on recherche la valeur en tout point de g_i) se fait traditionnellement par des techniques convolutionnelles. À chaque fonctionnelle est attachée une réponse impulsionale - ou fonction de Green - G . C'est-à-dire que, sur chaque domaine, \mathcal{D}_i , $g(x, y)$ est de la forme :

$$g(x, y) = \int \int_{\mathcal{D}_i} G(x, y, u, v) \cdot f(u, v) du dv$$

Par exemple en l'absence de contours, la membrane (équation 3.8) a pour fonction de Green K_o , fonction de Bessel modifiée de seconde espèce :

$$G(\rho = \sqrt{(x - x')^2 + (y - y')^2}) = \frac{1}{2\pi\lambda^2} K_o\left(\frac{\rho}{\lambda}\right)$$

équivalente à $\log(1/\rho)$ près de Γ et équivalente à $\frac{1}{\rho} \exp(-\rho)$ pour ρ grand. C'est la solution de l'équation sans second membre issue de l'équation 3.7 :

$$g - \frac{\alpha_2}{\alpha_1} \cdot \nabla^2 g = \delta(x, y)$$

La plaque mince (équation 3.9) a pour fonction de Green :

$$G(\rho) = \frac{i}{4\pi\mu^2} \left[K_o\left(\frac{\rho\sqrt{i}}{\mu}\right) - K_o\left(\frac{\rho}{\mu\sqrt{i}}\right) \right]$$

et elle est solution de l'équation :

$$g + \mu^4 \nabla^4 g = \delta(x, y)$$

avec $\mu^4 = \frac{\alpha_2}{\alpha_1}$.

Dans le cas où il n'y a pas de discontinuités de g le long des Γ , le problème est assez simple mais peu satisfaisant en termes de segmentation d'image [Grimson, 1981, Terzopoulos, 1983]. La solution est obtenue par des techniques de relaxation [Grimson, 1981] ou de relaxation multi-grilles [Terzopoulos, 1983]⁷.

Dans le cas discret on étend les solutions continues par des approches par éléments finis permettant de calculer de façon discrète les dérivées (on est alors dans un schéma très semblable à celui adopté pour les contours actifs vus au paragraphe 2.4).

On définit un processus de bord b booléen (cf figure ??) et on calcule les divers termes de l'énergie en fonction des valeurs $b = 0$ (pas de bord) ou $b = 1$ (présence d'un bord) de ce processus.

Si l'on ne connaît pas la position exacte des contours, on met en œuvre des techniques itératives et (souvent) sous-optimales qui partent de contours estimés et les améliorent. La résolution se fait par des techniques

⁷Les techniques multi-grilles sont des techniques d'optimisation qui utilisent la régularité spatiale de la solution pour trouver tout d'abord une solution exacte sur une grille d'échantillonnage très grossière, puis par un raffinement itératif de la solution, de la calculer sur des grilles de plus en plus fines pour aboutir à la résolution ultime de l'image.

de Markov avec processus de bord explicites [Geman et Geman, 1984] - cf paragraphe 3.1.5 - ou implicites et des schémas d'optimisation adaptés (par exemple la méthode GNC) ou des techniques à base de réseaux neuro-mimétiques de Hopfield. La méthode GNC (*Graduated Non Convexity* de Zisserman [Blake et Zisserman, 1987]), est particulièrement adaptée à ce problème. Elle remplace les fonctions d'énergie non convexes à optimiser par une famille de fonctions convexes qui, successivement optimisées par une technique de gradient, permettent d'atteindre l'optimum global.

Dans ces schémas, on peut aussi introduire un terme de pénalité sur la longueur des contours comme dans l'équation 3.6. On trouve dans la littérature des formes variées de cette énergie :

1. la plus simple est celle de Mumford et Shah et s'exprime par :

$$U = \int_{\Gamma} dl$$

2. une minimisation de la courbure des contours :

$$U = \int_{\Gamma} \left(\frac{d\theta}{dl} \right) dl$$

3. une pénalisation pour les seuls forts changements de direction des contours.

Des techniques où l'on mélange des termes de dérivées premières et secondes (comme dans les contours actifs - cf. paragraphe 2.4) sont également possibles, mais réputées moins bonnes.