ÉCOLE DOCTORALE SCIENCES ET INGÉNIERIE
de l'Université de Cergy-Pontoise

THÈSE

Présentée pour obtenir le grade de docteur d'université
Spécialité : Traitement des Images et du Signal

Modèles Multi-Échelles pour la Segmentation d'Images

par

Laurent Guigues

IGN - Laboratoire MATIS

le 12 Décembre 2003

Devant le jury composé de :

Mme Sylvie Philipp-Foliguet	Présidente
M. Jean-Michel Morel	Rapporteur
M. Philippe Salembier	Rapporteur
M. Jean-Pierre Cocquerez	Directeur de thèse
M. Hervé Le Men	Co-directeur de thèse
M. Jean-Michel Jolion	Examinateur

À ma mère

Remerciements

Ce mémoire présente des idées et méthodes auxquelles je suis parvenu après plusieurs années de recherches sur l'analyse automatique d'images, mais ne décrit pas la manière dont j'y suis parvenu. Ceux qui m'ont cotoyé durant ces années savent que ce mémoire est le résultat d'un cheminement de pensée, mais aussi d'un parcours personnel, l'un et l'autre parfois hasardeux ou difficile. Je remercie toutes les personnes qui m'ont aidé au cours de ce cheminement, de près ou de loin, à un niveau scientifique, technique ou humain. Cette thèse est aussi un peu la leur.

Je tiens en particulier à exprimer ma plus sincère gratitude à Hervé Le Men, qui m'a offert la possibilité de réaliser une thèse dans le cadre de mes activités à l'IGN et a accepté de diriger ces recherches pour l'IGN, et à Jean-Pierre Cocquerez qui a accepté d'être le directeur scientifique «externe» de ce travail. Cette co-direction a donné lieu à des échanges riches, à des confrontations de points de vue différents, et c'est réellement de la synthèse de ces points de vue que sont nées les idées et les méthodes développées dans ce mémoire. Par ailleurs, Hervé et Jean-Pierre m'ont accordé une grande liberté dans mes recherches. En particulier, le sujet d'origine était centré sur le développement d'une application spécifique, alors que ce mémoire pour sur des méthodes générales d'analyse d'image. Je leur suis profondément reconnaissant de la confiance qu'il m'ont accordé en autorisant mes recherches à prendre une orientation plus «fondamentale» que prévu. Je remercie également Serge Motet, chef du service de la recherche de l'IGN, ainsi que Daniel Piquet-Pellorce et Marc Pierrot Deseilligny, responsables successifs du laboratoire MATIS durant ma thèse.

Je remercie Sylvie Philipp-Foliguet, Jean-Michel Jolion, Jean-Michel Morel et Philippe Salembier pour avoir accepté de participer à mon juri, et ce en dépit du temps restreint que je leur avais accordé pour lire ce mémoire. Merci plus particulièrement à mes deux rapporteurs, Jean-Michel Morel et Philippe Salembier, pour leur lecture fine et leurs commentaires éclairants.

Je remercie également Emmanuelle Travet, responsable des thèses à l'université de Cergy-Pontoise, qui m'a été d'un grand secours dans la «dernière ligne droite».

Mes recherches se sont déroulées dans une équipe et j'ai tiré un grand bénéfice et un grand plaisir de mes échanges avec mes collègues et amis du MATIS, tant du point de vue personnel que scientifique. Je remercie toutes les personnes avec qui j'ai eu l'occasion de travailler ou d'échanger des idées, permanents, thésards ou stagaires.

Je remercie plus spécialement Olivier Jamet, sans lequel j'aurais certainement suivi une toute autre voie. Merci à Marc, qui m'a - entre autres - transmis son goût pour la théorie des graphes, l'optimisation combinatoire, l'algorithmique et la programmation «intelligente». Merci aux Fran(c)k, à Didier et à Jean-Marc, membres ou ex-membres du «noyau dur» du laboratoire. Je remercie plus particulièrement Frank : nos nombreuses discussions et nos recherches en commun m'ont beaucoup apporté. En mémoire des heures euphoriques où nous pourchassions un papillon et un domino... Je le remercie aussi pour ses commentaires sur une version «β» des premiers chapitres de ce document. Merci également à Franck le «jeune», qui a vécu de près les développements de l'escalade. Un grand merci à François, sans lequel le laboratoire ne serait pas ce qu'il est, et qui m'a été d'une aide précieuse en plusieurs occasions. Merci également à Alain et Marie-Claude.

Par ailleurs, je remercie affectueusement mon père, Jean-Luc, qui est aussi mon plus grand professeur. Je lui dois mon goût pour la recherche, l'interrogation et la réflexion en général. Plus ponctuellement, je le remercie pour sa relecture attentive et ses commentaires critiques sur ce mémoire. Je lui dois en particulier l'encadrement de la suite c de la proposition 6, qui est plus fin et s'obtient plus simplement que celui que j'avais proposé...

Un grand merci à Salomé, Julien, Fanfan, Valou, Jenny et Cyril pour leur hospitalité et leur soutien.

Un grand merci également à Jean-Phi, Seb Mus, Yann et Jean-Math.

Enfin, mille et un baisers à Karine, pour tout ce que je lui dois.

Table des matières

1 Introduction **1**

Questions de représentations **7**

2 Images et représentations graphiques **7**
2.1 Les notions d'image . . . 7
2.1.1 Images mathématiques . . . 7
2.1.2 Collections de mesures physiques . . . 7
2.1.3 Images numériques . . . 8
2.2 Deux procédés inverses . . . 9
2.2.1 Dématérialisation. . . . 9
Échantillonnage et résolution . . . 10
2.2.2 . . . et rematérialisation . . . 11
Coloration . . . 11
Projection et échelle de représentation . . . 11
2.3 Représentations cartographiques . . . 13
2.3.1 Généralisation cartographique . . . 13
2.3.2 Des images aux cartes . . . 14

3 Pour une approche multi-échelles de l'analyse d'images en régions **17**
3.1 Introduction . . . 17
3.2 Bas et haut niveau d'analyse visuelle . . . 18
3.2.1 Problèmes d'analyse de haut niveau d'une image . . . 19
Loconnaissance de formes . . . 19
Interprétation d'images complexes et vision artificielle . . . 20
3.2.2 Analyses de bas niveau et convergence vers une interprétation . . . 20
Descriptions de bas et de haut niveau . . . 21
Le principe d'économie de représentation de Mach . . . 22
Le haut niveau est le niveau utile au raisonnement . . . 23
3.3 Articulation entre bas et haut niveau d'analyse . . . 24
3.3.1 Groupements perceptuels . . . 24
«Le percept est donné à la conscience comme un tout» . . . 24
Le principe de Helmholtz . . . 25
3.3.2 La théorie de David Marr . . . 26
Différer l'«intelligence» . . . 26
Discussion . . . 27
3.3.3 Vision purposive et vision active . . . 28
3.3.4 Boucle analyse-synthèse . . . 29

3.3.5 Conclusion . . . 30
3.4 Traitements et analyses locales - points et contours . . . 30
3.4.1 Transitions plus ou moins rapides . . . 31
3.4.2 Théorie scale-space . . . 32
3.4.3 Pyramides multi-résolution et ondelettes . . . 34
Multi-résolution . . . 34
Analyses en ondelettes . . . 35
3.4.4 Points caractéristiques . . . 35
3.5 Segmentation en régions . . . 36
3.5.1 Hiérarchies de formes . . . 36
La hiérarchie sémantique . . . 36
Points de vue spécifiques . . . 37
Hiérarchisation perceptuelle ? . . . 38
3.5.2 Segmentation guidée par les connaissances . . . 39
Template matching . . . 39
Modèles déformables et ensembles actifs . . . 40
L'approche morphologique «classique» . . . 41
3.5.3 Stratégie segmentation-interprétation . . . 41
Le paradigme de partitionnement d'image . . . 42
Problèmes de fond posés par le partitionnement de bas niveau . . . 43
3.6 Conclusion : la thèse proposée . . . 45

4 Représentations mono et multi-échelle pour la segmentation d'images 49
4.1 Généralités - graphes et topologie discrète . . . 49
4.2 Treillis des parties et des partitions d'un ensemble . . . 50
4.2.1 Structures de treillis . . . 50
4.2.2 Parties et partitions . . . 50
4.2.3 Séquences et chaînes - couverture . . . 52
4.2.4 Représentation des treillis . . . 53
4.3 Représentation et manipulation de partitions d'images . . . 53
4.3.1 Généralités . . . 54
4.3.2 «Partitions» d'images idéales . . . 55
4.3.3 Partitions d'images numériques . . . 56
Frontières pixellaires et inter-pixellaires . . . 56
Graphe des frontières de régions (RBG) . . . 57
Graphes d'adjacence de régions (RAGs) . . . 57
RBAG . . . 58
4.4 Hiérarchies . . . 59
4.4.1 Définitions générales . . . 59
4.4.2 Père et fils, branches et infimum . . . 61
4.4.3 Hiérarchies et partitions - coupes . . . 61
Construction d'une hiérarchie à partir de partitions . . . 61
Construction de partitions à partir d'une hiérarchie - coupes . . . 62
4.4.4 Hiérarchies et pyramides régulières - Quad trees . . . 63
4.5 Espaces ultramétriques - hiérarchies indicées . . . 64
4.5.1 Distances et espaces ultramétriques . . . 64
4.5.2 Boules d'un espace ultramétrique et hiérarchies indicées . . . 64
4.5.3 Représentation graphique d'une hiérarchie . . . 65
4.5.4 Représentation et coupes naturelles d'une hiérarchie indicée . . . 66

Hiérarchies de cocons 71

5 Hiérarchies de cocons 71
5.1 Introduction . . . 71
5.2 Méthodes de partitionnement en régions homogènes . . . 71
5.2.1 Mesures d'hétérogénéité et prédicats d'homogénéité . . . 71
5.2.2 Consistance, croissance et maximalité . . . 72
5.2.3 Non unicité d'une solution . . . 73
5.2.4 Processus de construction de partitions . . . 73
5.2.5 Heuristiques de production d'une «bonne» solution . . . 74
5.2.6 Partitionnement par Quad Tree - Split and Merge . . . 74
5.2.7 Facteurs géométriques - procédures complexes . . . 75
5.2.8 Formulation orientée frontières . . . 75
5.3 Idées à l'origine de l'approche par cocons . . . 75
5.3.1 Principaux défauts des approches par prédicats . . . 75
5.3.2 Un problème d'ensembles . . . 76
5.3.3 Critères relatifs - contraste simultané . . . 76
5.4 Sous-graphes contrastés connexes (cocons) . . . 77
5.4.1 Définition . . . 77
Groupement sur un graphe attribué par une dissemblance . . . 77
Connexions, sous-graphes induits et cocycles . . . 78
Statistiques internes et externes . . . 79
Sous-graphes contrastés et cocons . . . 80
5.4.2 Interprétation . . . 81
5.4.3 Cocons triviaux . . . 81
5.4.4 Variantes : cocons complets et complets-complets . . . 82
5.5 La hiérarchie des cocons durs (max-min) . . . 83
5.5.1 Définition et interprétation . . . 83
5.5.2 Propriétés . . . 84
Hiérarchisation . . . 84
Distance ultramétrique . . . 85
Robustesse de la hiérarchie dure . . . 86
Un cas particulier : composantes connexes des sections d'une image . . . 87
5.5.3 Construction de la hiérarchie dure . . . 90
5.5.4 Résultats expérimentaux . . . 92
Mesures de dissemblances . . . 92
Exemples de cocons durs . . . 95
Partitionnement par coupes itérées de hiérarchies dures . . . 95
5.6 Hiérarchies de cocons équilibrés . . . 100
5.6.1 Familles de cocons non hiérarchisés . . . 100
5.6.2 f-Linkage et sous-ensembles hiérarchisés de f-cocons . . . 101
5.7 Relations avec d'autres méthodes de groupement . . . 104
5.7.1 Critère de contraste de Felzenswalb et Huttenlocher . . . 104
5.7.2 Critère de coupe normalisée de Shi et Malik . . . 105
5.7.3 Le problème de «comparaison avec le vide» . . . 106
5.8 Conclusion . . . 107

Analyse ensembles-échelle d'une image **111**

6 Présentation de l'approche **111**
6.1 Modélisation d'images par morceaux . . . 111
6.1.1 Recherche d'un modèle «optimal» . . . 112
Comparaison et unicité de solution . . . 112
Une question de compromis . . . 113
6.1.2 Formulations variationnelles . . . 114
6.1.3 Modélisations probabilistes . . . 115
Inférence bayesienne . . . 115
Champs de Markov . . . 115
6.1.4 Inférence par codage minimal . . . 117
Complexité de Kolmogorov . . . 118
Longueur de message ou de description . . . 119
Recherche de langages optimaux . . . 121
A priori et «objectivité» . . . 124
Critères de codage pour la segmentation . . . 126
6.1.5 Exploration de l'espace des partitions . . . 128
Optimisation combinatoire . . . 128
Recherche d'optimum local . . . 128
Recherche d'optimum global . . . 130
6.2 Modélisation multi-échelles d'images par morceaux . . . 131
6.2.1 Le cadre énergétique . . . 131
Énergies affines . . . 131
Un problème d'optimisation sous contrainte . . . 133
Compression avec pertes - théorie débit/distorsion . . . 134
Débit/distorsion et codage minimal - Régularisation par complexité . . . 135
Le principe d'indifférence d'échelle . . . 136
6.2.2 Le cadre structurel . . . 137
Causalité structurelle et non délocalisation de frontières . . . 137
Représentation ensembles-échelles d'une analyse multi-échelles . . . 138
6.2.3 Positionnement par rapport aux approches existantes . . . 140
Processus hiérarchiques ou multi-résolution . . . 140
Analyses hiérarchiques, pyramidales ou multi-échelles . . . 141
6.3 Description de la démarche . . . 144
6.3.1 Non monotonie a priori des minima d'une énergie affine . . . 144
6.3.2 Présentation de la démarche . . . 146

7 Minimisation d'énergies multi-échelles dans une hiérarchie **147**
7.1 Coupe minimale d'une hiérarchie . . . 147
7.1.1 Algorithme de programmation dynamique . . . 147
Discussion . . . 150
7.1.2 Modèles énergétiques séparables . . . 152
Fonctionnelle de Mumford-Shah . . . 152
Modèle séparable général . . . 152
Modèles non contextuels . . . 153
Deux degrés de séparabilité d'une énergie de frontières . . . 154
Invariance par focalisation . . . 156
7.1.3 L'espace des coupes d'une hierarchie . . . 158
Hiérarchies binaires parfaitement équilibrées . . . 158

Hiérarchies binaires totalement déséquilibrées . . . 160
Dimensions d'autres espaces de partitions . . . 160
Bilan comparatif . . . 163
7.1.4 Optimisation hiérarchique et descente de gradient . . . 164
7.2 Familles de coupes minimales monotones . . . 166
7.2.1 Condition de monotonie des coupes minimales . . . 167
Optimalité partielle et maximalité . . . 167
Sous-additivité d'énergies de parties et décroissance d'énergies de partitions . . . 167
Familles de solutions causales . . . 168
Apparition, disparition et persistance de régions . . . 169
7.2.2 Calcul des intervalles de persistance des régions . . . 170
Programmation dynamique fonctionnelle . . . 170
Hiérarchie persistante . . . 174
7.2.3 Discussion . . . 175
Sous-additivité et causalité . . . 175
L'échelle des échelles . . . 178
Version complémentaire pour les énergies sur-additives . . . 179
Extension multidimensionnelle . . . 180
7.2.4 Compression avec perte et minimisation sous contrainte . . . 181
Caricature visuelle optimale . . . 181
Résolution complète du problème de débit/distorsion . . . 182
7.3 Conclusion . . . 182

8 Escalade 185
8.1 Le principe d'escalade . . . 185
8.1.1 Complétions croissantes d'une pré-hiérarchie . . . 185
8.1.2 Escalade pure . . . 187
8.1.3 Escalade binaire et escalade binaire connexe . . . 188
8.2 Propriétés des hiérarchies d'escalade . . . 189
8.2.1 Invariance structurelle et covariance d'échelle . . . 189
Homothétie sur les énergies . . . 189
Transposition d'invariances . . . 191
8.2.2 Une solution unique aux deux problèmes duaux de débit/distorsion . . . 191
8.3 Interprétations du principe d'escalade . . . 193
8.3.1 Escalade et descentes . . . 193
8.3.2 Escalade et continuation . . . 194
8.3.3 Escalade et prédicats naturels . . . 194
8.3.4 Interprétation continue . . . 195
8.4 Formulation finale du problème de segmentation multi-échelles . . . 197
8.4.1 Principe d'indifférence d'échelle . . . 198
8.4.2 Minimiser la co-énergie totale d'une hiérarchie . . . 199
8.4.3 Invariance par homothétie des solutions . . . 200
8.4.4 L'escalade pure : une stratégie d'approximation localement optimale . . . 201
8.5 Persistance des régions . . . 203
8.5.1 Interprétation : cas de la fonctionnelle de Mumford-Shah . . . 204
8.6 Conclusion . . . 206

9 Résultats expérimentaux **209**
9.1 Modèle constant par morceaux de Mumford-Shah . . . 209
9.1.1 Représentation des résultats . . . 210
Dichotomie d'une description ensembles-échelle . . . 210
Ultramétrique et cartes de contours . . . 211
9.1.2 Résultats expérimentaux . . . 212
9.1.3 Étude de l'algorithme . . . 216
Influence de la sur-segmentation . . . 216
Escalade et bruit . . . 219
Résolution et échelle . . . 224
9.2 Énergies géométriques «orientées régions» . . . 227
9.2.1 Importance de la régularisation géométrique . . . 227
9.2.2 Énergies géométriques globales . . . 228
9.2.3 Nombre de traits d'une description graphique . . . 229
Forme non séparable . . . 229
Forme séparable . . . 229
9.2.4 Codage polaire de lignes polygonales . . . 230
Codage avec changements de direction équiprobables . . . 230
Codage avec changements de direction gaussiens . . . 232
9.2.5 Énergie de courbure/concavité . . . 233
Définition . . . 233
Variation de concavité par fusion . . . 235
9.2.6 Une variante séparable de l'énergie géométrique de Ballester et al. . . . 236
9.2.7 Résultats comparatifs . . . 237
9.3 Modèles radiométriques . . . 239
9.3.1 Codage des radiométries selon un modèle gaussien . . . 241
9.3.2 Entre trace et déterminant . . . 243
9.3.3 Résultats expérimentaux . . . 244
9.3.4 Critère MDL - débit/distorsion . . . 245
9.3.5 Introduction d'un terme de gradient . . . 247
9.4 Modélisation d'un MNE par facettes planes . . . 247
9.5 Conclusion . . . 250

10 Conclusion et perspectives **257**
10.1 Conclusion . . . 257
10.1.1 Ensembles et échelles . . . 257
10.1.2 Développer la dimension d'échelle . . . 258
10.2 Perspectives . . . 260
10.2.1 Perspectives théoriques . . . 260
10.2.2 Perspectives applicatives . . . 262
10.3 Conclusion finale . . . 264

Liste des communications **265**

Références bibliographiques **267**

Annexes **275**

A Sur-segmentation par ligne de partage des eaux **279**

B Algorithmes de cocons durs **283**
B.1 Calcul des cocons durs contenant un nœud donné . . . 283
B.2 Calcul de l'ultramétrique de cocons . . . 284
B.3 Calcul des cocons maximaux pour un critère d'homogénéité . . . 284
B.4 Complexité des algorithmes . . . 285

C Algorithme d'escalade **287**
C.1 Description de l'algorithme . . . 287
C.2 Complexité de l'algorithme . . . 292
C.2.1 Part structurelle de la complexité . . . 293
C.2.2 Coût des calculs d'énergie . . . 294
C.2.3 Complexité de la programmation dynamique fonctionnelle . . . 294
C.2.4 Optimisation par troncature des sommes d'affines par morceaux . . . 296
C.2.5 Bilan . . . 296

D Algorithme d'approximation polygonale **299**

Table des figures

2.1 Quatre exemples d'images 2D . . . 8
2.2 Grossissements successifs sur le port du Havre . . . 12
2.3 Cartes à différentes échelles . . . 14

3.1 Coupes dans le scale-space gaussien et détection de contours . . . 33
3.2 Régions perceptuellement pertinentes non hiérarchisées . . . 38

4.1 Infimum et supremum de deux partitions . . . 51
4.2 Treillis des parties d'un ensemble à quatre éléments . . . 53
4.3 Treillis des partitions d'un ensemble à quatre éléments. . . . 54
4.4 Graphe de frontières des régions d'une partition du plan . . . 55
4.5 Frontière discrète pixellaire et inter-pixellaire . . . 56
4.6 Graphes associés à des partitions d'images numériques 2D . . . 57
4.7 Dualité entre RBG et RAG géométrique . . . 58
4.8 Une région enclavée . . . 58
4.9 Fusion de régions dans un RBAG . . . 59
4.10 Parties hiérarchisées . . . 60
4.11 Représentation d'une hiérarchie sous forme de dendrogramme . . . 60
4.12 Branches et borne supérieure de deux nœuds d'une hiérarchie . . . 61
4.13 Construction d'une hiérarchie à partir d'une séquence de partitions . . . 62
4.14 Coupes dans une hiérarchie . . . 62
4.15 Une hiérarchie et une partition régulières . . . 64
4.16 Représentation et coupes naturelles d'une hiérarchie indicée . . . 67

5.1 Non unicité d'une segmentation en régions homogènes . . . 73
5.2 Sous graphe induit et cocycle . . . 79
5.3 Illustration de la définition de cocon dans le cas d'un graphe attribué par une distance . . . 80
5.4 Un cocon complet . . . 82
5.5 Arêtes considérées dans la démonstration de hiérarchisation des cocons durs . . . 84
5.6 Exemple de cocons durs et d'ultramétrique associée . . . 86
5.7 Ensembles de niveau ou sections d'une fonction . . . 87
5.8 Un pas de croissance d'un cocon dur (max-min) . . . 90
5.9 Une chaîne de cocons durs emboîtés . . . 96
5.10 Exemples de cocons durs . . . 97
5.11 Cocons homogènes itérés (1/2) . . . 98
5.12 Cocons homogènes itérés (2/2) . . . 99
5.13 Exemples de segmentation par cocons maximaux itérés . . . 100
5.14 Non hiérarchisation a priori des familles de cocons moyens et s'appuyant sur des quantiles 101

6.1 Une description ensembles-échelle et certaines de ses coupes 140
6.2 Exemple de non monotonie de partitions minimales en fonction de l'échelle 145

7.1 Hiérarchie partielle 148
7.2 Un pas de programmation dynamique pour le calcul d'une coupe minimale 149
7.3 Modèles énergétiques contextuels et non-contextuels 153
7.4 Les énergies locales non additives ne conduisent pas à des énergies de partitions séparables 155
7.5 Restriction sans occultation d'une partition à un domaine 157
7.6 Arbre binaire parfaitement équilibré BE_8 158
7.7 Arbre binaire totalement déséquilibré BD_8 160
7.8 Graphes considérés pour la programmation dynamique hiérarchique et pour le partitionnement d'un hypergraphe d'intervalles 162
7.9 Une 2-3-hiérarchie et le graphe de couverture correspondant 163
7.10 Séparation entre critère de fusion et critère de coupe 165
7.11 Calcul de l'énergie partielle E_c^* et de l'échelle d'apparition $\lambda^+(c)$ d'un nœud c 173
7.12 Les deux étapes de l'optimisation d'une énergie multi-échelles dans une hiérarchie 175
7.13 La représentation ensembles-échelle transforme la famille des coupes minimales d'une hiérarchie en une famille de droites horizontales 176
7.14 Fonction d'énergie des λ-coupes 177
7.15 Extension multidimensionnelle de l'optimisation hiérarchique 180
7.16 Résolution exacte du problème de débit/distorsion dans une hiérarchie 183

8.1 Une pré-hiérarchie 186
8.2 Union interne de deux hiérarchies disjointes 187
8.3 Stratégie d'escalade pure 187
8.4 Escalade binaire connexe 189
8.5 Escalade et descente 193
8.6 Escalade continue 196
8.7 Énergie et co-énergie totale d'une hiérarchie 198
8.8 La hiérarchie d'énergie minimale est toujours creuse (a) mais celle de co-énergie maximale ne l'est généralement pas (b) 199
8.9 Gain local net de co-énergie par escalade 202
8.10 Un objet o s'inscrit dans un objet plus global g, possède un environnement o, e_1, e_2, e_3 et des détails internes i_1, i_2, i_3 204
8.11 Sur et sous-structures d'un objet 205

9.1 Algorithme LQR_1 : dichotomies des ensembles-échelle 213
9.2 Algorithme LQR_1 : ensembles-échelle 214
9.3 Algorithme LQR_1 : ultramétriques / cartes de contours 215
9.4 Influence de la sur-segmentation : comparaison entre LQR_1, LQR_4 et LQW_2 (1/2) 217
9.5 Influence de la sur-segmentation : comparaison entre LQR_1, LQR_4 et LQW_2 (2/2) 218
9.6 Comparaison entre LQR_1 et LQW_2 sur l'image chromosomes 219
9.7 Comportement de l'algorithme LQR_1 sur une image bruitée par permutations aléatoires . 220
9.8 Comportement de l'algorithme LQR_1 sur une image bruitée par un bruit gaussien i.i.d. . . 223
9.9 Changement d'échelle et de résolution (1/2) 225
9.10 Changement d'échelle et de résolution (2/2) 226
9.11 Importance de la régularisation géométrique 227
9.12 L'énergie géométrique T_0 non séparable n'est pas monotone 230
9.13 Variation de l'énergie géométrique T_0 par fusion de couples de régions (cas de synthèse) . 231
9.14 Variation de l'énergie géométrique T_1 par fusion de couples de régions (exemple réel) . . . 232

9.15 Changements de direction le long d'une ligne polygonale 233
9.16 Exemples de valeurs d'énergie de concavité normalisée 234
9.17 Variation d'énergie de concavité par fusion 235
9.18 Comparaison des énergies géométriques sur l'image girl 238
9.19 Comparaison des énergies géométriques sur l'image champs 240
9.20 Comparaison des énergies radiométriques G, $\mathsf{Q}^{\frac{1}{2}}$, Q, $\mathsf{D}^{\frac{1}{2}}$ et D sur Lena 244
9.21 Escalades sur des critères de codage minimal 246
9.22 Introduction d'un terme de gradient dans l'énergie de la partition 248
9.23 Segmentation d'un MNE en facettes planes 249
9.24 Hangar en construction 251
9.25 Une ferme 251
9.26 Paysage de Beauce 252
9.27 Un bourg de campagne 253
9.28 Paysage rural 254
9.29 La tour Eiffel 255

10.1 Le couplage d'échelle peut-il induire des influences contextuelles et des phénomènes de formes illusoires ? 261
10.2 Un alignement de taches 262

A.1 Sur-segmentation par ligne de partage des eaux 281

C.1 Construction d'une hiérarchie d'escalade 288
C.2 Croissance du nombre de morceaux des fonctions affines lors d'une programmation dynamique fonctionnelle 295

Liste des tableaux

5.1 Critères de qualité de Johnson pour les algorithmes de groupement par similarité 87

6.1 Différentes interprétations d'une énergie affine 132

7.1 Tailles de différents espaces de partitions 164

8.1 Résumé des modèles multi-échelles étudiés 207

9.1 Algorithme LQR_1 : caractéristiques des images considérées, échelle d'apparition de l'image et temps de calcul 212
9.2 Comparaison entre LQR_1 et LQW_2 sur l'image chromosomes 218
9.3 Comportement de l'algorithme LQR_1 sur une image bruitée par permutations aléatoires . 221
9.4 Comportement de l'algorithme LQR_1 sur une image bruitée par un bruit gaussien i.i.d. . . 224
9.5 Résumé des énergies géométriques considérées 237
9.6 Résumé des énergies radiométriques considérées 243

C.1 Complexité combinatoire d'une escalade (bilan) 297

Rendre clair l'obscur et simple le complexe

Abu Ja'far Muhammad ibn Musa al-Khwarizmi,
Kitab al-muhtasar Hisab al-Jabr w'al-Muqabala,
Bagdad, début du IXème siècle.

Chapitre 1

Introduction

Ce mémoire décrit le résultat de recherches que nous avons menées au laboratoire MATIS[1] de l'Institut Géographique National (IGN) français.

Ces recherches ont été co-dirigées par Hervé Le Men, directeur technique adjoint de l'IGN, et Jean-Pierre Cocquerez, professeur à l'université de Cergy-Pontoise et directeur de l'Équipe de Traitement des Images et du Signal (ETIS) de l'ENSEA[2] au moment où cette thèse a commencé, et aujourd'hui directeur du département génie informatique de l'Université de Technologie de Compiègne (UTC).

Cette thèse porte sur la **segmentation d'images en régions**.

La segmentation est une étape algorithmique fondamentale de nombreux systèmes d'analyse automatique d'images. Son rôle est de délimiter dans l'image étudiée un ensemble de zones «pertinentes» pour l'interprétation ou la modélisation de la scène perçue. Il s'agit donc d'une étape-clé, et sa qualité conditionne fortement la réussite globale d'une entreprise d'analyse automatique d'image.

Nos recherches sur ce sujet ont été initialement motivées par une problématique d'application intéressant l'IGN, à savoir la mise à jour automatique de cartes d'occupation du sol à partir d'images aériennes. Notre travail s'est cependant petit à petit cristallisé autour du problème de segmentation d'image. En effet, telle que nous l'avons envisagée, la résolution de ce problème d'interprétation d'image nécessitait que l'on dispose d'un outil de segmentation d'images aériennes à la fois fiable et efficace, applicable à l'échelle d'un territoire observé à une résolution de l'ordre du mètre.

Or, nos expériences en segmentation nous ont montré qu'il était difficile d'obtenir un tel résultat en utilisant un algorithme «classique» de segmentation, ayant pour résultat une *partition* de l'image en régions. Plus précisément, le *paramétrage* du niveau de segmentation approprié à l'extraction du parcellaire agricole semblait un problème inextricable. Certains réglages permettaient d'extraire certaines parcelles très homogènes mais en sur-découpaient d'autres plus texturées. Par ailleurs, en relâchant les paramètres de l'algorithme, on pouvait obtenir les parcelles texturées, mais certaines zones faiblement contrastées par rapport à leur environnement étaient alors perdues. Il paraissait impossible d'obtenir une segmentation correcte du parcellaire agricole sur l'ensemble d'une scène aérienne *grâce à un réglage unique d'un algorithme de partitionnement*. En outre, que les conditions d'observation changent et les résultats s'en trouvaient modifiés. Par contre, une description correcte du paysage agricole semblait pouvoir systématiquement être composée *en mélangeant des objets obtenus pour différents réglages*.

Ces observations nous ont progressivement conduit à l'idée que ce réglage pouvait être relié à une notion d'échelle de description et que celle-ci ne pouvait pas être déterminée à bas niveau, par la méthode de segmentation. Nous aboutissons aujourd'hui à la conclusion que, pour pouvoir constituer un socle fiable pour un système d'interprétation d'images un tant soit peu complexes, une analyse en régions doit être conçue sous forme multi-échelles, c'est-à-dire fournir une description à différents degrés de détail de

[1] Méthodes d'Analyse et de Traitement d'Image pour la Stéréorestitution
[2] École Nationale Supérieure de l'Électronique et de ses Applications

l'image. Charge à chaque processus aval d'interprétation spécifique de trouver ses propres objets d'intérêt dans cette description.

La concentration de nos efforts sur le problème de segmentation nous a quelque peu éloignés de l'application qui était envisagée au départ. Un certain nombre de travaux lui ont toutefois été consacrés : appariement multivoque de régions de partitions, géoréférencement automatique de plans cadastraux, classification des zones de végétation arborée, pour ne citer que les recherches qui ont donné lieu à des communications. Ces travaux ne sont pas repris ici.

Nos recherches sur la segmentation elle-même ont en effet abouti à une méthodologie relativement générale d'analyse d'images en régions, que nous avons appelé la méthodologie d'analyse **ensembles-échelle** (scale-sets) d'une image. Le nom provient de la structure du résultat, qui constitue le pendant ensembliste d'une analyse espace-échelle (scale-space) issue du domaine du filtrage. Le modèle final possède de nombreux attraits à la fois théoriques et pratiques, et ouvre des perspectives applicatives qui dépassent largement le champ de la cartographie automatique. Certaines d'entre elles sont d'ailleurs relativement inattendues (compression, filtrage, caricature d'images). Aussi avons-nous finalement jugé bon de présenter ces travaux dans une perspective plus large que le contexte industriel spécifique qui les avait motivés. La plupart des exemples proviendra toutefois du domaine de l'imagerie aérienne et de la cartographie.

Ce document est divisé en trois parties. La première partie présente les idées générales qui sous-tendent notre approche de la segmentation. Les deux parties suivantes développent deux modèles de segmentation multi-échelles que nous avons successivement élaborés.

Le premier chapitre commence par examiner les différentes notions d'*image* (abstraction mathématique, collection de mesures, structure informatique, objet graphique...) et revoit deux notions qui sont étroitement liées : celle de *résolution* - caractéristique d'une opération de mesure, c'est-à-dire d'*acquisition* d'information - et celle d'*échelle* - caractéristique d'une opération de *description* d'information.

Le deuxième chapitre aborde la problématique de segmentation d'image et envisage la question sous un angle systémique : quel est le rôle d'une opération de segmentation dans une chaîne d'analyse d'images ? quelles doivent être ses caractéristiques externes (entrées-sorties, comportement qualitatif...) ? Dans la lignée de David Marr, nous proposons d'envisager la segmentation comme une opération de bas niveau dont l'objectif est d'élaborer une «première ébauche brute» (raw primal sketch) du contenu visuel de la scène, une *description* devant servir de base aux mécanismes ultérieurs d'interprétation. Empruntant également différentes idées aux théories de la Gestalt et du scale-space, nous argumentons alors que, pour pouvoir constituer une couche relativement générique et robuste d'analyse de bas niveau, une segmentation *doit* prendre la forme d'une description multi-échelles des données d'entrée, modélisant différents niveaux de perception possibles de la scène.

Pour conclure la partie, le troisième chapitre propose quelques rappels sur différentes représentations mathématiques et informatiques qui permettent de modéliser et de manipuler des familles d'ensembles structurés (graphes, treillis, partitions, hiérarchies, espaces ultramétriques...) et qui nous seront utiles par la suite.

La deuxième partie du document décrit une première approche de l'analyse en régions, qui s'appuie sur la définition de certains sous-graphes particuliers d'un graphe attribué, que nous avons baptisés **cocons** (pour *sous-graphes contrastés connexes*).

Une réflexion sur la nature des critères de segmentation nous amène à proposer qu'un critère «objectif» de décision (grosso-modo sans paramètre) quand à la «pertinence» d'une région doit être de nature *relative*, c'est-à-dire reposer sur une comparaison entre une région et son environnement. Formalisée dans le cadre de la théorie des graphes, cette idée nous conduit à la définition des *cocons durs*, dont nous montrons qu'ils s'organisent en *hiérarchies*. Le chapitre étudie en détails cette première famille de cocons.

Nous voyons en particulier que les hiérarchies de cocons durs correspondent à des sous-hiérarchies des hiérarchies obtenues par un algorithme classique de groupement - le groupement par lien complet - le critère de sélection des cocons reposant sur une notion de *maximalité locale* dans la hiérarchie. Les premiers algorithmes de segmentation d'image proposés s'appuient alors sur des *coupes* itérées dans des hiérarchies de cocons durs construites à partir d'une sur-segmentation de l'image par un algorithme de ligne de partage des eaux. Nous étudions ensuite d'autres algorithmes de la même famille que l'algorithme du lien complet, dont nous montrons qu'ils conduisent à d'autres types de cocons, que nous avons appelés *cocons équilibrés*.

La troisième partie du document (chapitres 6 à 9) développe ce que nous avons baptisé l'approche **ensembles-échelle** de l'analyse d'image en régions, et qui constitue le cœur de cette thèse.

Dans la lignée des approches variationnelles, bayésiennes ou par codage minimal de la segmentation, nous abordons l'analyse en régions comme un problème d'inférence d'un modèle de l'image, modèle défini par parties - i.e. s'appuyant sur une *partition* du domaine de l'image - et nous envisageons ce problème d'inférence comme une question de minimisation d'énergie sur l'espace des modèles considérés. Toutes les théories citées ci-dessus s'accordent sur le fait que le problème général de modélisation relève intrinsèquement d'un compromis entre deux facteurs : un facteur de «fidélité» du modèle aux données et un facteur de «simplicité» du modèle. Les énergies obtenues s'écrivent alors systématiquement comme une combinaison linéaire de deux termes énergétiques en opposition ; l'importance relative de ces deux termes, et donc l'optimum du problème, étant contrôlée par un paramètre réel, que nous notons λ.

Nous proposons alors que ce paramètre libre apparaissant dans les énergies ne correspond pas à une insuffisance des formulations mathématiques mais traduit l'existence d'un facteur inhérent à toute question de modélisation (ou de description) d'observations : le *paramètre d'échelle*. Dans le cadre de l'analyse d'images, pour que le paramètre λ se comporte comme une échelle, il faut alors que les partitions qui lui correspondent forment une séquence monotone : augmenter λ doit produire des descriptions de plus en plus grossières d'une image. Nous formulons donc le problème de segmentation multi-échelles comme celui de trouver une séquence de partitions emboîtées qui soit - dans un certain sens - aussi proche que possibles des minima d'une énergie à deux termes antagonistes. Argumentant alors, dans l'esprit scale-space, qu'aucune échelle de description n'est *a priori* meilleure qu'une autre, nous nous intéressons non pas à une seule solution du problème, mais à la *famille complète de ses solutions*, famille indicée par λ. Nous voyons en particulier qu'envisagée comme un tout, cette famille correspond également aux solutions de deux problèmes duaux de débit/distorsion opérationnels qui formalisent des questions de compression avec perte orientée régions.

Un élément clef de l'approche est alors une *représentation*, que nous avons baptisée la **représentation ensembles-échelles** des solutions du problème considéré. En effet, si les solutions forment une séquence de partition monotones alors elles peuvent être représentées et manipulées *en intégralité*, mais il faut pour cela les envisager de façon *implicite*, comme des *coupes* dans une structure de dimension supérieure (une hiérarchie indicée / un espace ultramétrique). A l'instar d'une description scale-space d'une fonction, une description ensembles-échelle possède un axe d'échelle *continu*. Elle permet - entre autres - d'obtenir instantanément toute solution particulière, à échelle fixe.

Cette démarche générale est exposée au chapitre 6. Les chapitres 7 et 8 étudient alors la minimisation hiérarchique d'une classe d'énergies relativement générale, qui englobe de nombreuses énergies habituellement considérées en segmentation. Nous aboutissons finalement au chapitre 8 à un principe d'optimisation, baptisé principe d'**escalade**, qui permet, pour une image et un couple d'énergies en opposition donnés, de construire une description ensembles-échelles remarquable de l'image associée aux énergies. La solution obtenue vérifie plusieurs propriétés théoriques, la plus importante étant une propriété de *covariance d'échelle* : régler λ a priori - dans la définition des énergies - ou a posteriori - par coupe dans la représentation ensembles-échelle obtenue - est strictement équivalent. *Le paramètre d'échelle est donc véritablement éliminé de l'étape d'analyse de bas niveau* et devient un moyen d'exploration au sens fort, un potentiomètre mis à disposition de l'utilisateur (homme ou machine) pour naviguer à travers

différents niveaux de descriptions d'une scène. D'un point de vue pratique, le principe d'escalade conduit à un algorithme *efficace et totalement exempt de paramètre* une fois les énergies spécifiées.

Cette méthodologie générale peut être déclinée pour toutes sortes de problèmes d'analyse d'images en régions via la spécification d'énergies adaptées à chaque type de problème (segmentation, compression) et d'images (photométriques, de distance...). Le chapitre 9 illustre la pertinence de la méthode en présentant des résultats expérimentaux obtenus pour différents types d'énergies et d'images. Il étudie en détail le comportement de l'algorithme et décrit différentes applications possibles des descriptions ensembles-échelle.

Après la conclusion, trois annexes détaillent les algorithmes proposés et étudient leur complexité.

Questions de représentations

Chapitre 2

Images et représentations graphiques

«J'appelle images[1], écrit Platon dans *La République*, d'abord les ombres, ensuite les reflets qu'on voit dans les eaux ou à la surface des corps opaques, polis et brillants et toutes les représentations de ce genre».

Une «image» est avant tout une représentation : elle est l'«image» *de* quelque chose. Ainsi, après la définition optique, le Petit Robert (ed. 1994) donne la définition courante de «représentation d'un objet par les arts graphiques ou plastiques, ou par la photographie». Dans le même esprit, Haralick et Shapiro (1991) entament leur *glossaire des termes de vision par ordinateur* en écrivant qu'«une image est une représentation spatiale d'un objet, d'une scène à deux ou trois dimensions, ou d'une autre image».

L'image est figurative. C'est ce qui conduit Marc Thivolet à qualifier l'art abstrait de «négation de l'image»[2]. Dans d'autres registres que le registre visuel, l'image du poète est inséparable de l'idée dont elle est la métaphore, l'image du mathématicien n'a pas de sens sans son antécédent...

Ce que représentent les images visuelles, ou plutôt ce que nous nous représentons qu'elles représentent, fera l'objet du prochain chapitre. Dans un premier temps, ce chapitre examine ce qu'elles *sont*.

2.1 Les notions d'image

2.1.1 Images mathématiques

Pour un mathématicien, une image est simplement la donnée des valeurs d'une fonction $I : \mathcal{D} \to \mathcal{V}$, c'est-à-dire la donnée d'un point de $\mathcal{V}^{\mathcal{D}}$.

$\mathcal{D}$ est une partie substantielle d'un espace généralement métrique, au moins topologique. $\mathcal{V}$ est la plupart du temps également métrique. Dans tout de ce document, à $\mathcal{D}$ et $\mathcal{V}$ fixés, nous noterons $\mathcal{I}$ l'ensemble des images, i.e. l'ensemble des fonctions de $\mathcal{D}$ dans $\mathcal{V}$. $\mathcal{D}$ sera appelé le support des images et $\mathcal{V}$ leur domaine de valeur.

Nous appellerons images «idéales» les images abstraites définies sur un support continu, typiquement sur une partie de $\mathbb{R}^n$, par opposition aux images concrètes, ou physiques, que nous aurons à manipuler, qui s'appuient sur un support discret, typiquement sur une partie de $\mathbb{Z}^n$.

2.1.2 Collections de mesures physiques

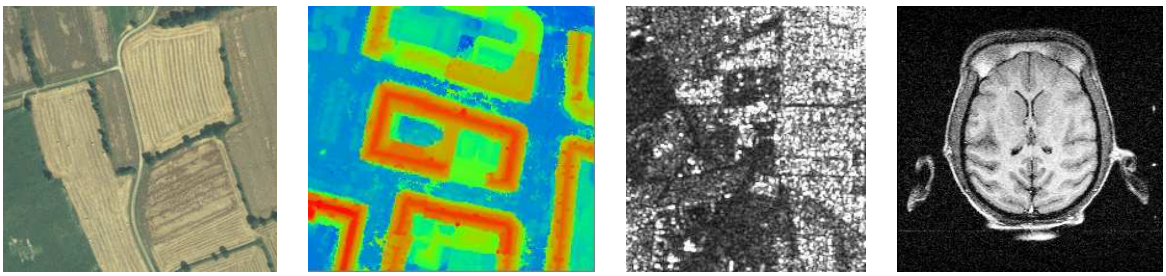
Pour un physicien, une image est une collection de mesures d'une grandeur physique réalisées selon une trame spatiale déterminée. Cette définition inclut les images au sens classique du terme - les images «visuelles» du type de celles acquises par notre rétine, un film photographique ou une matrice CCD - mais ne s'y restreint pas.

[1] εἰκόνας, qui a donné *icône* en français.
[2] Article *image* de l'Encyclopedia Universalis, ed. 1985.

Il y a toutes sortes de techniques d'imagerie. Les procédés passifs captent un rayonnement émis par une source externe. Les procédés actifs émettent une onde et enregistrent son écho (laser, radar, ultrasons, rayons X ...) ou mesurent une réaction particulière du milieu à l'excitation fournie (Imagerie par Résonance Magnétique (IRM), Tomographie par Émission de Positons (TEP), microscopie par balayage électronique...). Les observations consignées dans une image peuvent donc être de différentes natures : intensité lumineuse reçue dans une ou plusieurs bandes spectrales pendant un intervalle de temps donné, temps de retour d'une onde (et par conséquent distance aux surfaces réfléchissantes), déphasage, décalage doppler (et donc vitesse radiale des objets), etc. Notons que l'information que l'on cherche à acquérir n'est pas toujours la mesure brute effectuée par le capteur, mais peut nécessiter d'importants calculs (e.g. l'inversion d'une transformée de Radon en tomographie).

Par extension, une image est n'importe quel «crible» de mesures (pas nécessairement de nature électromagnétique). Ce qui caractérise une image est que le crible est réalisé selon une trame prédéterminée, fixée *a priori*. Une image peut s'envisager comme un ensemble de pixels[3], qui sont des couples *(site, valeur)* mais dont seule la valeur correspond à une observation. On ne qualifie par exemple pas d'image le relevé des couples *(lieu de résidence, âge)* dans la population d'une entreprise donnée, même si le lieu de résidence est une information de nature spatiale. En outre, dans sa définition habituelle, une image ne présente qu'une valeur par site, éventuellement composite (e.g. intensité de lumière rouge, verte et bleue) mais dont la composition est également prédéterminée.

Il existe donc de nombreux types d'images physiques. La figure 2.1 donne quatre exemples d'images acquises selon des modalités différentes (la deuxième est le résultat d'un algorithme mais le même type d'image de distance s'obtient par balayage laser).



(a) (b) (c) (d)

FIG. 2.1 – Quatre exemples d'images 2D. (a) Photographie aérienne d'un paysage rural. (b) Modèle Numérique d'Élévation (MNE) de bâtiments de la ville d'Amiens, calculé par un algorithme de stéréo-restitution automatique, i.e. qui estime les distances des objets à partir de plusieurs vues prises sous des angles différents. Les couleurs représentent les altitudes des points (les distances ont été converties en altitudes cartographiques), qui vont en croissant du bleu au rouge. (c) Image radar d'un quartier de la ville de Mexico prise par le satellite ERS-1 (outre l'intensité qui est représentée ici, un écho radar comporte une information de phase). (d) Coupe d'une IRM d'une tête humaine. L'IRM complète est en 3D.

2.1.3 Images numériques

Pour un informaticien, *image* est un type, une structure de données, qu'il appelle aussi structure de données «maillées» (ou «raster»).

Les images usuelles (celles qui peuvent être lues par un logiciel du commerce) s'appuient sur un support $\mathcal{D}$ dont la topologie est un maillage carré 2D régulier, du type . On assimile alors habituellement ce maillage à un rectangle de $\mathbb{Z}^2$: $\mathcal{D} = \langle 0, C\langle \times \langle 0, L\langle$[4].

[3]abbréviation de l'anglais «picture element».

[4]Nous notons entre chevrons les intervalles de $\mathbb{Z}$ (ou $\overline{\mathbb{Z}}$), i.e. $\langle n, m\rangle = \{n, n+1, \ldots, m\}$ et $\rangle n, m\langle = \{n+1, \ldots, m-1\}$.

Comme toutes les données manipulées par un ordinateur, les images informatiques sont numériques, ce qui signifie que leurs valeurs sont des *codes*. Dans ce sens (celui de l'anglais digital), numérique s'oppose à analogique et est synonyme de... symbolique ! Les codes contenus dans une image peuvent référencer des nombres (entiers, flottants, complexes...) mais pas nécessairement. Les valeurs d'une image d'étiquettes (ou image de «labels») peuvent ainsi renvoyer à des classes sémantiques.

La conséquence de ce codage est que le domaine de valeur $\mathcal{V}$ d'une image numérique est nécessairement fini (et donc discret). Les images usuelles sont à valeur dans $\langle 0, 256\langle$ pour les images en noir et blanc (niveaux de gris) ou dans $\langle 0, 256\langle^3$ pour les images en couleurs, à trois canaux rouge, vert et bleu.

Une image numérique n'est pas nécessairement une image de mesures physiques. Elle peut être le résultat d'un traitement informatique appliqué à une autre image, ou à d'autres types de données (e.g. image de synthèse). La transformée de Fourier d'une image est ainsi une autre image de mêmes dimensions mais de nature différente, son spectre. Les images issues de traitement peuvent alors être à valeur dans des espaces de grandes dimensions. Un exemple particulièrement frappant est le résultat de l'analyse texturale d'une image par une banque de filtres. L'image à analyser est traitée par un grand nombre n d'opérateurs locaux (plusieurs dizaines ou centaines) répondant chacun à la présence de structures particulières (e.g. à la présence d'une ligne d'épaisseur donnée dans une direction particulière). Chaque pixel de l'image finale porte un vecteur de $\mathbb{R}^n$ constitué des réponses ponctuelles à chacun des filtres. Dans le même ordre d'idées, certains traitements peuvent produire des images dont les pixels sont localisés dans un espace de grande dimension (e.g. l'application d'une transformée de Hough (1962) vers l'espace des paramètres d'une certaine famille de formes, voir (Maître 1985)).

2.2 Deux procédés inverses

2.2.1 Dématérialisation...

On présente habituellement les images numériques issues d'un processus de mesure physique comme résultant d'un échantillonnage suivi d'une quantification des échantillons d'une image doublement idéale, i.e. à support et domaine de valeur continus. Faisons deux remarques à ce propos.

Tout d'abord l'échantillonnage n'est pas une conséquence de la nature «numérique» des images, mais provient de la nature même de l'opération de mesure physique[5]. Les images analogiques, par exemple acquises par photographie argentique, sont également échantillonnées. Quand on numérise (on «scanne») une image analogique, comme une photographie, on procède à un rééchantillonnage, qui est limité par celui d'origine (on réalise l'image d'une image de même que notre oeil réalise l'image d'une image quand nous regardons un dessin ou une photographie).

En deuxième lieu, l'idée que l'étape de quantification correspond à une dégradation d'une mesure continue peut être fausse. Les capteurs numériques modernes sont en effet capables de compter les photons un à un. La quantification de l'image correspond alors à une réalité physique - la quantification de la lumière - et non à un artéfact, à une dégradation liée au processus de numérisation. L'IGN construit ainsi des caméras numériques s'appuyant sur des matrices CCD extrêmement sensibles qui réalisent quasiment des décomptes photoniques (dans les faibles intensités, le bruit de photon - poissonien - est alors prépondérant par rapport au «bruit» de capteur, i.e. à ses erreurs de mesure). Les images initialement acquises sur 12 bits (4096 niveaux) sont par la suite quantifiées sur 8 bits (256 niveaux) pour les utilisations habituelles, mais cette quantification peut être réalisée de manière «optimale».

Intéressons-nous plus particulièrement ici à l'échantillonnage et aux différents emplois du terme de *résolution*, qui lui est associé.

[5] L'idée de continuum est une abstraction purement mathématique.

Échantillonnage et résolution

Le Petit Robert (ed. 1994) indique que la **résolution** est l'«aptitude à mesurer une petite valeur d'une grandeur physique», puis définit la **limite de résolution** comme la «différence minimale perceptible entre deux valeurs d'une grandeur mesurée par un instrument». Ainsi, «la limite de résolution d'une lunette ou d'un télescope est un angle, celle d'un microscope est une longueur».

Nous confondrons pour notre part l'«aptitude» à mesurer des détails - la résolution - avec la taille des détails qu'un capteur est «apte» à mesurer - la limite de résolution.

En image, résolution est synonyme de précision d'analyse spatiale et correspond donc au pas d'échantillonnage du traitement du signal. On sous-entend alors souvent que l'échantillonnage est régulier et donc que ce pas est constant pour toute l'image. Pour certaines trames d'échantillonnage la résolution est cependant susceptible de varier sur l'étendue du domaine (notre rétine échantillonne par exemple plus finement les images au centre, dans la zone de la fovea).

Pour un dispositif optique à projection cônique (une caméra sténopé idéale), la limite de résolution a été définie ci-dessus comme un angle. C'est bien naturel : en perspective cônique, la taille apparente d'un objet diminue avec sa distance, par contre si l'échantillonnage du plan de projection de l'image optique est régulier, l'échantillonnage des directions spatiales est à peu près régulier[6].

Remarquons alors qu'une limite de résolution angulaire est une caractéristique intrinsèque au capteur alors qu'une limite de résolution spatiale correspond à une distance mesurée *sur l'objet observé*. Revenons donc sur l'exemple du microscope que donne le dictionnaire. Un microscope réalise également une projection cônique et pourtant sa résolution est définie comme une longueur. On connait en effet pour cet instrument la distance des objets observés - ils ont été disposés sur une plaque à cet effet - et une résolution angulaire correspond dans ce cas à une résolution spatiale à la surface de la plaque. Un microscope est alors conçu de manière à ce que cette résolution spatiale soit très proche d'une constante sur le champ de l'image.

En bref, dès que l'on peut, on exprime la résolution d'un instrument en terme de précision d'analyse spatiale des *objets observés* qu'il permet.

Pour un dispositif sténopé, ce n'est possible que si l'on dispose d'informations sur les distances potentielles des surfaces réfléchissantes, et cela n'a de sens d'indiquer une résolution spatiale, par exemple moyenne, que si les écarts possibles par rapport à cette valeur sont faibles. Il faut donc que les rapports entre les différentes distances susceptibles d'être observées soient proches de 1. On n'est absolument pas dans ce cas en observation astronomique, c'est pourquoi on indique des résolutions angulaires. On indique par contre des **résolutions au sol** en observation de la Terre par avion ou par satellite (on parle aussi de «taille de pixel sol»). Ce sont alors des résolutions moyennes (ou au centre du cliché).

Si l'on connait précisément les conditions de prise de vue et le relief du terrain photographié on peut transformer l'image de manière à ce que la résolution au sol soit identique pour tous les pixels, ce qui revient à annuler l'effet de perspective ou à simuler une image dans laquelle chaque point aurait été vu depuis sa verticale. On appelle l'image ainsi corrigée une ortho-photographie. Une ortho-photographie est une représentation topographique : on peut y mesurer des distances terrain et la superposer à une carte, voir (Boldo 2002).

Un informaticien parle également de résolution pour la numérisation de documents (le «scannage»), e.g. parle de numérisation à une résolution de 600 points par pouce (dot per inch, dpi). C'est dans ce cas l'inverse d'une résolution qu'il indique...

Un informaticien parle encore de résolution pour désigner le nombre pixels que contient une image ou le nombre de points affichés par un écran (e.g. 800×600). Pour éviter toute confusion, nous préférons employer le terme de **définition**, emprunté au domaine de la télévision numérique.

[6] Pour obtenir la meilleure régularité, il faut effectuer un échantillonnage polaire régulier sur une surface de projection sphérique. Notre oeil, qui adopte un tel procédé, place le pôle au milieu de la rétine, dans la ligne de visée de l'oeil.

2.2.2 ... et rematérialisation

Toutes les images numériques peuvent être visualisées, c'est-à-dire transformées en lumière visible ou en un objet pigmenté qui, éclairé, stimule sélectivement notre sens de la vue. Cette «matérialisation»[7] emploie un procédé d'impression, de projection, d'émission cathodique...

La visualisation d'une image passe par la définition de deux fonctions :

i) Une «projection» P du support $\mathcal{D}$ de l'image sur le support matériel choisi (papier, écran).

ii) Une «coloration» C qui associe à chaque valeur de l'image une «couleur» de représentation (quantité d'encre imprimée, de lumière rouge, verte ou bleue émise).

Soulignons qu'en elle-mêmes, les images numériques ne constituent ni des objets matériels, ni des objets visuels. Il ne s'agit que d'«information» et chaque image peut donner lieu à une infinité de rendus.

Si l'image est une collection de mesures physiques, en théorie, «restituer l'image» consisterait à reproduire le stimulus qu'a reçu le capteur au moment de l'acquisition, c'est-à-dire à inverser la fonction de transfert du système et resynthétiser un phénomène physique identique à celui qui a été enregistré. On voit que d'une part cela n'a pas de sens pour la plupart des procédés actifs, et que d'autre part nous ne verrions rien si l'instrument n'est pas un capteur de lumière visible. Bref, visualiser une image a pour seul but représenter le phénomène enregistré sous une forme qui nous est perceptible, afin que nous puissions l'interpréter, le «voir»...

Coloration

Les images numériques usuelles (fichiers tif, jpeg...) sont affichées par un ordinateur avec une coloration *par défaut* - une image à un canal est rendue par une émission lumineuse achromatique d'intensité croissante avec la valeur, une image à trois canaux par trois émissions colorées. Si l'image est une photographie usuelle, cette coloration par défaut pourra donner l'impression souhaitée : quelque chose qui ressemble à ce que nous voyons quand nous ouvrons les yeux. Mais encore une fois, il ne s'agit que d'un rendu possible de l'image : jouer avec la molette de luminosité de l'écran change par exemple l'objet visuel qui est produit par l'écran mais l'image sous-jacente reste la même. Il est ainsi souvent difficile de reproduire sur papier ce que l'on perçoit sur un écran. Quand les valeurs de l'image ne correspondent pas à des mesures lumineuses, comme pour les trois dernières images de la figure 2.1, le rendu a pour but principal de rendre sensible l'ordre des valeurs[8], ce qui peut se faire par une convention de couleurs relativement arbitraire (cf. e.g. 2.1(b)).

Nous venons de discuter de la coloration des images, intéressons-nous maintenant à leur rendu spatial, à leur projection.

Projection et échelle de représentation

Les images 2D usuelles s'appuient sur un maillage carré auquel les ordinateurs attribuent *par défaut* la géométrie d'un rectangle de $\mathbb{Z}^2$. Ils projettent donc ce rectangle sur un plan (écran, papier) en un autre rectangle (il s'agit encore une fois d'un choix fait par défaut). Outre le placement de l'image (translation et éventuellement rotation), le deuxième facteur de rendu est donc la taille de la représentation finale.

Pour un tirage papier, un informaticien parle à nouveau de résolution d'impression (e.g. à 600 dpi) pour ce qui est l'inverse d'une résolution (cette terminologie est toutefois consistante avec celle employée pour la numérisation). Sur un écran, les images sont injectées point pour point sur la trame de l'écran,

[7] C'est le procédé inverse d'une numérisation, qui est parfois appelée «dématérialisation» (un sujet d'actualité est par exemple la «dématérialisation» du cadastre français, c'est-à-dire la numérisation des plans papier).

[8] Notre capacité à interpréter un phénomène visuel à valeur mono-dimensionnelle dépend quasiment uniquement de l'ordre qu'il suggère. C'est ce type de remarque qui motive l'idée d'imposer à un algorithme d'analyse d'image d'être invariant par changement de contraste (tranformation croissante des valeurs de l'image). Nous discuterons de cette invariance à plusieurs reprises, principalement au chapitre 5.

elle-même un rectangle discret, mais peuvent être sur ou sous-échantillonnées avant affichage («zoomées»/«dézoomées» logiciellement). L'écran ayant certaines dimensions et une définition, on peut alors parler de résolution de visualisation.

Notons r_V la résolution de visualisation d'une image, exprimée comme une distance (e.g. 1/600 inch). Si une résolution d'acquisition r_A de l'image est également connue alors $e = r_V/r_A$ est l'**échelle** de la représentation, au sens où l'entendent les cartographes : e donne le rapport entre les distances figurées sur la représentation et les distances réelles sur l'objet (une échelle est donc sans dimension). Si la résolution r_A correspond au pas de numérisation d'un document ou d'une photographie alors l'objet de référence est le papier, le support physique du document que l'on a numérisé. L'échelle est alors le facteur d'agrandissement entre la représentation initiale et la nouvelle représentation (si le procédé de visualisation est une impression, nous venons de faire une photocopie). Si la résolution r_A renvoit à des distances sur des objets qui ne sont pas eux-mêmes des représentations (la Terre, une préparation microscopique...), l'échelle indique le rapport de taille entre la représentation et son contenu «ultime», le monde réel, qui est généralement celui qui nous intéresse.

La figure 2.2 illustre l'effet d'un changement d'échelle de représentation.

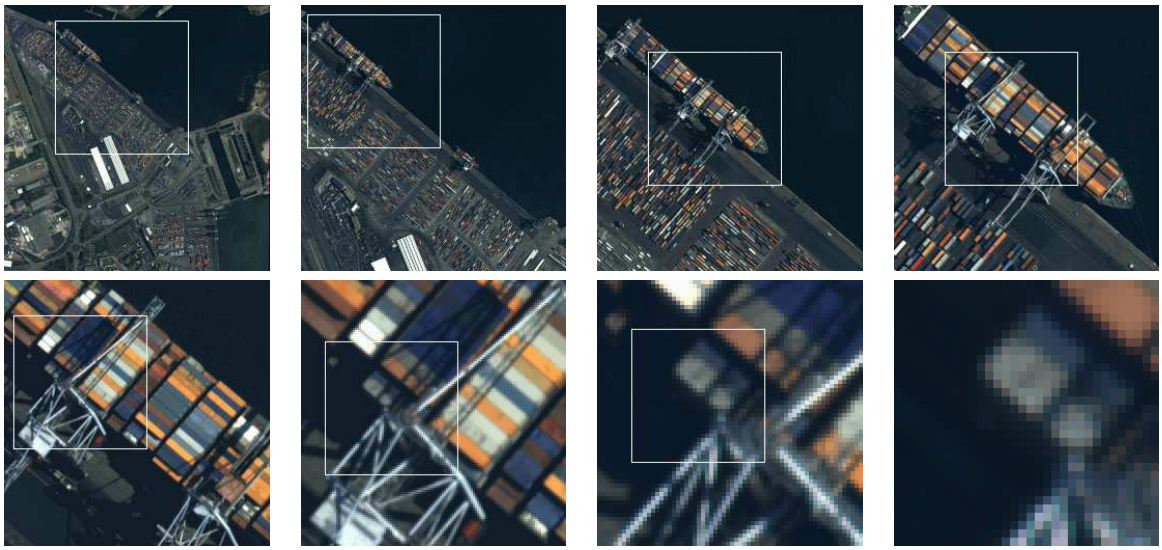


FIG. 2.2 – Grossissements successifs sur le port du Havre. L'image d'origine est une image 4096×4096 prise par une caméra numérique de l'IGN à une résolution au sol d'approximativement 35 cm. L'image en haut à gauche représente la scène complète ($\sim 1400 \times 1400$ m^2). Imprimée au format $A4$, cette image occupe un carré de $3,5$ cm de côté. L'échelle approximative de la représentation est donc de 1/40 000 (1 cm pour 400 m). Les images suivantes sont des grossissements successifs d'un facteur 2 (l'emprise de l'image $i+1$ est encadrée dans l'image i). La dernière image est à échelle 128 fois plus grande que la première, soit environ à l'échelle du 400ème sur un papier $A4$. On individualise alors clairement les pixels (il font environ 1 mm de côté).

Nous ne distinguons évidemment pas les mêmes choses dans deux représentations à des échelles différentes, ce qui est en premier lieu dû au fait que notre système visuel possède sa propre limite de résolution, qui est de l'ordre d'une minute d'arc. Au mieux, nous sommes donc capables de séparer deux points écartés de $0,1$ mm et placés à 30 ou 40 cm de l'œil. De plus en plus de détails deviennent ainsi perceptibles quand l'échelle de visualisation augmente (notre faculté à percevoir les détails dépend également d'autres facteurs : principalement de leur contraste et de notre capacité à leur attribuer une valeur sémantique. Ces facteurs seront discutés ci-dessous). Évidemment, quand la taille du pixel vu devient supérieure à la limite de résolution de l'œil, soit grosso-modo quand elle atteint $0,1$ mm (~ 250 dpi), il n'apparaît plus de nouveau détail. Le facteur limitant est alors la résolution de l'image et non plus celle de l'œil. Sur

la figure 2.2, cette limite se situe entre la $4^{\text{ème}}$ et la $5^{\text{ème}}$ image (la dernière de la première ligne et la première de la deuxième ligne), dont les tailles de pixel sont respectivement de $0,07$ mm et $0,14$ mm sur un format $A4$.

Pour conclure sur les «images visuelles» que l'on peut créer à partir d'images numériques, il est clair que ce que nous voyons dans ces représentations est très dépendant des choix de visualisation effectués. Avez-vous remarqué de quelle manière la belle palette bleu-rouge utilisée pour la représentation du MNE 2.1(b) sépare remarquablement bien le sol (bleu) du sur-sol (vert-rouge) ? Savez-vous que l'image du Havre de la figure 2.2 a subi une correction «gamma» de $2,5$ pour que nous percevions autre chose qu'un carré noir à l'impression ? Par ailleurs, comme l'illustre la promenade en échelle sur cette même image, ce que nous «voyons» dans une image dépend fortement de l'échelle à laquelle nous la regardons. Bien entendu, outre la limite de visibilité imposée par la résolution, il y a une question de perception locale ou globale, qui tient au fait que nous ne distinguons pas les stimuli ponctuels qui touchent notre rétine mais que nous voyons des touts, des ensembles de stimuli. Simuler ces différents niveaux d'analyse d'une scène est tout l'objet de cette thèse et sera amplement développé ci-dessous. Bornons-nous pour l'instant à constater qu'il y a de l'information dans une image numérique, mais qu'il nous faut jouer sur deux échelles[9] pour la rendre perceptible : l'échelle radiométrique (le contraste) et l'échelle spatiale (la taille).

2.3 Représentations cartographiques

Une carte est une représentation conventionnelle, généralement plane, de la répartition dans l'espace de phénomènes concrets ou abstraits. Lorsque l'on souhaite rédiger une carte à plus petite échelle à partir d'une carte à échelle donnée, la limite de résolution de notre système visuel donne lieu au problème de *généralisation cartographique*. Disons quelques mots de ce problème, qui est intimement lié aux questions d'analyse multi-échelles qui nous préoccupent dans ce document.

2.3.1 Généralisation cartographique

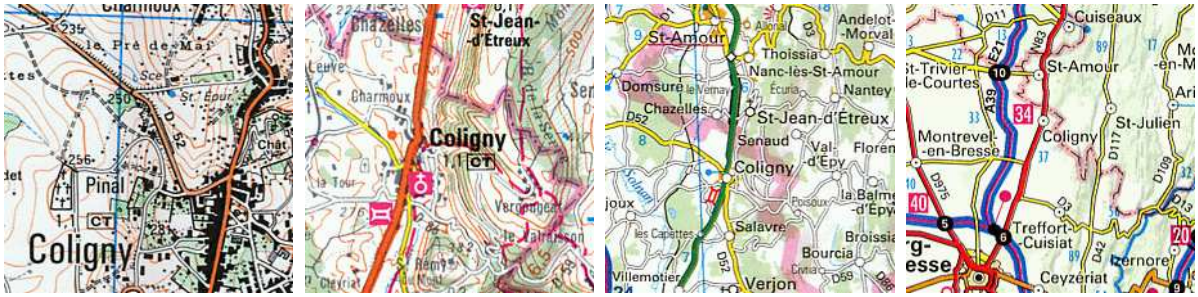
Pour être maniable, une carte doit avoir des dimensions humaines, soit environ 1 m^2. Pour être lisible sans loupe, un symbole sur une carte doit au moins avoir une taille d'un dixième de millimètre et deux symboles être distants d'au moins $0,1$ mm. Dans ces conditions, une carte de 1 m^2 à l'échelle du 1/25k (1/25 000, i.e. l'échelle d'une carte topographique IGN) couvre 25×25 km^2 et le plus petit détail que nous pouvons aprécier à l'œil nu représente $2,5$ m sur la Terre ; une carte au 1/100k couvre 100×100 km^2 (un département) et le système œil-carte a une résolution de 10 m ; au 1/250k, l'œil-carte couvre 250×250 km^2 (une région) à une résolution de 25 m ; au 1/1000k, il couvre 1000×1000 km^2 (la France métropolitaine) à une résolution de 100 m.

La figure 2.3 présente des extraits de cartes IGN à ces quatre échelles (approximativement...).

Leur contenu évolue nettement avec l'échelle : en raison de la limite de résolution de notre œil, il n'est pas possible de conserver toute l'information visible sur une carte au 1/25k dans des représentations à plus petite échelle. Il faut donc sélectionner, déplacer, déformer, agglomérer ou caricaturer des objets, ce que l'on appelle *généraliser* la carte. Un groupe de maisons proches peut par exemple être représenté comme une seule unité «îlot» quand l'espace entre les bâtiments, une fois mis à l'échelle finale, est trop faible pour pouvoir être discerné. On peut aussi écarter les symboles des maisons s'ils n'entrent pas en conflit avec d'autres objets environnants. De même, il faut supprimer ou distendre les virages trop serrés sur les routes, etc.

Globalement, une diminution d'échelle de représentation cartographique s'accompagne nécessairement d'une perte d'information (pour une même surface couverte). La représentation devient de plus en plus simple et simultanément de moins en moins fidèle à l'original. L'objectif de la généralisation est d'aboutir

[9] Échelle est ici employé au sens d'axe de mesure.



(a) 1/25 000 (b) 1/100 000 (c) 1/280 000 (d) 1/1 000 000

FIG. 2.3 – Cartes à différentes échelles. Si ce document est imprimé au format A4, ces extraits de cartes IGN sont aux échelles indiquées.

à une carte lisible et qui conserve ou respecte au mieux l'information pertinente, ce qui dépend bien sûr de la destination de la carte finale[10].

2.3.2 Des images aux cartes

Puisque nous avons tous les éléments en main, décrivons finalement la manière dont sont élaborées les cartes françaises.

L'IGN est chargé d'établir et maintenir à jour la carte topographique au 1/25k de l'ensemble du territoire français, également appelée carte de base[11].

Une grande part de l'information géographique qu'elle contient est à l'heure actuelle acquise par restitution photogrammétrique[12] manuelle à partir d'images aériennes. Disposant d'un couple d'images aériennes représentant la même zone terrestre vue sous deux angles différents et de données photogrammétriques sur ces images, un appareil de restitution photogrammétrique permet de perçevoir la scène en trois dimensions et d'y effectuer des mesures qui seront transposées en mesures terrestres. Grâce à un pointeur 3D contrôlé par deux manettes (xy) et une pédale (z), une personne dont c'est le métier et que l'on appelle un *restituteur* «dessine» alors *sur* le terrain virtuel les objets géographiques qu'elle voit et qui doivent apparaître sur la carte : elle détoure les bâtiments, suit les axes de routes, les courbes de niveau, etc. Si l'appareil de restitution est un appareil analogique, celui-ci reproduit simultanément le tracé du restituteur sur un papier - à l'échelle de la carte finale - par un système mécanique qui s'apparente à un pantographe. Les appareils modernes sont des appareils numériques, reliés à un ordinateur qui convertit et enregistre le traçé en coordonnées terrain, sous forme vectorielle[13] (et réinjecte aussi les traçés déja effectués dans la vue 3D). L'information acquise est gérée sous forme d'une base de données géographiques (BDG), c'est-à-dire grosso-modo par une structure de graphe attribué (avec certains aspects d'hypergraphe, plusieurs couches...).

Des travaux de terrain permettent de vérifier et compléter cette base par l'information qui n'a pas pu être acquise par photo-interprétation (relever les objets invisibles du ciel - e.g. en sous-bois - la nature

[10] Notons que si le pouvoir séparateur de notre œil était 40 fois plus important, nous pourrions cartographier la France entière sur 1 m^2 avec un contenu équivalent à celui d'une carte topographique au 1/25k. Il faudrait imprimer cette carte à une résolution de l'ordre de 10 000 points par cm.

[11] Elle constitue à proprement parler la base d'une hiérarchie de cartes...

[12] La photogrammétrie (de *gramma*, «dessin» et *-métrie*) est la discipline qui étudie les problèmes de mise en relation de mesures faites dans des images à des mesures dans l'espace physique tridimensionnel, aussi appelé espace «objet» ou «terrain». En robotique, les problèmes de photogrammétrie sont appelés problèmes de calibration.

[13] On distingue deux modes principaux de représentation numérique de données localisées : les représentation sous formes d'image, ou «raster», dont nous venons de parler, et les représentation dites vectorielles ou «vecteur», qui explicitent la géométrie des objets et éventuellement certaines relations entre objets, tel que des relations topologiques. Les SIG (Systèmes d'Information Géographique), les logiciels de DAO/CAO (Dessin/Conception Assistée par Ordinateur) ou de synthèse d'image manipulent tous des données vecteur.

des objets - e.g. bâtiment industriel ou habitation - la toponymie...).

Cette BDG de base s'appelle la BDTopo® (Base de Données Topographiques). Sa constitution emploie plusieurs centaines de personnes depuis 1989 et doit s'achever aux alentours de 2006.

Au même titre qu'une image numérique, une BDG ne constitue pas un objet graphique. Pour élaborer une carte à partir de l'information qu'elle contient il faut sélectionner les objets qui nous intéressent, leur affecter des symboliques, des couleurs, etc. Et évidemment, si l'échelle de la représentation finale est plus petite que l'échelle «nominale» d'utilisation des données[14] on tombe sur des problèmes de généralisation.

Les recherches conduites dans les trois principaux laboratoires de l'IGN (en terme de personnel) s'articulent logiquement autour de ce mode de production de cartes.

Le LOEMI (Laboratoire d'Opto-Électronique et de Micro-Informatique) est un laboratoire d'instrumentation qui conçoit des systèmes d'acquisition de données, principalement d'images aériennes pour ce qui nous concerne.

Le laboratoire MATIS (Méthodes d'Analyse et de Traitement d'Images pour la Stéréo-restitution) s'intéresse aux problèmes de photogrammétrie et à l'automatisation des processus d'acquisition et de mise à jour d'information géographique à partir d'images numériques (pas nécessairement de photographies aériennes, mais aussi d'images satellitaires, d'images de distance (e.g. laser), de cartes et de plans numérisés, etc.).

Le laboratoire COGIT (Conception Objet et Généralisation de l'Information Topographique) s'intéresse à tout ce qui concerne la conception et l'exploitation des bases de données géographiques. Leur exploitation à des fins cartographiques suscite en particulier d'importantes recherches sur l'automatisation des processus de généralisation cartographique, voir par exemple (Ruas 1999).

[14] La BDTopo® est établie à une précision terrestre de l'ordre du mètre. Son échelle nominale de représentation graphique est donc le 10 000ème. Une impression «brute» au 1/25k avec des symboles lisibles donne tout de même un résultat quasiment correct.

Chapitre 3

Pour une approche multi-échelles de l'analyse d'images en régions

What does it means, to see ? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the *process* of discovering from images what is present in the world, and where it is.

Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of *representing* this information -in all its profusion of color and form, beauty, motion and detail. The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions. This duality - the representation and the processing of information - lies at he heart of most information-processing tasks and will profoundly shape our investigation of the particular problems posed by vision.

David Marr, *Vision*[1]

3.1 Introduction

Entrons maintenant dans le vif du sujet, à savoir la segmentation d'images en régions. Tentons pour commencer quelques définitions générales.

i) Une *segmentation* est un ensemble de parties d'une image numérique. Les éléments d'une segmentation sont généralement appelées *régions*. Le terme de *région* fait habituellement référence à des parties de même dimension topologique que le support de l'image (des zones) - ou du moins qui symbolisent des phénomènes zonaux : 2D si l'image est en 2D, 3D si elle est en 3D, etc.

[1] Il s'agit des toutes premières lignes de (Marr 1982). On peut en proposer la traduction suivante : «Que signifie voir ? La réponse de l'homme de la rue (et celle d'Aristote aussi) serait : savoir où est quoi en regardant. En d'autres termes, la vision est le *processus* qui permet de découvrir à partir d'images ce qui est présent dans le monde, et où cela se trouve.

La vision est donc, d'abord et avant tout, une tâche de traitement d'information, mais on ne peut pas l'envisager seulement comme un traitement. En effet, si nous sommes capables de savoir où est quoi dans le monde, nos cerveaux doivent d'une manière ou d'une autre être capables de *représenter* cette information - dans toute sa profusion de couleur, de forme, de beauté, de mouvement et de détail. L'étude de la vision doit donc non seulement inclure l'étude de la manière d'extraire des images les différents aspects du monde qui nous sont utiles, mais aussi une interrogation sur la nature des représentations internes par lesquelles nous capturons cette information et la rendons ainsi utilisable pour nos décisions, nos pensées et nos actions. Cette dualité entre la représentation et le traitement de l'information est au cœur de la plupart des tâches de traitement de l'information et façonnera profondément nos recherches sur les problèmes spécifiques posés par la vision».

ii) Une *tâche de segmentation* est une tâche de délimitation de régions d'une image, c'est-à-dire de production d'une segmentation à partir d'une image. Réaliser une segmentation à partir d'une image s'appelle *segmenter l'image*. On parle aussi de *segmenter certains lieux spécifiques d'une image*, définis par certaines propriétés plus ou moins clairement formalisées, ce qui signifie déterminer - plus ou moins précisément - la géométrie des lieux en question, et seulement de ceux-ci. On parle ainsi de «segmenter les parcelles cultivées dans une image aérienne» ou de «segmenter une image en régions homogènes».

ii) Une segmentation peut être réalisée par une procédure manuelle ou automatique. Un *algorithme de segmentation* est une procédure automatique qui détermine une segmentation à partir d'une image et des valeurs d'un certain nombre de paramètres (la procédure est dite semi-automatique si certains de ces paramètres sont le résultat d'une interaction préliminaire ou en divers point de la procédure).

iii) Un *problème de segmentation* (automatique) est un problème d'analyse automatique d'image dont la réponse est un *algorithme* de segmentation. Un algorithme de segmentation peut être la réponse exacte ou approchée à un problème de segmentation, mais aussi bien n'être la réponse à aucun problème clairement formalisé.

iv) La *segmentation d'image* est la branche de l'analyse d'image qui étudie les problèmes de segmentation et/ou met au point des algorithmes de segmentation.

Le terme *segmentation* est un terme extrêmement général et flou. En analyse d'images, il englobe pratiquement toutes les problématiques qui touchent à la délimitation de zones. Forsyth et Ponce (2003, p. 301) indiquent qu'«il n'existe certainement pas à ce jour de théorie globale de la segmentation, et [que] le terme segmentation est utilisé de différentes façons par différents auteurs»[2]. Le point de départ de toute théorie est bien entendu une définition claire de l'objet de la théorie. Et si la «segmentation» est définie de différentes manières par différents auteurs, il peut difficilement en résulter une théorie unifiée...

L'objet de la discipline «segmentation» est la modélisation et l'étude des *processus* qui permettent réaliser des *tâches* de segmentation d'images numériques. Le point de départ d'une éventuelle théorie de la segmentation est donc la définition claire des tâches dont elle se préoccupe (ou de *la* tâche dont elle se préoccupe, dans l'éventualité où il existe une opération unique sur une image que l'on puisse appeler *la* tâche de segmentation).

L'objectif de ce chapitre est de définir clairement la tâche de segmentation à laquelle nous nous intéressons dans ce document, que nous appellerons la tâche de segmentation «générique» de «bas niveau» d'une image. En examinant le *rôle* de cette tâche dans un système complet d'interprétation d'images, nous argumenterons alors que le résultat qu'elle produit doit - au minimum - être une analyse multi-échelles : une hiérarchie de régions représentant des phénomènes visuels imbriqués. Dans un deuxième temps, les parties suivantes proposeront deux façons de modéliser la réalisation de cette tâche, ce qui passera systématiquement par deux étapes successives : i) formaliser mathématiquement le problème et ii) dégager un algorithme pour résoudre (exactement ou approximativement) le problème mathématique.

3.2 Bas et haut niveau d'analyse visuelle

Lorsque l'on parcourt la littérature consacrée à la segmentation, on constate que le terme *segmentation* est utilisé pour désigner deux types de tâches très différentes : des tâches de segmentation dites de «haut niveau» et des tâches dites de «bas niveau». Le fait que le même terme soit utilisé indifféremment pour ces deux types de tâches est source de confusions. Commençons donc par éclaircir la différence qui existe entre ces deux types de problèmes de segmentation.

[2] «There is certainly no comprehensive theory of segmentation at time of writing, and the term is used in different ways in different quarters.»

3.2.1 Problèmes d'analyse de haut niveau d'une image

Loconnaissance de formes

Olivier Catoni (2003) introduit la problématique de segmentation à travers la celle de reconnaissance de formes. Appuyons-nous sur ses propos :

«En un mot la reconnaissance de formes consiste à repérer dans des signaux compliqués (signal de parole ou images numérisées, mais aussi texte en langue naturelle ou séquence d'ADN...) la présence de certaines structures (ou objets, ou formes...) et à les classer en un nombre restreint de classes pertinentes pour l'application envisagée, le tout de façon automatique par un ordinateur. Cette définition peut paraître floue ; la discipline tire en fait sa cohérence de l'existence de difficultés techniques communes aux différents cas de figure possibles.

Indiquons tout d'abord par quelques exemples dans quels contextes pratiques on fait appel à la reconnaissance de formes. Il s'agit en général de reproduire de façon automatique des tâches d'interprétation et de discrimination réalisées par le cerveau humain avec un certaine facilité, mais par des voies difficiles à élucider. Quelques illustrations en vrac dans le domaine de l'imagerie : interpréter des images médicales et les classer en fonction des pathologies dont elles témoignent, interpréter automatiquement des images satellites (y reconnaître les différents types de sols, les réseaux routiers et fluviaux, les différents types de bâtiments...). [Suivent d'autres exemples ...]

[...] Il convient maintenant de mettre en lumière les points communs de ces applications à première vue disparates.

Essayons pour cela de recenser quelques difficultés qui fondent l'unité de la discipline. Une première difficulté vient du flou habituel avec lequel la tâche est définie : en fait il s'agit dans les exemples cités de l'automatisation d'une tâche effectuée par un être humain, si bien que le résultat souhaité est le plus souvent défini par l'exemple. [...]

Une [deuxième] difficulté générique est celle de la localisation de la structure à reconnaître : pour reprendre notre exemple, il ne s'agit pas uniquement de dire si une zone donnée d'une radiographie correspond à une structure pathologique, mais aussi de délimiter cette zone. Le sous-problème consistant à délimiter la zone correspondant à la structure recherchée s'appelle *segmentation*. La segmentation est un problème fondamental de la reconnaissance des formes, qui conditionne tout le reste de la marche à suivre et n'est à ce jour que très mal résolu»[3].

La «segmentation» vient d'être définie comme une composante essentielle de toute réponse à une question de reconnaissance de formes (RdF). La tâche de délimitation de régions est alors associée à une capacité de «haut niveau», à savoir interpréter les zones extraites, les ranger dans certaines classes sémantiques particulières. C'est sous cette forme que sont initialement posés la plupart des problèmes pratiques d'analyse d'image : il s'agit de reconnaître certains objets *spécifiques* - décrits en langage naturel - dans une image appartenant à une catégorie *spécifique*, e.g. reconnaître et délimiter les champs cultivés dans une image aérienne panchromatique à 80 cm de résolution sol. Il y a alors a priori autant de problèmes de «segmentation» différents que de types d'images (en nature et en sujet) et de types d'objets qui sont susceptibles de présenter de l'intérêt à nos yeux dans ces images. Malgré cette diversité, la RdF espère parvenir à des méthologies relativement génériques de reconnaissance de formes qui, une fois nourries des spécificités de chaque problème, débouchent sur des méthodes de reconnaissance adaptées à chaque objet d'intérêt (un argument pour soutenir cet espoir est qu'il serait étonnant que notre cerveau développe une stratégie spécifique pour chaque forme que nous sommes capable de reconnaître).

Comme l'indique Catoni, la délimitation des structures - «la» segmentation - est un problème fondamental de la RdF, qui a reçu énormément d'attention depuis les origine de la discipline[4] mais n'est

[3] Notons que Catoni ne relève pas de troisième difficulté générique.

[4] Henri Maître écrivait en préface de (Cocquerez et Philipp 1995) que «les travaux sur la segmentation constituent le corpus littéraire le plus important du traitement des images» et c'est probablement encore vrai aujourd'hui.

pourtant encore que très mal résolu. Pourquoi ? Un nœud du problème apparaît clairement dans la description de Catoni :

Un problème de RdF comporte deux aspects : *reconnaître* et *localiser* certains objets, et ces deux aspects sont profondément imbriqués. L'aspect localisation n'est pas un problème indépendant, inconditionnel. Il ne s'agit pas de délimiter n'importe quels ensembles, mais seulement quelques zones précises, celles qui correspondent aux objets recherchés. Il apparaît alors une sorte de dilemme d'«œuf et de poule» qui a alimenté de nombreux débats depuis les origines de la discipline. Littéralement re-connaître signifie «déterminer comme déja connu», c'est-à-dire déterminer que *quelque chose* est déja connu. Mais quoi ? Évidemment pas l'ensemble de l'image prise en bloc, mais bien certains lieux particuliers de l'image. Détecter la *présence* d'une forme connue dans un ensemble d'observations comporte nécessairement une part de localisation, même floue ou approchée. Inversement, localiser précisément un objet spécifique nécessite d'une manière ou d'une autre de déterminer que le lieu délimité correspond bien à l'objet recherché, c'est-à-dire comporte nécessairement une part de reconnaissance.

Dans le plus pur esprit d'analyse cartésien, on décompose habituellement la tâche globale de «loconnaissance» en deux - «localisation» et «reconnaissance» - mais dans le cas général, ces deux «sous-tâches» sont deux aspect indissociables d'un seul et unique problème.

Interprétation d'images complexes et vision artificielle

Face à une image, un algorithme de RdF cherche à répondre à une question du type «où est ça ?» dans laquelle «ça» désigne une classe d'objets spécifiques. L'attention est focalisée sur certains objets «pertinents pour l'application envisagée» et la méthode les recherche activement. L'opération de reconnaissance est alors globalement une opération *descendante* - des modèles vers les données - dans le sens où elle cherche à *retrouver* dans l'image un modèle d'objet connu. En pratique, si l'on a recours à une méthode de RdF - destinée à répondre à une question du type «où est ça ?» - c'est que l'on *sait* déja, pour une raison ou pour une autre, que «ça» se trouve quelque part dans l'image, ou du moins que «ça» a de fortes chances de s'y trouver : on cherche par exemple à reconnaître les lettres dans un document écrit, les visages sur une photographie de groupe, etc.

Sous sa forme générale, le problème d'*interprétation d'image* ou de *vision artificielle* est largement plus complexe. Il s'agit, face à une image relativement arbitraire, de répondre à la question : «que représente l'image ?» ou «que peut-on *voir* dans l'image ?», de la même manière que nous, humains, sommes capables de le faire. La différence essentielle avec les questions spécifiques de RdF est que - pour reprendre les termes de Marr cités en exergue - outre déterminer «où»[5], il faut également déterminer «quoi». Et si l'on ne sait pas a priori ce qui se trouve dans l'image, il faut le *découvrir*, ce qui implique nécessairement une part d'analyse *ascendante* des données. Dans le cadre du problème général d'interprétation d'image, l'interdépendance entre localisation et reconnaissance devient donc particulièrement forte et rejoint d'un certain point de vue la complémentarité entre processus cognitifs ascendants et descendants.

Notons que, de même qu'il est improbable que notre cerveau développe autant de stratégies de reconnaissance que d'objets connus, il est improbable qu'il réponde à la question globale «où est quoi ?» par énumération exhaustive des questions de type «où est ça ?» pour tous les «ça» connus.

3.2.2 Analyses de bas niveau et convergence vers une interprétation

La segmentation vient d'être envisagée comme une opération de «haut niveau», associée à une tâche d'interprétation. Intéressons-nous donc de plus près à la nature d'une opération d'interprétation.

[5] Notons que dans l'idée de Marr, déterminer «où» veut dire localiser les objets dans l'espace physique, tridimensionnel. La segmentation ne traite pas de localisation 3D à partir d'images 2D mais seulement de localisation d'ensembles dans l'espace image (en 2D si elles sont en 2D, en 3D s'il s'agit d'images volumiques). Le calcul de volumes apparents à partir d'images 2D relève de techniques de stéréo-vision (ou multi-vision) ou d'analyse des ombrages, de la texture, du mouvement... («shape from X», où X vaut shading, texture...).

Descriptions de bas et de haut niveau

Il est évident qu'aucun système cognitif, aussi perfectionné soit-il, ne peut être capable d'interpréter le contenu de *n'importe quelle image* qui est susceptible de lui être présentée.

Considérons par exemple l'image (b) de la figure 2.1, page 8. Si cette image était présentée sans aucune légende à une personne qui n'a jamais vu auparavant de MNE, pourrait-elle dire ce qu'elle représente ? Une portion de circuit imprimé ? Un observateur ne peut émettre d'hypothèse plausible sur le contenu d'une image que s'il réussit à établir certaines similitudes avec d'autres images ou représentations dont il détient les clés d'interprétation. Par exemple, la juxtaposition d'un plan cadastral de la ville d'Amiens au MNE 2.1(b) peut suffir à donner les clés d'interprétation du MNE. Face à ce couple d'images, un observateur établit spontanément une correspondance entre les zones orangées ou rougeatres du MNE et les polygones hachurés du plan cadastral, qui symbolisent des bâtiments. Si cet observateur sait interpréter un plan cadastral, il déduira alors une interprétation plausible du MNE. Une autre manière de fournir les clés d'interprétation du MNE 2.1(b) est simplement d'en décrire la nature (grille d'altitudes), le sujet (zone urbaine), la résolution, etc. Disposant de ces informations, quiconque sait ce qu'est une «ville» sera capable d'interpréter les tâches rouges de l'image comme représentant des bâtiments[6].

Cet exemple conduit à trois conclusions.

La première est une évidence : interpréter une scène, c'est-à-dire attribuer une valeur sémantique à ses constituants, est une opération de mise en correspondance entre des modèles et des observations.

La seconde est moins flagrante : même si l'on dispose de connaissances générales suffisantes pour correctement interpréter la scène - e.g. si l'on sait qu'un bâtiment se présente comme une excroissance au dessus du sol, aux formes relativement géométriques, etc. - ces connaissances générales ne sont en elles-mêmes pas suffisantes pour décider de la nature de ce qui est vu. Notre capacité à interpréter une image dépend également de connaissances a priori sur la scène observée, sur ce qui est susceptible de s'y présenter et sous quelle forme, quel aspect. Sans ce type de connaissance - que nous qualifierons de connaissance *contextuelle* - le problème est généralement sous-contraint et plusieurs interprétations sont possibles. Ce point est lié à un phénomène que les psychologues de la perception appellent la «probabilisation»[7] : selon ce que l'on s'attend à voir dans une image, certains schema cognitifs sont plus ou moins activés ou inhibés (Jimenez 1997). Ce phénomène se manifeste par exemple par une dépendance au contexte visuel du temps que nous mettons pour reconnaître un objet. Plus la présence de l'objet dans la scène est inhabituelle ou inattendue, plus le temps mis pour l'identifier est important. Cet activation sélective de modèles ne dépend pas que de l'environnement visuel, mais de l'ensemble du contexte mental au moment de l'observation.

Un troisième point ressort de la discussion qui précède sur le MNE d'Amiens : même si nous ne sommes pas toujours capables d'effectuer une interprétation sémantique, de haut niveau, d'une image, il semble en revanche que nous soyons systématiquement capables d'en faire une description purement *visuelle*, en termes de formes géométriques, de couleurs et de leurs relations. Face à l'image 2.1(b), une personne non avertie de ce qu'elle représente pourra tout de même identifier des formes, par exemple distinguer un objet rouge ayant la forme d'un chiffre '9' couché ( ) au centre de l'image.

Nous dirons qu'une analyse qui produit uniquement des éléments de description visuelle d'une scène, en termes de formes géométriques et/ou de caractéristiques d'aspect (couleur, clarté, texture...) est une analyse de **bas niveau**. Une analyse de bas niveau pourra par exemple décrire un *'ovale rose'* là où nous voyons un *'visage'*. Préciser que ce lieu correspond à un *'visage'* constitue quant à soi une description de haut niveau.

[6] Après avoir lu ces lignes, un relecteur étranger au domaine nous a fait remarquer qu'il ne voyait pas ce que représentait la quatrième image (d) de la figure 2.1. Nous n'avions pas encore indiqué en légende qu'il s'agissait d'une image médicale d'une tête humaine ! L'image radar 2.1(c) est certainement encore plus «obscure» pour qui ne dispose d'aucune clé de lecture.

[7] On peut effectivement modéliser ce phénomène sous un angle bayesien, par l'introduction d'une probabilité a priori sur l'ensemble des interprétations possibles (voir 6.1.3 p. 115).

Telle quelle, la différence entre ces deux niveaux de description n'est pas claire. Cette différence ne tient par exemple pas au fait que la description emploie ou non un langage symbolique (*ovale rose* et *visage* sont par exemple deux expressions symboliques). Pour clarifier cette différence, faisons un détour par l'idée d'économie de représentation d'Ernst Mach.

Le principe d'économie de représentation de Mach

Nous nous représentons donc un espace peuplé d'objets auxquels nous attribuons des significations. Si l'on s'interroge sur *comment* notre cerveau peut bien construire une telle représentation - ou comment l'on peut simuler artificiellement une telle construction - il peut être utile de se demander *pourquoi* : quel intérêt peut-il y avoir à former une telle vision du monde ?

Selon Ernst Mach, «la plus grande fréquence, la suprématie des constantes sur les variables [dans nos sensations], nous imposent une économie de la représentation et de la dénomination, pour une part instinctive, pour une part volontaire et consciente, qui s'exprime dans la pensée et le langage quotidiens. Ce qui a été représenté de manière *unique*, reçoit une *seule* dénomination, un *seul* nom» (Mach 1886). Ainsi, d'après Mach, c'est la structure même de nos perceptions qui nous *impose* une représentation mentale en termes d'objets. Nos perceptions présentent des permanences, des redondances, des régularités, des cohérences, symétries... qu'il est possible de représenter de manière économique et nous le faisons. En langage moderne, nos perceptions sont compressibles et notre cerveau les compresse. Grouper par paquets, exploiter la redondance et utiliser un dictionnaire sont les mécanismes de base qui permettent de compresser l'information, de faire des descriptions succintes de données.

Le principe d'économie de représentation - ou d'économie de pensée - de Mach est lié au principe de parcimonie, ou principe du rasoir d'Occam[8], qui est souvent caricaturé par l'expression «l'explication la plus simple est la meilleure». Le principe de parcimonie peut être considéré comme un des principes de base de la pensée scientifique. On peut toutefois en distinguer deux formes différentes. La première forme est colorée de métaphysique et correspond à l'idée d'Aristote selon laquelle «la nature emprunte le chemin le plus simple». La seconde forme est purement pragmatique et correspond à la célèbre phrase d'Einstein : «tout devrait être rendu aussi simple que possible, mais pas plus simple». Sous cette dernière forme, l'argument n'emploie aucune hypothèse relative à une éventuelle réalité sous-jacente aux observations : de deux modèles qui rendent également compte des observations, le plus concis est le plus aisé à mémoriser, le plus pratique à manipuler ; il doit donc être préféré pour des raisons purement pragmatiques. Notons que le principe de parcimonie renvoie également à un sentiment esthétique : nous jugeons qu'un modèle (mathématique, physique...) est d'autant plus «beau», «élégant», que le ratio entre sa «complexité» et celle des phénomènes qu'il permet de décrire est faible. Définir la «complexité» est un problème épineux sur lequel nous reviendrons dans la troisième partie du manuscript[9].

Au delà d'une simple règle pragmatique, Mach voyait dans le principe d'économie un mécanisme premier de l'élaboration de nos représentations mentales[10]. Le principe d'économie fournit par exemple un argument simple en faveur de la thèse réaliste, celle de l'existence d'un monde matériel qui serait la cause de nos perceptions. Prenons l'exemple de la vision. Une difficulté majeure du problème est qu'un même objet physique peut donner lieu à un nombre arbitrairement grand d'apparences visuelles, quand

[8] Du nom d'un moine franciscain anglais du XIVème siècle, William d'Occam (ou d'Ockham), pour qui «les entités ne devraient pas être multipliées sans nécessité».

[9] Une question fondamentale consiste à savoir si l'idée de «complexité» absolue, «universelle» ou intrinsèque, i.e. indépendante d'un l'observateur, possède un sens.

[10] On peut proposer un argument évolutionniste, Darwinien, en faveur de la thèse de Mach : Toutes choses égales par ailleurs, un organisme est d'autant plus apte à survivre qu'il possède i) des organes de perception développés (e.g. vue) et, ii) une capacité de mémoire importante. Plus sa mémoire est importante, plus il peut se rappeler d'expériences passées et de règles d'action adaptées à chaque situation. Cependant, la capacité à acquérir de l'information (brute) sur le monde extérieur est incomparablement plus importante que la capacité à la mémoriser ; il nous serait par exemple impossible de mémoriser tel quel le flot gigantesque de stimuli visuels qui nous parviennent à chaque minute. Un organisme doté de vue et de mémoire sera alors d'autant plus performant qu'il sera apte à former des représentations mentales concises de ses perceptions.

les conditions d'observation varient : point de vue, distance, éclairage, etc. Attribuer toutes ces apparences à une même origine - la présence d'un objet physique unique d'un espace matériel 3D dont nous serions l'observateur - permet de simplifier considérablement la représentation mentale de nos sensations, de représenter une quantité de vues différentes par le modèle d'un objet «réel» unique.

Une telle représentation mentale s'appuie sur deux éléments : i) la distinction entre un *observateur* et un *observé*, i.e. *notre* propre représentation mentale en tant qu'observateur détaché d'une réalité autonome ; et ii) la modélisation de cette réalité extérieure comme un monde tridimensionnel. Un objet physique correspond alors à une classe d'équivalence d'aspects sensoriels, obtenue lorsque l'on fait abstraction des conditions d'observation, c'est-à-dire précisément quand les sensations sont objectivées, quand sujet et objet, observateur et observé, sont découplés. Le mécanisme d'objectivation est, certainement par voie de conséquence, un des piliers centraux de la physique classique (lire à ce propos Erwin Schrödinger (1956, ch.3, pp.183-194))[11]. En ce qui concerne le deuxième point, à savoir la géométrie que nous attribuons à l'espace, appuyons-nous sur les propos d'Henri Poincaré. Selon Poincaré, c'est l'existence de solides, et la remarque de certaines corrélations ou décorrélations entre nos sensations *visuelles*, *tactiles* et *motrices* qui explique la formation primitive du concept d'espace géométrique : «si donc il n'y avait pas de corps solides dans la nature, il n'y aurait pas de géométrie» (Poincaré 1902, p. 86). Poincaré attribue alors explicitement le fait que nous donnions trois dimensions à l'espace physique à un principe de parcimonie : «En résumé, l'expérience [quotidienne] ne nous prouve pas que l'espace a trois dimensions ; elle nous prouve qu'il est commode de lui en attribuer trois, parce que c'est ainsi que le nombre de coups de pouce est réduit au minimum» (Poincaré 1905, pp. 93-94). Les «coups de pouce» dont parle Poincaré sont les tapotements sur les objets physiques grâce auxquels nous faisons l'expérience de l'espace.

Nous verrons dans la troisième partie du mémoire que certaines théories de l'inférence statistique de modèle (Minimum Description Length - MDL (Rissanen 1978), Minimum Message Length - MML (Wallace et Freeman 1987)) argumentent du principe de parcimonie pour dériver une méthodologie d'inférence s'appuyant sur un critère de compression optimale *sans perte*. Le même principe nous conduira pour notre part à une formulation du problème de modélisation comme un problème d'optimisation sous contrainte, correspondant à toute une famille de problèmes de compression *avec perte*, de type débit/distorsion (Shannon 1959). Dans le cadre de la segmentation, nous proposerons d'envisager cette famille de problèmes comme un tout, représentant tout un continuum de modèles - à différentes échelles - de l'image.

Le haut niveau est le niveau utile au raisonnement

Revenons maintenant sur la différence de fond qui existe entre le «bas» et le «haut» niveau de description. À un objet sémantique qui possède une réalité perceptuelle, e.g. à un '*visage*', correspondent deux types de caractéristiques :

1. Un ensemble d'aspects possibles provenant de nos expériences sensorielles de l'objet (et en particulier, pour un objet visuel, un ensemble d'apparences possibles dans une image).

2. Un ensemble de caractéristiques non perceptuelles, relatives à ses propriétés propres (nature physique, fonction, comportement possible...) ou aux relations qu'il entretient avec d'autres phénomènes (e.g. pour un '*visage*' : '*corps*', '*yeux*', '*chapeau*'...). Ces caractéristiques peuvent également provenir d'expériences sensorielles mais aussi bien avoir été apprises par d'autres biais, e.g. par un apprentissage oral ou écrit[12].

[11] Poussons à son terme l'argument évolutionniste proposé en note 10. D'après ce que nous venons de voir, le mécanisme d'objectivation, i.e. le développement d'un modèle de «soi», est un moyen de simplifier ses représentations mentales du monde et donc d'accroître ses performances. De ce point de vue, et en supposant que le projet de l'Intelligence Artificielle dans son sens le plus dur de rendre «consciente» une machine soit possible, il semble qu'une étape fondamentale soit de rendre «consciente de soi» une machine, c'est-à-dire de parvenir à concevoir une machine qui forme spontanément une représentation d'elle-même comme étant un observateur autonome d'une réalité extérieure.

[12] Selon Rudolf Arnheim, tous les mots ou expressions renvoient in fine à des expériences sensorielles, y compris les notions les plus abstraites tels que les concepts mathématiques ou physiques, et nous sommes assez enclins à penser comme lui. Par exemple, «la notion de profondeur de pensée dérive de la profondeur physique. Mieux encore, la profondeur n'est pas simplement une métaphore commode destinée à décrire le phénomène mental : c'est le seul moyen possible de parvenir à

Lorsque nous associons un nom à une apparence dans une image nous établissons alors une relation entre cette apparence (une caractéristique de type 1 de la classe sémantique) et tout un réseau complexe de connaissances (l'ensemble des autres apparences possibles de l'objet et l'ensemble de ses caractéristiques de type 2). En substance, cette mise en correspondance permet de transformer un fait visuel *ponctuel* - le fait observé - en un fait *étendu*, à la fois spatialement et temporellement. Elle permet de relier l'apparence observée à un certain nombre d'autres apparences voisines possibles (aspect spatial) et d'apparences passées ou futures possibles (aspect temporel). Ce fait étendu permet alors d'effectuer des raisonnements, e.g. de déduire une évolution, un comportement possible du phénomène. A contrario, une description de bas niveau véhicule uniquement des informations concernant l'apparence ponctuelle d'une forme et ne permet pas d'effectuer de raisonnement.

Un exemple bien choisi vallant mieux qu'une longue explication : voir un objet brun en forme de tube sur le sol est inutile en soi ; s'il s'agit d'une '*branche*' elle ne risque pas de nous mordre, mais s'il s'agit d'un '*serpent*' il vaut mieux l'avoir reconnu avant de poser le pied dessus (aspect temporel). De même, reconnaître qu'il s'agit d'une branche permet de déduire la présence probable de '*feuilles*' à proximité (aspect spatial).

Bref, le haut niveau de description est le niveau utile au raisonnement ou à l'action.

3.3 Articulation entre bas et haut niveau d'analyse

De manière générale, l'objectif d'un système automatique d'interprétation d'images est de construire une description de haut niveau de la scène perçue, c'est-à-dire de passer d'une image - un paquet de valeurs numériques - à une description en termes d'objets sémantiques visibles, exploitable par des mécanismes de raisonnement symbolique de haut niveau. Nous avons vu qu'interpréter une image n'était pas toujours possible. Par contre, il semble que nous soyons systématiquement capables d'effectuer des descriptions de *bas niveau* des images, ce qui suggère que l'élaboration de telles descriptions puisse intervenir de façon systématique, en tant que préliminaire à l'analyse de haut niveau des perceptions. Cette idée a d'abord été formulée par les psychologues Gestaltistes dans la première moitié du XXème siècle, puis a été reprise par David Marr à la fin des années soixante-dix, dans le cadre de la vision artificielle.

3.3.1 Groupements perceptuels

«Le percept est donné à la conscience comme un tout»

«La psychologie Gestaltique est un courant phénoménologique pour lequel le percept est donné à la conscience comme un tout, une configuration qui a des propriétés émergentes qui sont différentes de la simple somme de ses parties» (Holender 2001, pp. 41-42).

On retient habituellement de la théorie de la Gestalt la deuxième partie de cet énoncé, qui correspond à la phrase célèbre de Wertheimer (1924) : «La «formule» fondamentale de la théorie de la Gestalt peut être exprimée de la manière suivante. Il y a des touts, dont le comportement n'est pas déterminé par celui de leurs éléments individuels, mais dans lesquels les processus partiels sont eux mêmes déterminés par la nature intrinsèque du tout. La théorie de la Gestalt nourrit l'espoir de déterminer la nature de tels touts»[13].

Toutefois, un deuxième aspect essentiel de la théorie de la Gestalt est que i) ce sont ces touts que manipule la «conscience», i.e. qui sont donnés aux processus d'interprétation, et que ii) l'élaboration de ces percepts, ou primitives, relève uniquement de quelques propriétés de *bas niveau*, i.e. qui ne font intervenir

concevoir cette notion. Sans la conscience de la profondeur physique, la profondeur mentale est purement est simplement inconcevable. D'où la qualité figurative de tout discours théorique» (Arnheim 1969, pp. 245-246).

[13]Le texte original de Wertheimer est en Allemand mais notre traduction française s'appuie sur la traduction anglaise de Willis D. Ellis : «The fundamental «formula» of Gestalt theory might be expressed in this way. There are wholes, the behaviour of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. It is the hope of Gestalt theory to determine the nature of such wholes.».

aucune connaissance sémantique : «l'hypothèse Gestaltique est que le processus primaire de constitution [d'une description visuelle générique] et celui secondaire de reconnaissance sont indépendants» (D'Ales, Froment, et Morel 1999).

Le principe de Helmholtz

Par de nombreuses expériences de psychologie perceptuelle, les psychologues Gestaltistes ont établit une liste de propriétés visuelles de bas niveau[14] qui sont supposées pouvoir modéliser toutes les configurations d'éléments que nous percevons tous[15] comme des touts : connexité, identité chromatique, proximité, similitude, continuité de direction, etc. (Kanisza 1997; D'Ales et al. 1999). De nombreux chercheurs en vision artificielle se sont intéressés à la *construction* de ces groupements perceptuels, par des techniques variées, voir par exemple le livre de Lowe (1985). Prenant en quelque sorte le contrepied des approches habituelles, Desolneux, Moisan et Morel se sont récemment attachés à répondre à la question : étant donnée une hypothèse de groupement - qui a pu nous parvenir par un moyen quelconque - comment peut-on décider s'il s'agit d'un groupe objectivement pertinent, au sens précis où nous le verrions tous dans l'image ? Ils proposent de répondre à cette question en appliquant un principe qu'ils attribuent à Helmholtz.

Le test de pertinence est relatif à *une* propriété donnée d'un ensemble de percepts élémentaires (alignement, parallélisme...), c'est-à-dire porte sur ce qu'ils nomment une Gestalt Partielle (Desolneux et al. 2001b). Supposons que l'on observe une Gestalt Partielle dans un ensemble de percepts élémentaires, e.g. trois points alignés dans un ensemble de dix points. Grossièrement, cette Gestalt est significative, i.e. nous la percevons, dès que la probabilité qu'une telle configuration soit apparue par hasard est négligeable. Le critère précis de décision est simple : quel serait l'espérance du nombre d'occurrences d'une telle configuration si les percepts élémentaires avaient été produits par un processus purement aléatoire (uniforme selon tous les paramètres) ? Si cette espérance est inférieure à 1, la configuration testée est pertinente : le fait d'en observer *une* correspond en soi à une déviation par rapport à l'aléatoire[16].

Desolneux et al. appliquent également un autre principe pour élire au rang de Gestalt Partielle une configuration, le *principe de maximalité* (Desolneux et al. 2000) : pour être remarquable, une configuration visuellement pertinente ne doit pas être un sous-ensemble d'une configuration également pertinente. Ils montrent alors pour certains modèles de formes (segments) que les configurations pertinentes maximales constituent une partition de l'espace. Nous verrons qu'un tel principe de maximalité intervient - plus ou moins explicitement - dans toutes les formulations de la segmentation. Nous proposerons au chapitre 5 de relâcher ce principe en un principe de «maximalité locale», ce qui nous conduira à des solutions multi-échelles.

L'objectif affiché de Desolneux et al. est de parvenir à des algorithmes de traitement d'image s'appuyant sur des principes «objectifs», ce qui d'un point vue pratique signifie parvenir à des algorithmes *sans paramètre* (visible ou caché). Dans (Desolneux et al. 2001a), ils positionnent ce point comme un «challenge» de la discipline. Tous les praticiens du traitement d'image savent en effet que la difficulté de paramétrage est un éceuil majeur à l'opérationalité de nombreux algorithmes. Nous verrons ci-dessous qu'il s'agit d'un des problèmes fondamentaux en segmentation d'images.

[14] Pour garantir qu'aucun mécanisme de reconnaissance sémantique n'intervient dans la perception des formes, les figures visuelles utilisées sont purement abstraites.

[15] Pour que la méthode soit correcte, il faut éliminer toute subjectivité, ce qui ne peut se faire qu'en considérant les formes que nous voyons tous, c'est-à-dire intersubjectives (voir à ce sujet la section 5.3.3 p. 76).

[16] Notons que si l'on définit les signaux aléatoires à la manière de Kolmogorov ou Chaitin, comme étant les signaux incompressibles (cf. 6.1.4 p. 118), le principe de Helmholtz peut s'envisager comme une variante du principe d'économie de Mach.

3.3.2 La théorie de David Marr

Différer l'«intelligence»

Synthétisant des idées de neurobiologie[17] et de psychologie Gestaltique, et s'attaquant au problème général de vision artificielle, David Marr a proposé que les processus de cognition liée à l'information visuelle puissent être décomposés en deux grandes étapes d'analyse : une étape de bas niveau produisant une description générique de la scène perçue, en termes de volumes en mouvement non interprétés, suivie d'une étape de haut niveau, d'interprétation de ces volumes en termes d'objets connus.

Cette décomposition s'appuie sur deux *hypothèses* fortes :

i. **«Voir» ne nécessite pas de «comprendre»** : des mécanismes de bas niveau sont capables de correctement analyser les formes en n'employant qu'un corpus réduit de connaissances *génériques* sur ce qui est susceptible de se présenter dans une image (point de vue réaliste) ou qui est susceptible d'être utile à la compréhension d'une image (point de vue positiviste).

ii. **«Comprendre» ne nécessite que de «voir»** : chaque mécanisme *spécifique* d'interprétation de de haut niveau peut s'appuyer «aveuglément» sur les résultats des analyses de bas niveau, sans avoir à revenir aux perceptions brutes.

À l'époque de Marr, l'opinion générale était que les mécanismes - naturels ou artificiels - d'interprétation de scène devaient nécessairement reposer sur l'emploi de toutes les connaissances disponibles, et ce dès les premières étapes de traitement. Autrement dit, on ne dissociait pas nettement «voir» de «comprendre». Le problème de la vision était alors jugé comme tellement imbriqué avec les mécanismes d'interprétation et de raisonnement de haut niveau que l'on doutait que des solutions réalistes (même très partielles) puisse intervenir avant la résolution du problème de l'Intelligence Artificielle (IA) dans sa globalité (l'histoire des impasses de l'analyse d'image est fortement liée à celle des illusions de l'IA). En outre, la jeune discipline du traitement automatique des images était caractérisée par un fort empirisme et la production de solutions ad hoc à des problèmes spécifiques (ce que Marr (1977) appelle théories de Type 2) plus que par un souci de réflexion globale (Type 1). En proposant que l'«intelligence» d'une scène - au sens de sa modélisation sémantique de haut niveau - puisse être repoussée à une étape tardive du traitement de l'information visuelle, la théorie de Marr a clairement délimité la discipline de vision par ordinateur - en particulier en la démarquant de l'IA - et l'a dotée d'objectifs propres et d'une démarche pour les atteindre.

Marr et ses collaborateurs se sont principalement attachés à imiter nos propres capacités visuelles, liées à nos *deux* yeux, qui nous confèrent une perception en trois dimensions de l'espace. Ils envisageaient essentiellement le problème de vision dans un contexte robotique : rendre un robot doté de deux yeux capable d'utiliser ses perceptions pour agir sur son environnement, s'y déplacer, etc. Ce problème de vision prend donc en entrée deux séquences temporelles d'images dont les champs visuels se recouvrent et inclut les problèmes de stéréoscopie et d'analyse du mouvement.

Dans ce cadre, Marr a proposé que le rôle d'un système de vision soit avant tout de produire une description tridimensionnelle de la scène observée. C'est sur cette «topographie» brute de l'espace, et sans retour aux perceptions primaires, que doivent s'appuyer les mécanismes de haut niveau qui conduisent à l'établissement d'une véritable «carte», formée de symboles qui renvoient à une légende.

L'optique et la photogrammétrie nous enseignent que le problème inverse de remonter aux causes du monde physique qui ont pu produire les observations contenues dans des images est un problème qui, sous sa forme générale, n'admet pas de solution unique, et ce quel que soit le nombre de vues disponibles. Marr

[17] Marr a été frappé par des tests d'aptitude effectués sur des patients ayant subi des lésions cérébrales, qui mettent en évidence la modularité de notre système de cognition visuelle : certaines lésions cérébrales spécifiques entrainent la perte de la capacité à reconnaître les objets perçus alors que subsiste la capacité à percevoir leur présence et à les localiser, à saisir leur forme (Marr 1982, pp. 34-36). Boucart (1996) relate d'autres expériences qui pointent également un caractère modulaire, mais à plus haut niveau cognitif. Certaines lésions rendent par exemple incapable de *nommer* les objets alors que subsistent les connaissances relatives à leur usage, fonction, mode de fabrication, etc.

postule alors que l'ajout de quelques hypothèses génériques sur la nature physique des objets observés ou sur leur aspect dans les images peut permettre de régulariser le problème :

1. Existence de surfaces lisses texturées,
2. Presque partout continuité spatiale,
3. Continuité des discontinuités (existence de lignes de discontinuité)
4. Continuité du flot optique.

Il ajoute également à ces hypothèses deux principes de nature différente :

5. Le principe de similarité relative (ou principe de contraste simultané) qui a trait à la manière dont nous pouvons remonter aux propriétés des objets réels à partir de leurs apparences (point de vue réaliste) ou à la manière dont nous pouvons émettre des jugements intersubjectifs sur les observations (point de vue positiviste). Ce principe est à la base de notre modèle de cocon et sera détaillé au chapitre 5.

6. Le principe Gestaltique d'organisation hiérarchique des percepts sur lesquels s'appuient notre vision, sur lequel nous reviendrons en 3.5.1 ci-dessous.

L'obtention d'une description 3D de la scène est alors décomposée en une série de descriptions intermédiaires, qui s'enchaînent en utilisant chacune certaines de ces hypothèses :

1. La «première ébauche» (*primal sketch*) est une description bi-monoscopique de la scène en termes de contours (*raw primal sketch*), de points caractéristiques et de groupements perceptuels de contours (*full primal sketch*).

2. L'«ébauche $2D\frac{1}{2}$» ($2D\frac{1}{2}$ *sketch*) explicite des éléments volumiques de la scène dans une représentation centrée sur l'observateur (distance à l'observateur, discontinuités en profondeur...). Les primitives sont encore peu structurées.

3. L'«ébauche $3D$» (*3D sketch*) est une représentation de la scène en termes de primitives volumiques génériques attribuées par des caractéristiques visuelles : couleur, texture, etc.

Discussion

On réduit souvent la théorie de Marr à l'idée que la vision puisse procéder par un mécanisme purement ascendant, allant «des pixels aux prédicats» (Vézien 1995). Horaud et Monga (1995, p.19) écrivent par exemple que «dans la pratique, le paradigme de Marr se traduit par trois étapes : segmentation, reconstruction et interprétation»[18]. Mettons l'accent sur quelques autres idées de Marr qui nous semblent importantes.

«La vision est un processus qui produit à partir d'images du monde extérieur une description qui est utile à l'observateur et qui ne s'embarrasse pas d'information non pertinente»[19]. Marr met ici en avant deux aspects centraux de sa conception de la vision :

En premier lieu, la description finale - comme celles produites par les étapes intermédiaires - représente une perte d'information par rapport au stimulus initial. La vision est un processus irréversible ; l'image d'origine ne peut pas être reconstruite à partir des descriptions élaborées. Quand, sur une carte, nous repérons le symbole d'un bâtiment, nous ne pouvons pas reconstruire l'aspect qu'aurait le bâtiment sur une image aérienne, disons à 50 *cm* de résolution. Le symbole sur la carte nous indique principalement «où est quoi». Si l'on disposait de l'hypothétique description mentale intermédiaire - e.g. en termes de lignes et de surfaces - qui a servi à élaborer la carte à partir de l'image aérienne, on pourrait se faire une image plus précise de ce bâtiment, mais qui serait encore une caricature de l'image aérienne. D'un point de vue informationnel, on peut envisager la vision comme une cascade de filtres : chaque étape produit une description de moins en moins fidèle aux perceptions brutes.

[18] Horaud et Monga entendent ici *segmentation* au sens large de toute opération de bas niveau d'extraction de primitives visuelles monoscopiques et le terme *reconstruction* fait référence à la modélisation tridimensionnelle de la scène.

[19] « Vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information» (Marr 1982, p.31).

Ceci nous mène au deuxième point, qui concerne le rôle de cette cascade, et celui de ses constituants individuels. Ce rôle est d'**expliciter l'information utile** à la compréhension d'une scène, ou plus généralement à la prise de décision, à la pensée ou l'action. Un système de vision doit donc être appréhendé comme un tout. Chaque opération intermédiaire n'est pas une fin en soi, et la pertinence de son résultat ne peut s'évaluer que par son utilité globale. Par ailleurs, l'expression «expliciter l'information» est à prendre au sens fort : dans l'idée de Marr, les mécanismes de vision correspondent essentiellement à des **changements de représentation**, à des recodages sous une forme différente, mieux adaptée à l'accomplissement des tâches ultérieures. Chaque recodage s'accompagne alors d'un filtrage des éléments qui apparaissent superflux *a posteriori*, du point de vue de la nouvelle représentation.

Mach résume ces différents points à travers l'idée de **dualité entre le traitement et la représentation de l'information**. Prenons quelques lignes pour commenter cette idée, qui nous parait fondamentale.

D'un point de vue mathématique, un même objet peut être représenté de plusieurs façons équivalentes. De nombreuses avancées, à la fois théoriques et pratiques, sont liées à l'exhibition d'une nouvelle manière de représenter un objet déja connu, qui met en avant un aspect différent de l'objet. Pensons aux révolutions qu'ont constitué la transformation de Fourier puis celle en ondelettes dans le domaine de l'analyse fonctionnelle[20]. Notons que l'utilité pratique d'une représentation est conditionnée par la capacité à l'obtenir rapidement. Ainsi, les analyses fréquentielles, puis en espace/échelle n'ont été utilisées pour le traitement automatique de données que quand on a pu disposer de transformées rapides (FFT, FWT). Mais avant l'aspect pratique, la manière dont *nous nous représentons* un objet, tel que la solution d'un problème, influe grandement sur la façon dont nous l'appréhendons. Dans la formalisation et l'étude d'un problème, les structures mathématiques jouent un rôle déterminant. La méthode d'analyse multi-échelles qui est proposée dans la troisième partie du mémoire s'appuie de manière fondamentale sur la structure de la solution recherchée et la façon dont elle peut être représentée.

D'un point de vue algorithmique, il est bien connu que les structures employées pour représenter des données influent sélectivement sur la capacité à accéder à l'information. Certaines structures favorisent l'accès à certains types d'informations, et d'autres structures privilégient l'accès à d'autres types d'informations. Un exemple particulièrement frappant de l'influence de la représentation est donné par la méthode de codage de graphes adaptée à la recherche d'isomorphismes proposée par Bruno Messmer et Horst Bunke (Messmer et Bunke 1995). Trouver les isomorphismes entre sous-graphes de deux graphes donnés est un problème NP-complet. Cependant, Messmer et Bunke ont montré qu'un prétraitement de coût exponentiel d'un graphe permettait de le représenter de telle façon que toute requête ultérieure d'isomorphisme de sous-graphes avec un autre graphe puisse être résolue en temps polynômial. La complexité calculatoire est donc reportée sur une étape de prétraitement, et est en quelque sorte factorisée car une fois le changement de représentation effectué, un nombre arbitraire de requêtes différentes peut être soumis. Bien entendu, la possibilité d'effectuer une telle transformation dépend de la nature du problème.

Le choix de représentations ou l'adoption de points de vues adaptés est un point clé qui conditionne le succès et/ou la rapidité de succès de n'importe quelle entreprise de résolution de problème. Nous pensons comme David Marr que traitement et représentation de l'information doivent être envisagés comme deux éléments indissociables.

3.3.3 Vision purposive et vision active

Le programme de Marr est évidemment extrêmement ambitieux et n'a pas abouti à ce jour.

Jugeant ce programme utopique et adoptant une attitude opératoire, les théories de vision **purposive**[21] et de vision **active** (Aloimonos et Badyopadhyay 1987) ont proposé de modéliser la vision comme un système dynamique qui résout successivement des problèmes spécifiques d'analyse, en réponse à des intérêts ponctuels. Pour ces théories, la cognition visuelle ne s'appuie pas sur des descriptions génériques

[20] Lire à ce propos le livre de Barbara Burke Hubbard (1995) : *Ondes et ondelettes. la saga d'un outil mathématique.*

[21] Purposive (anglais) : calculé, à dessein.

de l'espace environnant mais relève essentiellement de processus **descendants**, guidés par les modèles et par la tâche à accomplir. Ces approches mettent en avant le caractère dynamique, attentif, de notre vision (et de notre pensée). Notre regard se focalise tour à tour sur des aspects précis de l'environnement perçu, en réponse à des intérêts ponctuels. Par exemple, quand nous conduisons une voiture, notre vision passe successivement d'un niveau d'appréhension globale du trafic à l'observation d'un véhicule particulier, et se focalise ponctuellement sur certains détails très précis (un clignotant ...). En outre, notre manière d'appréhender le monde extérieur change selon nos dispositions mentales. Pour la conduite elle-même, seul le gabarit et le mouvement des véhicules a de l'importance, cependant notre attention passe parfois à d'autres registres : nous relevons la marque d'une voiture, apprécions sa couleur ou sa ligne. La théorie purposive n'exclue pas l'utilisation de procédures d'analyse de bas niveau des données - qui semblent indispensables, au moins dans un esprit cartésien de décomposition résursive de tâches en sous-tâches - mais pense que de telles procédures sont expressement dirigées par la tâche globale à accomplir.

L'idée de vision active est essentiellement une idée de robotique, qui met en avant l'importance de la coopération entre système moteur et système visuel dans l'exploration de l'environnement. Un robot mobile employant des stratégies de vision active peut par exemple décider de se déplacer pour voir une partie cachée d'un objet qui l'intéresse. L'idée de mouvement actif du regard ne se restreint cependant pas à la robotique. Une analyse active d'une image peut ainsi consister à rechercher un objet spécifique en un lieu précis pour corroborer la nature d'un autre objet proche sur lequel le système aurait un doute.

Ces théories renouent en quelque sorte avec l'approche traditionnelle de l'IA de résolution de problème : l'accent est mis sur la question du contrôle, sur l'utilisation apropriée de connaissances ou de méthodes de résolution, et sur la coopération de processus. On retrouve alors la difficulté classique : la résolution de problèmes réalistes nécessite de disposer et de savoir manipuler une quantité énorme de connaissances spécifiques.

3.3.4 Boucle analyse-synthèse

Également vers la fin des années quatre-vingt, l'idée de boucle **analyse-synthèse** a constitué une alternative à la démarche Marrienne de reconstruction 3D par des mécanismes purement ascendants. Cette approche, qui a particulièrement été développée dans le cadre du projet SYNTIM de l'INRIA, est liée aux progrès effectués dans les domaines de l'infographie et de la synthèse d'image. L'idée centrale est qu'analyse et synthèse d'une image sont deux opérations réciproques : l'analyse d'une image réelle produit un modèle «synthétique» et la synthèse d'un modèle «réel» crée une image synthétique. On peut donc imaginer une boucle de type prédiction-vérification : partant de l'image perçue, un premier modèle grossier est inféré à partir duquel une image de synthèse est produite, puis comparée à l'image réelle. Les défauts observés permettent de boucler sur une nouvelle phase d'analyse guidée par les données, ayant pour but de proposer des corrections au modèle. On peut espérer ainsi converger vers un modèle fidèle, une explication possible de l'origine des observations. De manière fort appréciable, cette théorie définit un critère de réussite clair et directement accessible (convergence de l'image de synthèse vers l'image réelle). Un point intéressant est que l'infographie nous renseigne sur les éléments nécessaires à la synthèse d'une scène : il faut modéliser les surfaces, leurs réflectances, textures, les positions et caractéristiques des sources de lumière, etc. Pour pouvoir participer à une boucle analyse-synthèse, un modèle doit pouvoir être *synthétisable*.

Telle qu'elle vient d'être présentée, la «reconstruction» d'une scène par analyse-synthèse procède par boucle itérative plutôt qu'en une seule passe ascendante, mais ne déroge pas à l'idée Marrienne de produire un modèle générique de l'environnement visuel. Cette école met cependant en avant le fait que des connaissances externes (des modèles d'objets connus) peuvent naturellement s'insérer dans la boucle (Vézien 1995).

Le principe d'inférence de modèle par un mécanisme d'analyse-synthèse, ou de prédiction-vérification, dépasse bien entendu le cadre de la modélisation 3D. Il a par exemple donné lieu à une méthode très intéressante pour la modélisation de textures par des processus stochastiques (Zhu, Wu, et Mumford

1996).

3.3.5 Conclusion

D'un point de vue neurobiologique, il est aujourd'hui pleinement attesté que le système visuel humain réalise un certain nombre d'analyses de manière purement mécanique (analyse des contours, de la couleur, perception de la profondeur, du mouvement...). Il semble que l'interaction de ces percepts élémentaires avec les zones de mémoire à long terme se fasse relativement tardivement, au niveau du cortex inféro-temporal (Boucart 1996).

D'après Jimenez (1997, p.22), «résumant des positions issues d'un grand nombre de travaux, on peut proposer (...) trois étapes générales de traitement de l'information perceptive : sensorielle, figurative et cognitive». Les psychologues de la perception semblent donc aujourd'hui s'accorder pour placer la «cognition» (l'interprétation) en aval d'une étape «figurative», qui correspond à ce que nous avons appelé l'analyse de bas niveau des sensations.

Du point de vue de la vision artificielle, certains traitements dont on pensait encore très récemment qu'ils relevaient en grande partie de processus «intelligents» s'avèrent pouvoir être simulés de manière réaliste par des mécanismes simples, n'utilisant aucune modélisation sémantique. C'est le cas par exemple du mécanisme de complétion des occultations, qui conduit Caselles, Masnou, et Morel (2001) à qualifier la vision de «machine géométrique».

La part «intelligente» du traitement de l'information visuelle semble donc pouvoir être repoussée à une étape relativement tardive, étayant ainsi - au moins partiellement - les thèses de la Gestalt et de David Marr (il n'est toutefois pas évident que les traitements de bas niveau puissent aller jusqu'à une reconstruction 3D complète). Deux questions se posent alors :

1) Jusqu'à quel point l'introduction de connaissances de haut niveau peut-elle être différée ? ou encore à quel degré de structuration générique de l'espace perceptuel peut-on parvenir uniquement sur la base de critères de bas niveau ?

2) Quelle forme doivent prendre les descriptions de bas niveau ? Quelles sont les conditions pour qu'elles remplissent, même partiellement, leur rôle de couches génériques d'analyse des perceptions ?

3.4 Traitements et analyses locales - points et contours

Il y a deux façons équivalentes de *décrire* un phénomène localisé (un ensemble d'un espace topologique correct) : en décrire les limites, les frontières ou en décrire les constituants, l'intérieur. Les descriptions «orientée contour» et «orientée région» d'un ensemble sont interchangeables (voir 4.3 p.53). Cependant, s'il s'agit d'*analyser* une image pour y localiser des phénomènes spatiaux, rechercher les limites des phénomènes ou rechercher les points qui participent aux phénomènes n'est en règle générale pas équivalent. Le seul cas où les deux approches - l'analyse des «transitions» dans les observations et celle des groupes d'observations «cohérentes» - coïncident strictement est celui où l'image analysée correspond au modèle simpliste d'une image constante par morceaux dans laquelle chaque zone plate[22] correspond à un objet d'intérêt. Identifier les lignes de changement de valeur ou les zones de constance définit dans ce cas les mêmes ensembles. Toutefois, dès que l'on s'écarte de ce cas de synthèse, une analyse «orientée contours» peut révéler des bribes de frontières à l'intérieur des objets ou ne pas révéler certaines portions de frontières entre objets. Les éléments de «contours» ne délimitent alors généralement pas des zones. De même, une analyse «orientée régions» peut mettre en évidence des zones «cohérentes» à l'intérieur des objets ou chevauchant des limites entre objets voisins. La dualité région-contours disparaît.

Bien que les approches «contours» et «régions» soient a priori deux approches complémentaires d'un même problème - mettre en évidence les structures significatives d'une image - elles ont connu des déve-

[22] Une zone *plate* d'une image définie sur un domaine topologique est une partie connexe maximale sur laquelle l'image est constante. Les zones plates d'une image la partitionnent (Salembier et Serra 1995).

loppements relativement distincts. Commençons par revoir dans cette section les approches «contours» et les théories qui en ont dérivé avant d'aborder le problème de délimitation de régions.

3.4.1 Transitions plus ou moins rapides

Les discontinuités d'une image d'intensité lumineuse idéale (non échantillonnée) sont des événements remarquables : elles correspondent nécessairement à une discontinuité dans un des phénomènes responsable de la formation de l'image. Si l'on suppose que la mesure (idéale, i.e. de résolution nulle) n'introduit aucun artefact discontinu, les discontinuités de l'image sont nécessairement dues à une discontinuité soit dans les caractéristiques de la surface ayant réfléchi la lumière (distance à l'observateur, réflectance, orientation de sa normale...) soit dans celles de la lumière éclairant la surface (ombre portée...).

Ce raisonnement s'appuie sur un principe fondamental que l'on doit à Pierre Curie : «Lorsque certains effets révèlent une certaine dissymétrie, cette dissymétrie doit se retrouver dans les causes qui lui ont donné naissance» (Curie, 1894, cité dans Lochak (1994)). Le complémentaire est faux : observer une symétrie dans les effets ne permet pas de conclure à l'existence de la même symétrie dans les causes. Ce principe de dissymétrie de Curie traduit évidemment la dissymétrie profonde entre cause et effet, c'est-à-dire l'orientation fondamentale d'une chaîne causale. «C'est la dissymétrie qui crée le phénomène», résume merveilleusement Curie.

La dissymétrie observée le long d'une ligne de discontinuité est un défaut d'isotropie locale de l'image : une dissymétrie entre le comportement de l'image dans la direction de la fracture et dans la direction orthogonale. Plus généralement, si l'on observe un gradient dans une image on peut conclure à la présence d'un gradient dans une des causes de la formation de l'image (la même conclusion tient pour le gradient de n'importe quelle caractéristique, par exemple texturale, calculée à partir de l'image par un procédé «honnête», i.e. qui n'invente pas d'information).

Notons donc une différence qualitative importante entre approches «contours» et approches «régions» : trouver une limite dans les observations assure l'existence d'une limite réelle, par contre ne pas en trouver (e.g. trouver une région «homogène») n'assure pas qu'il n'existe pas de limite dans le monde physique : plusieurs phénomènes peuvent se compenser et produire une apparence de non-phénomène.

Cependant, d'une part une image discrète est par nature partout discontinue, d'autre part, les transitions lentes sont également pertinentes, e.g. même dans une image idéale, la limite d'une ombre portée n'est pas ponctuelle mais diffuse. De manière plus générale, les «limites» des phénomènes qui nous intéressent peuvent avoir un aspect plus ou moins net ou flou, être plus ou moins localisées ou étendues. Marr fait alors «la remarque essentielle (...) que, à leurs échelles propres[23], ces objets peuvent tous être pensés comme s'ils étaient localisés spatialement» [24] : vue d'assez loin, une transition étendue paraît ponctuelle. Cette remarque a conduit Marr et Hildreth (1980) à proposer que l'analyse des transitions radiométriques devait s'effectuer à différentes échelles.

La diminution d'échelle est modélisée comme une opération de lissage de l'image d'origine, gommant progressivement les détails et simulant un éloignement progressif de la scène. Différents arguments conduisent Marr et Hildreth à un filtre gaussien isotrope pour réaliser cette opération. Si I est l'image d'origine, l'image «floue» I_σ d'«échelle» σ est obtenue par convolution $I_\sigma = G_\sigma \star I$ avec le noyau gaussien

$$G_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{||x||^2}{2\sigma^2}}$$

où $x \in \mathbb{R}^d$ et l'écart-type σ est appelé le paramètre d'«échelle»[25]. Outre les arguments neuromimétiques, un argument fondamental est que le filtre gaussien est l'unique filtre linéaire qui est localisé

[23] Marr entend ici «échelle» au sens d'extension spatiale caractéristique d'un phénomène.

[24] «the critical observation here is that, at there own scale, these things can all be thought of as spatially localized» (Marr 1982, p. 69).

[25] Notons que σ est homogène à une longueur et correspond rigoureusement à une notion de résolution d'analyse : plus σ augmente, moins l'on discerne de détails. Inversement, plus l'échelle d'une carte augmente, plus des détails fins apparaîssent. Tous les paramètres d'«échelle» dont nous parlerons dans ce document agiront dans le même sens que ce σ, c'est-à-dire

de manière optimale à la fois dans le domaine fréquentiel et spatial, ce qui indique qu'il est le moins susceptible d'introduire des variations qui n'étaient pas présentes dans l'image d'origine.

Marr et Hildreth ont alors proposé de localiser des «contours» aux lieux d'annulation du Laplacien de l'image lissée[26]. L'extraction de ces points de contour à quelques échelles différentes (2 ou 3), l'analyse des coïncidences entre les différentes échelles (nous reviendrons sur cette idée), le chaînage des points de contour en lignes et leur modélisation par des segments de droite conduit au *raw primal sketch*, qui est la couche de base du primal sketch.

Canny (1986) a par la suite formalisé le problème de détection de contours comme un problème de filtrage optimal. Un contour est modélisé comme une discontinuité rectiligne entre deux plages constantes dans l'image idéale (fonction créneau). L'image observée est échantillonnée et supposée noyée dans un bruit gaussien. Le filtre recherché doit alors donner une réponse unique, forte et localisée à la présence d'un créneau. Canny a montré qu'il n'y avait pas de solution analytique à ce problème dans le domaine des filtres à réponse impulsionnelle finie (FIR) mais qu'une bonne approximation était obtenue par convolution par la dérivée d'une ... gaussienne ! Notons qu'ici le rôle de la convolution gaussienne n'est pas de détecter les transitions douces mais de prendre en compte un bruit gaussien additif. Il est toutefois remarquable que les approches de Marr-Hildreth et Canny aboutissent à des solutions très voisines. Deriche (1987) a ensuite proposé une solution exacte au problème de Canny sous la forme d'un filtre à réponse impulsionnelle infini (IIR). Cet opérateur est alors très bien approximé par un filtre auto-régressif d'ordre faible, qui est très efficace. C'est ce filtre que nous avons utilisé pour estimer des «gradients» d'images numériques (il dépend d'un paramètre α qui règle l'extension caractéristique du filtre, voir la figure 3.1).

3.4.2 Théorie scale-space

L'idée de considérer l'échelle comme un paramètre *continu* est due à Witkin (1983), qui a proposé la première représentation **scale-space** (échelle-espace) d'une fonction. La théorie scale-space s'est par la suite considérablement développée pour aboutir à une théorie relativement unifiée.

Si I est une image

$$\begin{array}{rcl} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ x & \mapsto & I(x) \end{array}$$

une représentation scale-space de I est une autre image L

$$\begin{array}{rcl} \mathbb{R}^d \times \mathbb{R} & \rightarrow & \mathbb{R} \\ (x, \sigma) & \mapsto & L(x, \sigma) \end{array}$$

telle que $L(\cdot, 0) = I(\cdot)$.

La dimension supplémentaire modélise une *dimension d'échelle*. Witkin préconise lui aussi une convolution par une gaussienne, i.e. $L(x, \sigma) = (G_\sigma \star I)(x)$, en s'appuyant sur le fait que pour les signaux 1D, les extrema locaux de la fonction ne peuvent que s'annihiler au cours d'une augmentation de σ. Cette propriété est cependant fausse dès la dimension 2.

La formalisation du scale-space gaussien en dimension arbitraire a été réalisée par Koenderink (1984) avec l'introduction du **principe de causalité**.

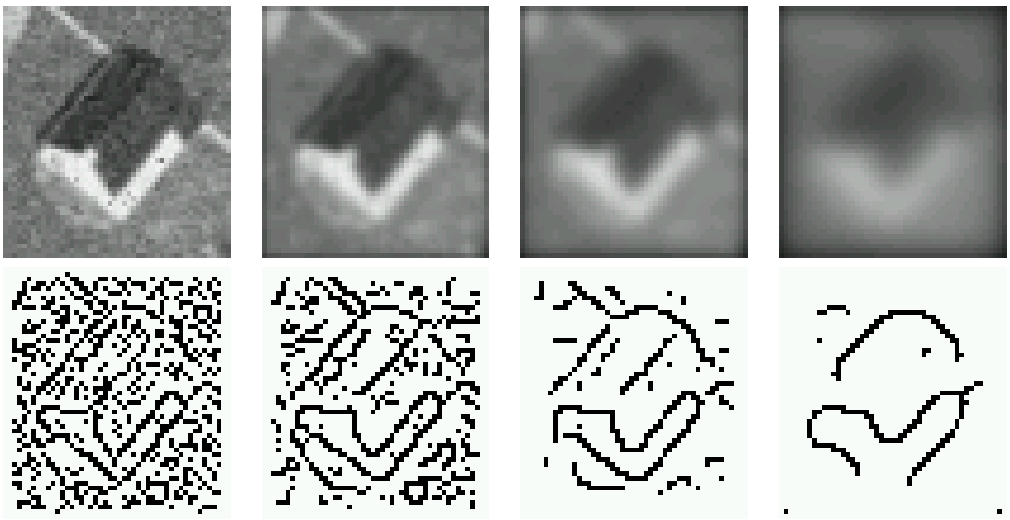
Ce principe stipule que les structures à échelle grossière doivent trouver une cause dans les échelles fines (σ faible). Koenderink propose que les «structures» d'une image sont ses courbes de niveau, ses isophotes, et traduit le principe de causalité par le fait qu'aucune nouvelle courbe ne doit apparaître quand σ augmente. Une isophote ne peut pas apparaître du néant pour construire dans le scale-space une surface dont la concavité serait dirigée vers les échelles croissantes. En ajoutant à la condition de

en sens inverse de la notion d'échelle cartographique. Toutefois, lorsque nous discuterons de la notion d'«échelle» de façon informelle, il faudra souvent entendre le terme au sens cartographique de précision d'analyse ou de description.

[26] Le Laplacien $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$, est l'opérateur différentiel isotrope d'ordre le plus faible.

causalité celles d'isotropie et d'homogénéité du scale-space et en se restreignant à des solutions linéaires, Koenderink démontre que la famille d'images est solution de l'équation de diffusion de la chaleur dans un milieu homogène, avec pour condition initiale l'image I. Le paramètre d'échelle est le temps du processus de diffusion. Il se trouve alors que la fonction de Green associée à l'équation de la chaleur est la... gaussienne ! La coupe à σ du scale-space s'obtient donc par convolution de I avec G_σ (voir la figure 3.1).

Notons que l'idée de scale-space continu est une construction théorique. En pratique, on est obligé de l'explorer par des *coupes*, de l'échantillonner en échelle. Si l'on doit réaliser une série de coupes, les progressions naturelles sont alors les progressions géométriques et l'origine naturelle des progressions est la résolution de l'image.



$-$ $\sigma = 1$ $\sigma = 2$ $\sigma = 4$

FIG. 3.1 – Coupes dans le scale-space gaussien et détection de contours. Dans la théorie du scale-space linéaire, l'image est convoluée par un noyau gaussien d'écart-type σ croissant. Nous montrons ici les résultats de l'application d'un filtre auto-régressif de Deriche pour trois valeurs de son paramètre α : 2 ; 1 et 0,5. Les résultats correspondent approximativement à des convolutions par des gaussiennes de paramètre $\sigma = 2/\alpha$. La seconde ligne représente les maxima locaux du module du gradient dans la direction du gradient qui ont une valeur supérieure à 10.

L'approche de Koenderink est purement axiomatique.

Le principe de causalité peut être considéré comme le principe de base des théories multi-échelles. Il traduit l'idée qu'une diminution de précision d'analyse (une augmentation de σ) ne peut pas créer d'information. Si au contraire, en diminuant la finesse d'analyse, on pouvait voir apparaître de nouvelles formes, qui ne sont pas le résultat d'une «dégradation» des formes existant à échelle plus fine, alors l'outil d'analyse serait inutile : rien ne permettrait de discerner un artefact d'analyse d'une véritable forme. Le principe de causalité stipule donc l'irréversibilité fondamentale d'une diminution d'échelle : on ne peut pas inventer les détails que l'on ne voit pas, l'axe des échelles est orienté (notons qu'il s'agit stricto-sensu de l'idée de dissymétrie de Curie). Dans le cas du scale-space gaussien, les formes analysées sont les isophotes, les courbes de niveau de la fonction[27]. Ces courbes de niveau se déforment continuement pour finir par disparaître à une certaine échelle. Il est important de souligner que la causalité porte sur les isophotes car il semble que l'idée que les *contours* - e.g. les maxima locaux du module du gradient - se raréfient nécessairement dans le scale-space gaussien soit assez répandue. C'est globalement vrai, mais pas systématiquement, de nouveaux bassins versant peuvent par exemple apparaître sur le gradient, voir

[27] Notons que les courbes de niveau d'une fonction constituent elle-mêmes une structure multi-échelles. Ce point sera développé en 5.5.2 p.87.

par exemple (Olsen 1996).

Par ailleurs, Koenderink s'appuie sur trois **invariances fondamentales** pour un traitement de bas niveau : les invariances par translation, rotation, et changement d'échelle. En effet, sans information a priori ni focalisation particulière, une opération de bas niveau doit être *indifférente* au lieu, à l'orientation et à la taille des structures (uncommited comme disent les anglo-saxons). La construction scale-space a par la suite été motivée à partir de l'idée de «frontal visuel» (visual front-end)[28], c'est-à-dire de couche générique de traitement initial d'une image, voir par exemple Lindeberg (1994). Les conditions d'invariance découlent alors directement de la condition de généricité du traitement.

La dernière condition requise par Koenderink, la linéarité de l'opérateur, correspond à requérir l'additivité des stimuli. C'est le point le plus criticable. Caselles, Coll, et Morel (1998) ont proposé de modéliser la formation d'une image comme le résultat de deux types d'opérations : occultations et transparences. L'occultation (masquage d'un objet par un autre) est de loin l'opération la plus fréquente et elle est hautement non linéaire (étant absolument non commutative, c'est peut-être même l'opération la plus non-linéaire possible).

En outre, un problème essentiel du scale-space linéaire est qu'en lissant l'image de manière homogène, il lisse également les transitions et délocalise les limites, comme le montre la figure 3.1.

Les modèles scale-space ultérieurs se sont alors développés en supprimant la condition de linéarité.

Perona et Malik (1990) ont proposé d'utiliser une forme non linéaire de l'équation de diffusion de la chaleur qui comprend un coefficient de conductance dépendant du gradient de l'image dans la direction de diffusion. La diffusion n'est donc plus isotrope, ni homogène. L'idée - qui peut sembler assez paradoxale - est de lisser les régions les plus lisses en priorité, et d'éviter le mélange de zones séparées par d'importants gradients, afin de préserver les discontinuités. La contrepartie de ce schéma de diffusion est que le temps de diffusion ne peut plus être relié de façon simple avec la dimension spatiale des objets analysés. En outre, différents paramètres doivent être réglés pour obtenir un résultat satisfaisant (forme de la fonction de conductance, nombre d'itérations...). Cette méthode connait néanmoins des applications importantes en détection de contours (Perona et Malik 1990), en segmentation (Dang, Jamet, et Maitre 1994; Kervrann, Hoebeke, et Trubuil 1999) ou en restauration d'images (Chanas 2001).

D'autres schéma de propagation régis par des équations non linéaires ont été proposés, dépendant par exemple de la courbure locale des lignes de niveau, comme dans le modèle d'Osher et Setian (1988).

Alvarez, Guichard, Lions, et Morel (1993) ont finalement procédé à une remarquable unification du domaine en montrant que toutes ces représentations scale-space pouvaient se déduire d'axiomatiques précises sur les opérateurs de traitement recherchés. Ces différentes solutions dérivent - souvent en tant que solution unique - de différents jeux de propriétés, principalement des propriétés d'invariance. Cette approche axiomatique aboutit à une classification complète des différents scale-space possibles, ce qui permet aux auteurs de dégager des schéma de propagation nouveaux qui «manquaient à l'appel». Un modèle important - proposé indépendamment par Sapiro et Tannenbaum (1993) - est le scale-space morphologique affine (Affine Morphological Scale-Space, AMSS) dont le nom provient du fait que c'est l'unique scale-space qui vérifie à la fois l'invariance par changement de contraste - qui est fondamentale en morphologie mathématique (Matheron 1975) - et l'invariance par transformation affine du domaine de l'image, qui est une invariance projective approximative.

3.4.3 Pyramides multi-résolution et ondelettes

Multi-résolution

Comme nous l'avons vu au chapitre précédent, résolution et échelle sont deux notions proches, mais distinctes. La résolution est une caractéristique d'une opération de mesure, d'observation, alors que l'échelle

[28] Dans un contexte informatique, l'adjectif *frontal* désigne «tout élément d'un système, qu'il soit matériel ou logiciel, qui sert d'interface conviviale à l'interaction entre l'utilisateur et le système, dont il cache la complexité» (office de la langue française, 2001). L'idée de *frontal visuel* est celle d'une interface générique «conviviale» permettant à des processus d'analyse arbitraires d'«interagir» avec les perceptions brutes.

est un attribut d'une description (ou modélisation). Les théories scale-space sont bien des théories multi-échelles, dans le sens où elles proposent de décrire la «structure» d'une image par une famille d'images simulant tout un continuum de perceptions, allant de la perception des structures les plus fines - ou locales - aux plus grossières - ou globales. Les structures présentes dans l'analyse deviennent de plus en plus grossières au fur et à fur que le paramètre d'échelle augmente mais la définition des images reste celle de l'image d'origine. Les formes à échelle grossière sont donc définies à la résolution d'origine (ce point est indépendant du problème de délocalisation des contours dans le scale-space gaussien).

A contrario, les analyses multi-résolution, dont l'idée remonte à Tanimoto et Pavlidis (1975), proposent de simuler les images que l'on aurait pu acquérir avec des capteurs moins résolus. Une pyramide multi-résolution est une séquence $(I_0, I_1 \dots I_p)$ d'images dont la taille décroît régulièrement, généralement d'un facteur 2 à chaque étape (pyramide dyadique). I_0 est l'image d'origine et classiquement, l'image I_{k+1} est obtenue à partir de l'image I_k par enchaînement de deux opérations : un filtrage passe-bas (moyennage sur un voisinage $l \times l$, lissage gaussien, filtrage morphologique...) et une opération de sous-échantillonage, généralement par «décimation», i.e. rétention d'un échantillon sur d.

Il y a donc deux différences essentielles entre une représentation scale-space et une multi-résolution :

1. Une multi-résolution propose une analyse selon un axe *discret* de résolutions alors qu'une représentation scale-space propose une analyse selon un axe *continu* d'échelles[29].

2. Une multi-résolution propose une séquence d'images de définitions décroissantes alors que le scale-space propose une séquence d'images de définitions identiques à l'image d'origine.

Les pyramides multi-résolution ont initialement été imaginées pour optimiser les traitements d'images. Elles ont ensuite connu des applications pour la compression et la transmission progressive d'images (Burt et Adelson 1983). Avant de s'intéresser plus particulièrement aux pyramides morphologiques, Florence Laporterie (2002) propose une bonne synthèse sur l'ensemble des techniques pyramidales.

Analyses en ondelettes

Les analyses en ondelettes continues sont des analyses multi-échelles, au sens où la projection sur une base d'ondelettes constitue un changement de mode de description des observations (la description est sans perte et non redondante quand la base est orthogonale). Les ondelettes sont les «mots» de la description, qui sont obtenus par décalages et homothéties à partir d'un mot primitif, l'ondelette «mère». La transformée en ondelettes discrète de Mallat (1989), qui s'appuie sur une multi-résolution dyadique correspondant au filtrage itéré selon la «fonction d'échelle» associée à l'ondelette «mère», peut être vue comme une sorte d'hybride entre analyses multi-échelles et multi-résolution.

3.4.4 Points caractéristiques

Pour terminer sur l'analyse locale et le filtrage multi-échelles ou multi-résolution, notons qu'hormis les points de variation rapide, d'autres points caractéristiques liés à la structure différentielle d'une image (idéale) s'avèrent intéressants (Schmid 1996). La répartition des points de Harris dans des images aériennes est par exemple suffisament informative pour réaliser des aéro-triangulations automatiques (Boldo et al. 2001). Dufournaud (2000, 2001) a récemment utilisé la théorie du scale-space linéaire pour définir l'échelle «propre» d'un point de Harris. Ces points d'intérêt qualifiés par une échelle (en fait par une ellipse : une orientation et deux extensions spatiales) permettent de mettre en correspondance des images d'une scène observée à des résolutions et selon des points de vue très différents.

Différents points d'intérêt peuvent également être définis à partir des contours ou des lignes de niveau d'une image : angles (discontinuité de la direction de la tangente le long d'un contour), terminaisons, jonctions (intersections de plus de deux contours) et particulièrement jonctions en 'T', caractéristiques d'un cas d'occultation d'un objet par un autre (Kanisza 1997; Caselles, Coll, et Morel 1998).

[29] On devrait d'ailleurs parler de représentation *pan-échelles*, à *toutes* les échelles, à propos du scale-space.

3.5 Segmentation en régions

Intéressons-nous maintenant à la segmentation, c'est-à-dire à l'analyse en régions des images.

Résumons tout d'abord les principaux traits des approches que nous venons de décrire. Nous sommes partis de l'idée d'extraction de contours, d'ensembles (1D) représentant les interfaces entres phénomènes perceptuels d'une image. En continu, l'idée de contour (local) est modélisée par des caractéristiques différentielles, e.g. l'existence d'un extremum dans les variations de la fonction. Le passage aux images numériques, l'existence de bruit et/ou le fait que les transitions lentes, plus globales, sont également pertinentes conduit alors à introduire un lissage de l'image avant l'analyse de ses extrema. Généralisée, cette idée donne lieu à la notion de scale-space. Les théories scale-space proposent de simuler la manière dont évolue l'aspect d'une image en fonction de la distance à laquelle on la regarde ou - ce qui est assimilé - en fonction du niveau de perception, local ou global, adopté. Le paramètre régissant l'évolution est globalement appelé «échelle» et modélisé comme une quantité continue. À côté du scale-space, nous avons distingué une deuxième grande famille de théories, qui proposent de modéliser l'effet de changements de résolution ou, ce qui est relativement proche, de modéliser une image comme une composition de signaux élémentaires de différentes résolutions (ondelettes).

Le point commun à ces différentes théories est qu'elles modélisent une image (une fonction) par d'autres images (d'autres fonctions), c'est-à-dire qu'elles constituent essentiellement des traitements d'images[30]. L'analyse à proprement parler, i.e. l'identification de formes, d'événements localisés, intervient a posteriori, par exemple par détection de «contours» ou seuillage des coefficients d'ondelettes.

Une segmentation en régions, en tant que description explicite d'une image en termes de phénomènes zonaux, constitue directement une analyse. Du point de vue zonal, la notion d'échelle est alors associée à une nouvelle idée : celle d'une **organisation hiérarchique des structures**. Intuitivement, augmenter une quantité qui se comporte comme une «échelle» d'analyse ne peut que raffiner la description, i.e. conduire à percevoir certains détails *internes* aux objets vus à l'échelle courante. La structure globale des unités présentes dans la description est donc une hiérarchie : les objets fins correspondent à des sous-parties de parties plus globales, composant elles-mêmes des objets de taille plus importante, etc. Commençons par argumenter la plausibilité d'un tel modèle d'organisation des zones significatives d'une image.

3.5.1 Hiérarchies de formes

La hiérarchie sémantique

«Dans le domaine de l'astrophysique extragalactique et de la cosmologie, un des principaux problèmes posés actuellement est celui de l'existence de structures hiérarchisées sur une large gamme d'échelles. Un enjeu important de cette science est ainsi la construction d'une théorie de la formation et de l'évolution des grandes structures de l'Univers.

Il est remarquable de rencontrer, dans d'autres champs que ceux de la physique, qu'on pourrait penser si différents a priori, en particulier dans le domaine biologique, le même type de problèmes et de préoccupations. Qu'on pense simplement à la structure hiérarchisée des organismes vivants complexes : ADN dans le noyau, noyau dans la cellule, cellule dans le tissus, tissus dans l'organe, organe dans l'organisme. La complexité est explicite ici à travers l'existence de niveaux d'organisation imbriqués, chaque niveau

[30] On peut faire la distinction suivante entre *traitement* et *analyse* d'image : un traitement est une opération sur une image qui produit une autre image, c'est-à-dire toujours une collection non structurée de valeurs (l'image produite n'est pas nécessairement de même nature que l'image d'entrée, e.g. l'estimation d'un gradient d'une image de réels produit une image de complexes). A contrario, une analyse produit un résultat différent d'une image, qui de manière assez générale consiste à expliciter certaines *relations*, soit entre différents points de l'image (e.g. appartenir à un même ensemble, tel qu'un «contour»), soit entre certaines parties, voire toute l'image, et d'autres données (e.g. des modèles). Notons que le résultat d'une analyse peut être physiquement représenté par une autre image (e.g. une image binaire de points de contours). Le résultat est alors implicite mais peut être explicité par un mécanisme d'analyse extrêmement simple, comme la rétention des points de valeur supérieure à un seuil donné.

étant caractérisé par l'existence d'une taille typique, souvent même matérialisée dans le cas du vivant par des parois ou frontières. Cette hiérarchie de structures se poursuit d'ailleurs au niveau de l'organisation sociale et politique» (Nottale 2001). Elle se poursuit également dans les modèles de l'organisation des phénomènes géographiques (Auger 1992). Plus généralement, deux lieux du monde physique auxquels nous attribuons une valeur sémantique (i.e. grosso-modo auxquels nous donnons des noms) sont en règle générale dans un rapport hiérarchique, c'est-à-dire soit disjoints, soit emboîtés. Pour s'en convaincre à échelle humaine, il suffit de faire l'expérience d'examiner son environnement visuel immédiat et d'identifier des parties de l'espace que l'on est capable de désigner par un substantif. On constate alors qu'il est extrêmement difficile de trouver deux objets A et B que l'on puisse nommer et qui possèdent une intersection propre, i.e. tels que $A \cap B \notin \{A, B, \emptyset\}$.

Notre décomposition du monde en unités sémantiques est essentiellement hiérarchique. Cette représentation a très certainement une origine physique, dans l'agencement des structures significatives de la nature ; mais les organisations hiérarchiques sont également associées à deux démarches intellectuelles complémentaires, la démarche analytique et la démarche constructive. Plaquons-nous une grille d'analyse hiérarchique sur le monde ou s'agit-il d'une propriété intrinsèque des phénomènes qui s'y déroulent ? Il est difficile de trancher... mais ce n'est pas ce qui nous préoccupe ici. Ce qui nous importe est que *nous* voyons globalement le monde de cette façon et que c'est cette vision que nous cherchons à reproduire.

Une conséquence fondamentale pour notre problème est alors l'extrême rareté des ensembles qui nous apparaissent pertinents[31] : une hiérarchie s'appuyant sur un ensemble à N éléments contient au plus $2N-1$ ensembles (dont $N+1$ ensembles triviaux : les singletons et l'ensemble entier) ce qui, comparativement aux 2^N ensembles possibles, est extrêmement faible. Cette linéarité est essentielle : elle signifie qu'il est combinatoirement envisageable d'expliciter *tous* les objets sémantiques visibles dans une image, le volume de la représentation est proportionnel à la taille de l'image.

Points de vue spécifiques

L'intuition que chaque point de l'espace peut être associé à la présence d'une unité sémantique *unique* est pourtant assez forte. Examinons-en deux raisons.

Tout d'abord, il y a ces unités sémantiques particulières que sont les objets *solides*. Les objets solides indépendants, libres de mouvement, sont effectivement mutuellement exclusifs et si on leur ajoute l'espace complémentaire on aboutit à une partition de l'espace physique. Le problème de l'interprétation *mécanique* d'une scène peut donc être posé comme un problème de partitionnement[32]. L'impression que les unités sémantiques partitionnent l'espace est quotidiennement renforcée par le fait que les objets de la vie courante, au sens commun d'objet manufacturé, sont des solides libres que nous regardons la plupart du temps comme des touts. Le point de vue mécanique n'est évidemment pas toujours pertinent. Par exemple, l'analyse mécanique d'une image de télédétection ne révèle généralement qu'un objet : la Terre !

Par ailleurs, nous réalisons effectivement *tour à tour* des associations verbales ou mentales univoques entre certains lieux et unités sémantiques. Quand nous pointons un lieu du doigt et demandons à une autre personne de nous dire ce qu'elle y voit, celle-ci ne répond jamais quelque chose comme : une voiture *et* une roue *et* un pneu *et* un crampon (Morel et Solimini (1995) indiquent que nous pouvons percevoir dans une scène usuelle environ quatre niveaux d'imbrication sémantique). Cependant, chaque association

[31] Pour étayer cette considération, faites l'expérience de tracer au hasard des ensembles sur une photographie. La probabilité pour qu'un de ces ensembles coïncide, même très approximativement, avec le contour d'un objet de la scène est pratiquement nulle. Évidemment, les ensembles que nous traçons naturellement à la main sont loin d'être quelconques (formes régulières, connexes, sans trou), mais tout comme le sont la plupart des objets qui nous intéressent...

[32] Remarquons que la modélisation mécanique d'une scène s'appuie sur des représentations mentales relativement élaborées, qui ne peuvent pas reposer sur la seule expérience visuelle. En particulier, il faut disposer d'un rapport *tactile* et *moteur* à l'espace, car l'identification d'un objet en tant qu'unité solide libre ne peut s'effectuer que par l'expérience effective de son déplacement, au moins par la pensée. Nous avons vu que Poincaré plaçait l'existence de solides à la base du concept même d'espace géométrique. Si les solides ont effectivement ce rôle fondamental dans la formation de nos représentations mentales, on comprend que les structures qu'ils induisent puissent avoir une telle valeur normative.

n'est univoque que *ponctuellement* dans le temps et le fait qu'elle le soit résulte simplement du fait que nous ne sommes pas capables de réaliser plusieurs associations simultanées, aussi bien par le langage que par la pensée. La *focalisation*, à chaque instant, sur un niveau unique de perception est guidée par un intérêt, une attention ou une intention spécifique. En outre, ces différents regards portés sur le monde peuvent changer très rapidement : rappelez-vous de la situation de conduite automobile que nous avons décrite ci-dessus.

Hiérarchisation perceptuelle ?

Peut-on transposer la structure hiérarchique observée au niveau sémantique au niveau purement perceptuel - c'est-à-dire au niveau des hypothétiques primitives Gestaltiques ? Non ; mais on constate que les contre-exemples à la hiérarchisation sont plutôt rares.

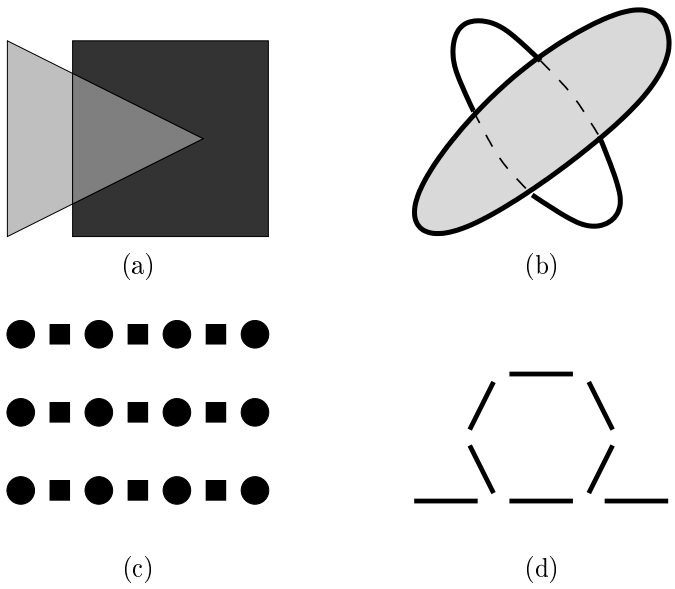


FIG. 3.2 – Régions perceptuellement pertinentes non hiérarchisées. (a-b) Cas liés à la formation de l'image. (a) Objets vus en transparence et ombres portées. (b) Occultation d'un objet par un autre. (c-d) Groupements conflictuels. (c) Motifs texturaux. Le groupement des points élémentaires en lignes horizontales, qui s'effectue essentiellement sur un critère de proximité, comme leur groupement en lignes verticales, sur un critère de similarité de forme, sont deux vision «orthogonales», hiérarchiquement incompatibles. (d) Partage de géométrie. L'hexagone et la ligne droite horizontale partagent un segment.

Les cas de conflits de hiérarchisation que nous avons pu identifier se rangent dans deux catégories principales :

1. Non hiérarchisation liée au processus de formation de l'image. Ces cas correspondent au modèle de formation d'une image proposé par Caselles et al. (1998).
   a. Objets transparents et ombres portées (figure 3.2(a)).
      • Des objets peuvent être hiérarchisés dans l'espace physique, sans que leurs empreintes dans l'image le soient, typiquement en cas de vision par transparence. Cette situation provient de la projection qui est réalisée par passage à l'image.
      • Une ombre portée n'est pas un objet du monde physique mais un artefact du procédé d'imagerie. Cet artefact possède cependant une réalité perceptuelle et la région qui lui correspond n'est a priori pas hiérarchisée avec les régions pertinentes auxquelles elle se superpose (on peut encore assimiler ce cas à un phénomène de transparence).
   b. Occultations (figure 3.2(b)).

Lorsque l'on complète mentalement un objet partiellement occulté par un autre, l'objet complété et l'objet occultant peuvent ne pas être hiérarchisés. Toutefois, cette non hiérarchisation est «virtuelle» (on peut à nouveau le voir comme un cas de vision par transparence). Nous interprétons d'ailleurs les objets comme deux objets disjoints en 3D, l'un devant l'autre, et non comme deux objets «interpénétrés». Un cas particulier intéressant est le cas de déconnexion par occultation : si les objets 3D pertinents sont connexes, l'occultation est le seul moyen d'aboutir à des unités pertinentes non connexes dans une image (la formation du célèbre triangle de Kanisza peut par exemple s'expliquer par compositions de processus de complétion amodale et de reconnexion - de parties réelles ou illusoires).

2. Gestalts partielles conflictuelles.
   a. Motifs texturaux (3.2(c)).
   Des cas de non hiérarchisation peuvent apparaître lorsque différents critères Gestaltiques (proximité spatiale, proximité de forme, de couleur...) permettent de placer un texton dans plusieurs groupes distincts «orthogonaux». Faisons toutefois deux remarques. En premier lieu, les textures d'objets réels qui présentent de telles propriétés sont principalement artificielles (motifs décoratifs, imprimés ...) et la décomposition de la texture globale en sous motifs texturaux ne présente pas d'intérêt essentiel. Les deux niveaux importants de description sont celui de l'objet global et le niveau fondamental, dans lequel tous les textons sont séparés. En outre, pour un homme, même si certains conflits de groupement se présentent, les différentes possibilités n'ont pas toutes la même saillance visuelle et les groupements les plus saillants ont tendance à *masquer* les autres (Kanisza 1997). Il existe alors toujours une décomposition hiérarchique privilégiée, qui suit à chaque étape l'ordre des motifs les plus saillants. Par exemple, sur la figure 3.2(c), les lignes horizontales ont une saillance visuelle plus forte que les lignes verticales.
   b. Partage de limite (3.2(d)).
   Il s'agit d'un cas particulier de groupement conflictuel. Il met toutefois en avant le fait qu'une hiérarchisation de régions n'implique pas une hiérarchisation de leurs contours. Ce point est à relier au problème de partage de géométrie en modélisation de bases de données géographiques[33].

Il est intéressant de noter que dans tous ces cas de non hiérarchisation, l'intersection (propre) entre les ensembles non hiérarchisés est systématiquement soit virtuelle (cas des occultations) soit elle-même une unité visuellement pertinente.

3.5.2 Segmentation guidée par les connaissances

Comme nous l'avons vu, les problèmes pratiques de segmentation en régions sont initialement posés comme des problèmes de «haut niveau», de délimitation de certains objets possédant un intérêt pour une application spécifique. Outre une question de localisation, se pose alors également une question de discrimination. Il y a globalement deux manières de concevoir la réalisation d'une telle tâche :

1) Introduire la connaissance spécifique dans le processus même de segmentation, i.e. coupler reconnaissance et localisation.

2) Différer l'introduction des caractéristiques spécifiques au problème à une étape ultérieure, i.e. envisager la segmentation comme une étape de bas niveau d'extraction de primitives régionales.

Examinons pour commencer le premier type de stratégie.

Template matching

Quand la classe d'images à analyser et/ou l'apparence des objets à identifier est suffisament contrainte, la «loconnaissance de formes» peut s'envisager par une approche purement descendante, i.e. par apparie-

[33] Notons que dans ces modélisations, aucun cas de «conflit de propriété» ne se présente comme un conflit pour un objet de nature *surfacique*, ce qui est bien entendu dû au modèle hiérarchique d'organisation des entités géographiques surfaciques.

ment direct d'un modèle des formes à reconnaître avec l'image (template matching, e.g. par corrélation ou transformée de Hough généralisée (Ballard 1981)). On peut par exemple traiter la reconnaissance de caractères d'imprimerie sur un document numérisé par template matching. On ne parle généralement pas dans ce cas d'un véritable problème de «segmentation» car la géométrie des formes à localiser est soit totalement déterminée par avance, soit décrite par un modèle paramétrique assez simple pour que la détermination des paramètres correspondant aux formes observées puisse être attaquée de manière purement analytique ou combinatoire. On parle plutôt dans ce cas d'un problème de recalage (matching).

Modèles déformables et ensembles actifs

La limite entre problèmes de «recalage» et problèmes de «segmentation» est toutefois relativement ténue : le recalage d'un modèle déformable, par une technique de contour actif (snake) ou de région active est par exemple généralement considéré comme une segmentation. La distinction tient essentiellement au fait que la gamme des formes géométriques que peuvent décrire ces modèles est extrêmement large, voire arbitraire. La détermination des formes effectivement observées est alors en grande part une opération ascendante, guidée par les données.

Introduits par Kass, Witkin, et Terzopoulos (1987), les *snakes* sont des courbes déformables dont l'objectif est de s'adapter au mieux aux contours de certains objets (il s'agit donc d'une approche contours, mais non locale). La qualité d'une solution - une position du snake - est mesurée par une énergie, que l'on cherche à minimiser. Cette énergie comporte deux types de termes : une énergie «interne» privilégiant les courbes régulières géométriquement, typiquement de faible courbure, et une énergie «externe» ou d'attache aux données, privilégiant typiquement le positionnement sur des points de fort gradient dans l'image (un snake peut également s'«attacher» aux lignes claires ou sombres de l'image, etc.). La minimisation est généralement effectué par itération de déformations locales du snake en suivant l'équation d'Euler-Lagrange associée au problème variationnel, ce qui assure de trouver un minimum local. Cohen (1991) a transformé les serpents en «ballons», en introduisant une force de «gonflage» dans l'énergie, visant à contrecarrer certains problèmes d'effondrement des snakes classiques. Notons également la méthode de Fua et Brechbühler (1996) qui permet d'imposer des contraintes «dures» à un snake : passer par un point, être tangent à une droite, présenter un angle droit, etc., ce qui nous ramène au fait que la limite entre patrons géométriques «rigides» et objets «mous» - i.e. entre recalage et segmentation - est relativement artificielle.

Les méthodes d'ensembles actifs se sont remarquablement enrichies avec l'introduction de la théorie des ensembles de niveaux (level sets) par Osher et Setian (1988), qui considère la représentation implicite d'un ensemble comme une *coupe* d'un ensemble de dimension supérieure. Cette représentation permet de gérer naturellement les «catastrophes» topologiques de connexion et de déconnexion. Les méthodes d'ensembles de niveau ont été initialement formulées en termes d'EDP modélisant la propagation de fronts, provenant de modèles de propagation d'incendies, mais elles peuvent également s'écrire sous forme variationnelle. Osher et Fedkiw (2001) proposent un panorama récent sur ces méthodes.

Les snakes et ensembles de niveaux classiques ne possédent qu'une énergie d'attache aux données localisée sur la frontière de l'ensemble. Plus récemment, ces modèles se sont transformés en *régions actives* qui intègrent (dans tous les sens du terme) des énergies sur les frontières et à l'intérieur des régions. Mariés avec les représentations par ensembles de niveaux, ces modèles conduisent à des méthodes de segmentation efficaces et robustes à l'initialisation (Paragios 2000).

Notons également que l'idée de courbe déformable s'étend à celle de graphe déformable (Jasiobedski 1993).

Ces méthodes sont cependant dédiées à la délimitation d'un seul ensemble (éventuellement non connexe quand il est envisagé comme un ensemble de niveau) ou d'un faible nombre, fixe, d'ensembles. Pour que chaque ensemble extrait corresponde à un objet d'intérêt il faut donc que la méthode soit focalisée, d'une manière ou d'une autre, sur la forme à trouver. Cette focalisation peut être implicite - par exemple provenir du fait que l'on sait par avance qu'il n'y a qu'un seul objet proéminent dans l'image - ou explicite - par

exemple consister en la donnée d'une solution approchée dans le cas des snakes classiques. De même, bien qu'une segmentation par région active soit plus robuste à l'initialisation géométrique, si l'image comporte plusieurs formes potentiellement pertinentes, la méthode doit nécessairement inclure des critères de discrimination spécifiques (e.g. l'objet à trouver est le seul objet en mouvement, possède une texture connue...). L'aspect générique de ces méthodes réside donc explicitement dans la capacité des modèles à prendre en compte des caractéristiques variées concernant les objets d'intérêt. Ces caractéristiques sont cependant intégrées «en dur», dans les énergies ou les équations d'évolution considérées.

L'approche morphologique «classique»

L'approche morphologique «classique» de la segmentation en régions (Beucher et Lantuejoul 1979; Beucher 1990; Meyer et Beucher 1991; Beucher et Meyer 1993) s'appuie également sur une stratégie de type «interprétation-segmentation», dans laquelle les connaissances spécifiques aux objets d'intérêt sont placées en amont de la technique de segmentation elle-même. Dans cette approche, interprétation et segmentation sont toutefois physiquement découplées ; elles correspondent à deux étapes successives d'analyse bien distinctes.

La partie «segmentation» repose sur un mécanisme systématique de partitionnement en zones d'une image (d'une fonction réelle sur un domaine topologique), à savoir le partitionnement selon ses *bassins versants*, les limites entre zones correspondant à la *ligne de partage des eaux* (*LPE*) associée à la fonction (voir l'annexe A p. 279).

La partie «interprétation» correspond alors à une étape préliminaire ayant pour rôle d'extraire des *marqueurs* des objets d'interêt, c'est-à-dire un ensemble de parties connexes qui soit en bijection avec les objets recherchés et tel que chaque partie soit strictement incluse dans un objet. La technique de LPE sert alors à déterminer la géométrie précise des frontières entre objets. Elle est le plus souvent appliquée au module d'un gradient de l'image d'origine, modifié de manière à ce que chaque marqueur donne lieu à un et un seul bassin versant le contenant (modification d'homotopie). Grosso-modo, les frontières trouvées passent par les contours les plus saillants entre les marqueurs.

La partie géométrique est donc systématique mais une procédure de marquage différente doit être imaginée pour chaque problème particulier, en exploitant ses spécificités (e.g. les objets à extraire se présentent comme des tâches claires dans l'image, de telles dimensions...). Cette approche suppose donc que l'on soit capable de traduire toute la connaissance nécessaire à la discrimination dans une étape préliminaire de localisation approchée des formes. La reconnaissance intervenant avant la segmentation, on voit donc qu'elle ne peut pas s'appuyer sur des critères liés à la géométrie des objets. En outre, sous sa forme classique, le critère de détourage des zones repose uniquement sur l'intensité du gradient sur les frontières, soit uniquement sur un critère de contraste local. Cette méthode «classique» de segmentation morphologique est cependant simple et méthodique. Elle a fait ses preuves pour résoudre un certain nombre de problèmes de reconnaissance automatique de formes relativement contraints par la nature des images et des objets à identifier, e.g. pour l'extraction de cellules en microscopie. Elle peut également être utilisée comme outil de segmentation semi-automatique, demandant la saisie manuelle des marqueurs.

Toutefois, pour traiter des scènes plus complexes, ou proposer des segmentations à usage moins dédié, certaines approches morphologiques récentes (Meyer 1999b; Meyer 1999a; Meyer 2001) mettent en avant la nécessité de différer l'introduction de connaissances spécifiques (au moins en partie, l'idée d'analyse de bas niveau totalement générique étant bien entendu un non-sens). Elles aboutissent alors des segmentations multi-échelles qui sont similaires sur la forme aux analyses ensembles-échelle que nous développons dans ce document. Nous les analyserons au moment opportun.

3.5.3 Stratégie segmentation-interprétation

Les approches que nous venons de décrire envisagent explicitement la segmentation comme une opération dédiée à une tâche spécifique. Elles sont apparues autour des années 90, plus ou moins à la suite d'un

constat d'échec des stratégies inverses, i.e. des stratégies segmentation-interprétation. Haton et Haton (1993, p.106-107) écrivent ainsi que «[dans les années 70 et 80,] un certain nombre d'expériences furent menées à partir du paradigme de segmentation-interprétation selon lequel le processus de vision consiste en l'interaction entre une segmentation de bas niveau de l'image et une interprétation symbolique de haut niveau des résultats de la segmentation. Ce point de vue n'a cependant pas donné les résultats espérés».

Nous pensons que les difficultés rencontrées dans ces expériences ne proviennent pas du paradigme de segmentation-interprétation lui-même mais de l'utilisation conjointe d'un autre paradigme : celui de **partitionnement d'image**.

Le paradigme de partitionnement d'image

Le fait de formuler la segmentation comme un problème de partitionnement d'image, i.e. de décomposition de l'image en régions disjointes, remonte aux années 70 et a acquis depuis lors un statut paradigmatique. Dans leur *glossaire des termes de vision par ordinateur*, Haralick et Shapiro (1991) écrivent par exemple : «*La segmentation d'image* est un processus qui habituellement partitionne le domaine spatial d'une image en sous-ensembles disjoints, appelés régions, chaque région étant uniforme et homogène par rapport à une certaine propriété telle que la teinte ou la texture, et telle que la valeur de cette propriété diffère significativement de celle de chaque région voisine»[34]. La fin de cette définition renvoie aux formalisations par prédicats d'homogénéité de la segmentation, qui datent des années soixante-dix et sont relativement obsolètes (nous les détaillerons au chapitre 5). Toutefois, même pour les formalisations plus modernes, e.g. variationnelles ou bayesiennes, que nous détaillerons au chapitre 6, l'objectif reste très majoritairement de déterminer une partition de l'image[35].

Or, comme nous l'avons vu, **il n'existe a priori pas d'unique partition pertinente d'une image**, aussi bien en terme de pertinence sémantique qu'en terme de pertinence perceptuelle. En effet, en règle générale des unités sémantiques sont visibles à différentes échelles dans une image (e.g. groupe, personne, visage, yeux). Partitionner une scène à haut niveau correspond donc à en faire une lecture à un degré de détail donné, une lecture «plate», orientée par un intérêt spécifique. De même, et peut-être plus encore, les unités perceptuellement pertinentes ne partitionnent généralement pas une scène. Par exemple, dans une image usuelle, pratiquement toute zone «cohérente» (e.g. possédant un certain degré d'homogénéité et maximale pour cette propriété) d'une certaine taille présente des sous-parties «cohérentes», des «tâches», des irrégularités internes qui possèdent une pertinence visuelle (nous pouvons les «voir») et peuvent présenter de l'intérêt (nous sommes susceptibles de focaliser notre attention sur elles). Un mécanisme de partitionnement sélectionne donc - d'une manière ou d'une autre - un sous-ensemble des formes visibles potentiellement pertinentes.

Poser les questions d'*interprétation* d'image comme des problèmes de partitionnement n'est pas très contraignant : la plupart des applications pratiques d'analyse d'image visent effectivement à réaliser une certaine lecture «plate» de la scène, à en dresser une carte comportant des entités mutuellement exclusives (e.g. une carte des personnes visibles). Cependant, même si le résultat final - de haut niveau - est une partition, et ce par la *nature* des objets recherchés, transposer cette contrainte à bas niveau pose des problèmes de fond.

[34] «*Image segmentation* is a process which typically partitions the spatial domain of an image into mutually exclusive subsets, called regions, each one of which is uniform and homogeneous with respect to some property such as tone or texture and whose property value differs in some significant way from the property value of each neighboring region.»

[35] Quelques auteurs - d'influences diverses - ont proposé, à partir des années 90, des méthodes de segmentation produisant un résultat plus riche qu'une simple partition (Montanvert, Meer, et Rosenfeld 1991; Jolion et Montanvert 1992; Koepfler, Lopez, et Morel 1994; Fuchs 2001; Serra et Salembier 1993; Meyer 1999a; Salembier et Garrido 2000). Nous décrirons ces propositions aux chapitres suivants, quand nous serons en mesure d'examiner les points communs et différences avec notre approche.

Problèmes de fond posés par le partitionnement de bas niveau

Supposons dans un premier temps que la segmentation ait pour rôle de délimiter *en une seule pièce* les objets qui intéressent une application donnée. Dans ce cas, le fait de réaliser une segmentation de bas niveau par partitionnement est incompatible avec la volonté de différer l'introduction des caractéristiques spécifiques aux objets d'intérêt à une étape ultérieure d'«interprétation». En effet, l'action de partitionnement impliquant un choix entre différentes unités potentiellement pertinentes, pour que la partition produite corresponde au résultat escompté, il faut bien que la méthode qui l'effectue intègre des connaissances spécifiques aux objets recherchés, c'est-à-dire intègre des connaissances que l'on se proposait de réserver à une étape ultérieure. La conclusion est simple : il y a une contradiction logique interne dans le fait de simultanément adopter une architecture Marrienne, repoussant l'introduction des connaissances spécifiques à l'étape d'interprétation, et une segmentation de bas niveau par partitionnement.

Une autre manière d'exprimer ce point consiste à remarquer que c'est la *nature* spécifique des objets recherchés qui induit le fait qu'ils partitionnent l'espace. Nous obtenons par exemple une description plate parce que nous ne cherchons pas à inclure simultanément les villes et les bâtiments ou les visages et les yeux dans la description. Si c'est la nature spécifique des objets qui entraine la structure de partition alors cette structure doit découler de la *reconnaissance effective* de la nature spécifique des formes qui composent la description.

Il y a deux façons de corriger l'inconsistance du schéma (partitionnement de bas niveau $\rightarrow$ interprétation) tout en conservant une segmentation par partitionnement :

i) Soit en revenant explicitement à un couplage entre segmentation et interprétation. Il faut alors donner à la segmentation les moyens effectifs de discriminer les objets qui intéressent l'application.

ii) Soit en relâchant la contrainte de correspondance biunivoque entre les régions segmentées et les objets d'intérêt de l'application.

Une autre façon de corriger cette inconsistance est d'*abandonner la contrainte de rechercher une partition à bas niveau.* C'est l'approche que nous proposons. Examinons tout d'abord les deux autres approches afin de dégager les intérêts de cette solution.

La stratégie i) revient à concevoir la segmentation comme une opération dédiée. Pour que l'approche soit concluante, il faut alors que la méthode de segmentation soit effectivement apte à prendre en compte des caractéristiques discrimantes concernant les objets d'intérêt. Or, en pratique, les algorithmes de partitionnement d'image s'appuient sur des critères de cohérence de forme extrêmement génériques : des critères d'«homogénéité», de «régularité» ou de «cohérence» texturale des valeurs dans chaque région (tout ceci à un certain degré), et des critères de taille ou des critères très généraux de «simplicité» géométrique de la silhouette de chaque région (rapport isopérimétrique faible, courbure faible...), etc. De toute évidence, ces critères ne permettent pas a priori de discriminer différentes classes d'objets sémantiques. Être homogène à un certain degré n'est par exemple pas une caractéristique propre à l'aspect de certains objets précis, e.g. propre à l'aspect des routes ou des champs. Le fait de pouvoir posséder - toujours à un certain degré - une des caractéristiques citées ci-dessus tient essentiellement aux conditions particulières d'observations de l'objet. C'est alors uniquement la *restriction à un contexte applicatif très spécifique qui est susceptible de rendre ces critères discriminants.* Précisons : une méthode d'analyse d'image est destinée à être appliquée à un sous-ensemble $\mathcal{J}$ de l'ensemble $\mathcal{I}$ des images possibles. $\mathcal{J}$ comporte des images d'un certain type, qui ont été acquises dans certaines conditions particulières et qui représentent certaines scènes bien particulières (e.g. images microscopiques de grains d'alumine). En outre, on cherche à identifier certaines formes spécifiques $\mathcal{O}$ dans ces images (e.g. les grains). Appelons $\mathcal{C} = (\mathcal{J}, \mathcal{O})$ le *contexte applicatif*[36]. La question est alors de savoir s'il existe un *réglage correct* de l'algorithme de partitionnement, c'est-à-dire un réglage fixe de l'algorithme qui assure que les objets d'intérêt soient

[36]Cette terminologie est à relier à l'idée de «connaissance contextuelle» de la page 21.

systématiquement délimités en une seule pièce, pour toute image de $\mathcal{J}$[37]. Pour qu'un tel réglage existe il faut alors que le contexte $\mathcal{C}$ soit suffisamment restreint pour contraindre drastiquement l'aspect des objets recherchés dans les images considérées : il faut que ces objets se présentent systématiquement dans la même gamme de tailles, de contrastes, textures, complexité de silhouette ... Ces conditions peuvent être réalistes dans des contextes expérimentaux très restreints (analyse d'images très spécifiques, présentant des scènes simples, etc.), cependant, dès que les images, leur sujet ou ce que l'on recherche possède un minimum de variabilité, l'existence d'un réglage correct devient fort improbable.

Par exemple, si dans l'ensemble des images de $\mathcal{O}$, un objet d'intérêt peut être vu à une distance ou dans des conditions d'éclairage variable, alors des critères de taille ou de contraste absolus ne sont pas discriminants. De même, vu d'assez près, un objet peut apparaître texturé, mais si l'on s'en éloigne, il peut sembler lisse. Un critère d'homogénéité n'est alors pas discriminant.

Notons que ce type de caractéristique peut varier non seulement d'une image à l'autre, mais également au sein d'une même image. Plusieurs instances d'un même objet situées en différents lieux d'une image peuvent par exemple avoir des contrastes très différents, typiquement si une instance est au soleil et l'autre à l'ombre. Donc, *même pour une seule image*, il n'est absolument pas évident qu'il existe un réglage correct d'un algorithme de partitionnement s'appuyant sur des critères génériques.

Enfin, même dans le cas où les conditions d'observation sont relativement stables, plusieurs représentants de la classe des objets d'intérêts peuvent avoir des attributs visuels très différents. L'aspect des parcelles agricoles sur une orthophotographie aérienne à 50 cm de résolution est un bon exemple : certaines parcelles sont claires, d'autres sombres, certaines sont très homogènes, d'autres très texturées (champs labourés, vignes, vergers, bois), certaines parcelles ont des formes très régulières, e.g. rectangulaires, d'autres ont des formes très tortueuses, peut-être même fractales sur une certaine gamme d'échelles (limites de certains bois). Dans ce contexte, on conçoit bien que les objets pertinents doivent être recherchés dans différentes gammes de contrastes, tailles, textures, etc.

En pratique, dès que l'on s'écarte de conditions expérimentales que l'on pourrait dire «de laboratoire», partitionner une image sur des critères liés à l'aspect des formes (critères qui en outre sont généralement très frustres) conduit aux difficultés qu'évoquaient Haton et Haton. Il est pratiquement impossible de trouver un réglage correct[38] ; selon l'image et même la position dans la scène, certains objets d'intérêt sont sur-découpés ou d'autres sont fusionnés à une partie de leur environnement. Si l'étape d'interprétation aval fait l'hypothèse que les unités auxquelles elle s'intéresse correspondent systématiquement à une seule région (e.g. si c'est une simple classification des régions segmentées), alors l'ensemble de l'analyse échouera pour un certain nombre d'objets d'intérêt. L'échec de la stratégie partitionnement-interprétation a trait à deux questions extrêmement sensibles en analyse automatique d'images : la question du paramétrage des algorithmes et celle, intimement liée, de l'invariance des algorithmes.

Appuyer une interprétation d'images sur un partitionnement de bas niveau a bien entendu connu des succès dans certains contextes applicatifs suffisamment contraints. Toutefois, le «challenge» actuel de l'analyse d'image est de dépasser ce type de contextes, et par exemple de pouvoir analyser automatiquement des images collectées au hasard sur internet, ou des séquences vidéo variées. Dans de tels contextes, l'invariance des algorithmes aux conditions d'observations est fondamentale ou encore, le succès d'une analyse sur une image ne doit pas reposer sur le réglage ad hoc de paramètres.

On peut alors purement et simplement abandonner l'idée de segmentation de bas niveau en régions, et juger que ces techniques de segmentation achoppent parce qu'elles ne prennent en compte que des critères trop frustres, précisément des critères de bas niveau, relatifs à certaines apparences ponctuelles

[37] Notons en passant qu'une méthode de partitionnement sans paramètre spécifie d'une manière ou d'une autre un niveau de lecture «optimal» d'une image. La méthode est donc nécessairement dédiée à une application unique. Cette remarque nous conduira ci-dessous à une analyse critique du principe orthodoxe de Minimum Description Length (MDL) de Rissanen (1983) appliqué au partitionnement d'images.

[38] Même si un tel réglage existe, en trouver un est un problème en soi, dont la difficulté augmente bien entendu avec le rapport entre la taille de l'espace des paramètres et la taille de l'ensemble des paramétrages corrects.

possibles des objets dans les images et non à des invariants des formes. Comme il n'existe certainement pas d'invariant d'aspect qui soit commun à tous les objets sémantiques, on revient alors à l'idée d'incorporer certaines connaissances spécifiques dans le processus même de segmentation, soit à une approche globalement descendante, supervisée, ou guidée par les modèles.

Sans abandonner totalement l'idée d'analyse de bas niveau, une autre solution est de restreindre les ambitions de l'étape de partitionnement, en adoptant l'approche notée ii) ci-dessus.

Conçue comme une première étape d'une séquence d'opérations devant conduire à une interprétation, la segmentation n'a évidemment d'intérêt que si elle permet de progresser dans l'accomplissement de l'objectif global, c'est-à-dire si elle facilite l'exécution des tâches ultérieures. Sous un angle purement combinatoire, on peut envisager le rôle d'une segmentation comme celui de réduire l'espace des localisations possibles d'objets pertinents. Une approche naturelle est alors d'adopter une stratégie *constructiviste*, à la David Marr, et de cantoner le rôle d'une opération de segmentation à celui d'une extraction de primitives. Les objets d'intérêt devront dans ce cas être recherchés sous forme de *compositions* des primitives régionales trouvées à bas niveau. Si l'on envisage toujours la segmentation sous forme d'une partition en régions, il s'agit alors de trouver systématiquement une *sur-segmentation* des objets d'intérêt, à partir de laquelle tout objet pourra être déterminé sous la forme d'une union de primitives. On retrouve cependant les problèmes de paramétrage qui viennent d'être décrits. Si l'aspect des objets d'intérêt est assez variable, il peut être difficile de trouver un réglage qui assure d'obtenir systématiquement une sur-segmentation. La seule solution sûre consiste à considérer ce que l'on peut appeler la partition en «atomes perceptuels» d'une image : si l'on admet l'hypothèse de hiérarchisation des régions visuellement pertinentes alors il existe dans toute image des *régions primitives élémentaires*, qui sont les plus petites régions pertinentes perceptibles, celles pour lesquelles on ne peut pas distinguer de sous partie pertinente. Ces «atomes» perceptuels constituent une sorte d'échelle fondamentale de lecture d'une scène. Cette segmentation atomique correspond cependant à une importante sur-segmentation par rapport aux échelles typiques d'interprétation recherchées. Les objets texturés y sont par exemple représentés comme une constellation de petites primitives visuelles. On voit donc qu'envisager la partition atomique comme le résultat de l'analyse générique de bas niveau prive cette étape d'une grande part de son intérêt pratique[39].

Si l'on souhaite développer une analyse de *bas niveau* en régions qui soit à la fois *fiable* et *utile*, il faut donc abandonner la contrainte de partition sur les régions produites. On aboutit alors naturellement à l'idée de segmentation multi-échelles et ... à la conclusion de ce chapitre.

3.6 Conclusion : la thèse proposée

David Marr a proposé d'asseoir les mécanismes de vision sur une description de bas niveau des images en termes de *contours* et a montré l'importance de s'intéresser aux transitions de différentes échelles. Les idées de Marr ont eu une influence importante sur les recherches ultérieures, qui ont débouché sur plusieurs théories multi-échelles extrêmement cohérentes : scale-space, ondelettes[40].

De son côté, l'analyse de bas niveau en régions des images - la segmentation - est traditionnellement conçue comme un problème de partitionnement d'image. Une telle formulation trouve son origine dans une confusion ancienne entre bas et haut niveau d'analyse visuelle, entre perception et interprétation. Or, les critères employés en segmentation sont bien des critères de bas niveau et non des critères de discrimination sémantique : ils sont absolument *génériques*, i.e. ne sont l'apanage d'aucune classe sémantique particulière, et relatifs à l'*apparence* des objets dans les images, et donc très dépendants des conditions d'observation (résolution, éclairage, etc.) - contrairement par exemple à notre capacité à reconnaître les objets. Par ailleurs, il existe en règle générale dans une image des unités pertinentes emboîtées les unes dans les autres. En conclusion, les méthodes de partitionnement d'image ne sont ni des méthodes robustes

[39] La segmentation atomique correspond grosso-modo aux zones plates d'une image ou au résultat obtenu par calcul d'une ligne de partage des eaux sur un gradient fin (cf. annexe A p.279).

[40] Lire par exemple Yves Meyer (2000) à propos de l'influence de Marr sur la théorie des ondelettes.

de reconnaissance de formes spécifiques, ni des méthodes génériques d'extraction de primitives. Nous y voyons l'explication des difficultés rencontrées pour résoudre l'une ou l'autre de ces tâches en s'appuyant sur une segmentation (dès que l'on s'écarte de contextes de RdF suffisamment contraints).

Nous proposons de transposer rigoureusement le paradigme Marrien au cas d'une analyse de bas niveau en régions. Nous aboutissons alors à une idée identique à celle de *«frontal visuel»* de la théorie scale-space : pour convenablement jouer son rôle d'interface *générique* entre les données et les processus de haut niveau d'interprétation, une segmentation de bas niveau doit relever les détails les plus ténus comme les plus contrastés, les plus fins comme les plus grands, les plus homogènes comme les plus texturés... bref, être *indifférente à l'apparence* des phénomènes. La sélection des unités pertinentes *pour une application donnée* doit s'effectuer *a posteriori* sur la base de critères fiables de reconnaîssance d'objets, qui sont nécessairement *spécifiques*. Une analyse de bas niveau en régions doit donc - au minimum - prendre la forme d'une **description hiérarchique**, une famille d'ensembles imbriqués qui modélise l'existence d'unités «cohérentes» à différentes «échelles». Réciproquement, comme nous l'avons argumenté, une telle structure semble être une bonne approximation de l'organisation de *toutes* les zones perceptuellement pertinentes d'une image.

Il y a alors un élément fondamental que se doit d'identifier toute théorie multi-échelles, à savoir le **paramètre d'échelle** lui-même, le paramètre monodimensionnel qui permet d'engendrer la description, puis de l'explorer. Les différentes théories multi-échelles dont nous avons parlé ci-dessus identifient toutes clairement ce paramètre (facteur de dilatation d'une ondelette mère, variance d'une gaussienne, temps d'un processus de diffusion ou de déplacement de courbe...). Outre l'aspect structurel - l'organisation hiérarchique des parties - une théorie multi-échelles de l'analyse en régions se doit également de définir l'«échelle».

Nous trouverons nos paramètres d'«échelle» directement dans les paramètres habituels des méthodes de partitionnement d'image : les seuils sur l'«homogénéité» des régions ou les poids de certains termes énergétiques qui règlent la «complexité» ou la «fidélité» à l'image de la solution trouvée. Mais, alors que dans les approches par partitionnement, ces paramètres sont *fixés a priori*, nous proposerons de les considérer comme des *variables* et d'étudier à quelles conditions ils engendrent effectivement des descriptions multi-échelles, descriptions que nous proposerons alors de calculer *in extenso*.

Comme nous le verrons, *expliciter* une dimension d'«échelle» permet de résoudre la délicate question du **paramétrage** des algorithmes de segmentation de bas niveau, et d'obtenir des descriptions **invariantes** par différentes transformations sur les images, deux problèmes qui conditionnent directement la robustesse d'une chaîne d'analyse d'images s'appuyant sur une analyse de bas niveau en régions. L'idée de segmentation multi-échelles n'est donc pas simplement une «belle» idée théorique mais touche à un problème extrêmement concrèt : réaliser des systèmes **fiables** d'interprétation d'images.

L'idée d'analyse multi-échelles touche également à un rôle fondamental de la vision de bas niveau, à savoir son rôle **exploratoire**. Tel que nous le concevons, le rôle des analyses de bas niveau est essentiellement de *découvrir* les formes perceptibles dans l'image, et de les décrire aux mécanismes de haut niveau, chargés d'interpréter les percepts en termes de ce qui est présent dans le monde. Comme nous le verrons, ne pas fixer *a priori* les paramètres d'un algorithme de segmentation peut permettre de trouver *a posteriori* les valeurs des paramètres qui sont pertinentes *pour l'image observée*. Nous proposerons que les formes «significatives» d'une image sont les formes **stables** en échelle, celles qui apparaissent sur d'importantes gammes de valeur du paramètre d'«échelle». Nous verrons qu'une telle approche peut permettre de répondre à une question aussi simple que : «y a-t-il quelque chose à voir dans l'image ?». Bien entendu, certaines scènes peuvent ne comporter qu'une seule échelle pertinente d'analyse. Mais encore une fois, ce n'est qu'*a posteriori*, après avoir *vu*, analysé l'image, que l'on peut en décider.

Finalement, une des caractéristiques principales de la vision - et de la pensée - est son aspect **dyna-**

mique, **attentif**. Une approche multi-échelles de l'analyse de bas niveau en régions nous semble nécessaire si l'on souhaite mettre en oeuvre des idées de vision active ou développer des mécanismes attentionnels s'appuyant sur des primitives régionales, c'est-à-dire si l'on souhaite parvenir à des modélisations réalistes de la vision.

Les deux modèles développés ci-dessous permettrons d'étayer ces idées générales.

Le modèle de cocons de la deuxième partie montrera comment en tentant de qualifier de manière «objective», i.e. grosso-modo non paramétrique, certaines zones cohérentes d'une image on aboutit à une hiérarchie de structures dans laquelle l'on navigue grâce à un paramètre de contraste.

La théorie ensembles-échelle développée dans la troisième partie du document établit une sorte de réciproque. Nous montrerons comment en imposant une structure hiérarchique aux solutions de problèmes classiques de modélisation d'image par parties formulés en termes d'optimisation, on parvient à différer totalement le réglage d'un paramètre intrinsèque à ces formulations, en lui conférant le sens d'une échelle. Ce paramètre d'échelle reliera la taille ou la complexité des formes décrites à leur contraste ou leur texture. Naviguer sur l'axe des échelles permettra alors à la fois de «rendre clair l'obscur et simple le complexe». Nous aboutirons à une méthodologie d'analyse non paramétrique (une fois une énergie spécifiée) dont le résultat est une description multi-échelles ensembliste d'une image, un «frontal visuel» orienté régions qui représente le pendant structurel d'une description scale-space.

Pour commencer, le chapitre qui suit revoit différentes structures mathématiques et informatiques qui entreront dans nos modèles.

Chapitre 4

Représentations mono et multi-échelle pour la segmentation d'images

Ce chapitre rappelle les définitions et les principales propriétés des structures qui nous seront utiles par la suite. Par structures nous entendons à la fois les structures mathématiques (abstraites) utiles pour la modélisation et l'étude de notre problème (graphes, treillis, hiérarchies, hiérarchies indicées et espaces ultramétriques) et les structures informatiques utiles pour la manipulation concrète de segmentations (essentiellement instances particulières de graphes attribués : graphe de frontières, d'adjacence de régions d'une partition, arbres...).

Ce chapitre traite donc de questions de représentations - abstraites ou concrètes - de familles d'ensembles dotées de structures particulières.

La plupart des notions qui sont abordées ici sont classiques - nous nous sommes principalement inspirés de Serra (1982, Ch. 1), Simon (1985, Ch. 4) et Schmitt et Mattioli (1994, Ch. III). Toutefois la terminologie et les notations varient d'un auteur à l'autre. Cette présentation nous permet d'introduire les notres. Par ailleurs, certaines notions ou façons d'envisager les choses nous semblent originales. Nous insistons également sur certains aspects qui nous paraissent parfois mal connus ou mal compris dans le domaine du traitement d'images - concernant en particulier les relations qu'entretiennent entre elles différentes structures (e.g. partitions et hiérarchies) - et qui seront centraux pour nos développements ultérieurs.

Ce chapitre est par nature relativement aride - il s'agit essentiellement d'une succession de définitions et de propriétés[1] - et un lecteur familier avec les objets concernés pourra le parcourir relativement vite.

4.1 Généralités - graphes et topologie discrète

Dans l'ensemble de ce document, le symbole $\triangleq$ signifie «égal par définition».

Si X est un ensemble, nous notons $|X|$ le cardinal de X et $\mathcal{P}(X)$ l'ensemble des $2^{|X|}$ parties de X.

Si X est muni d'une topologie $\mathcal{T}$, nous notons $\mathcal{P}_C(X)$ l'ensemble des parties de X *connexes* pour $\mathcal{T}$.

Rappelons qu'un graphe orienté (resp. non orienté) est un couple $G = (X, U)$, où X est un ensemble dénombrable (fini pour nous) dont les éléments sont appelés les nœuds du graphe et U est un ensemble de couples (resp. de paires) de nœuds, appelés arêtes du graphe.

Les graphes sont les structures combinatoires qui permettent de modéliser des relations binaires entre éléments d'un ensemble discret. En conséquence, si $\mathcal{R}$ est une relation binaire sur X, nous noterons $(X, \mathcal{R})$ le graphe orienté (X, U) dans lequel $U = \{(x, y) \in X^2 \mid x\mathcal{R}y\}$.

[1] Dans la suite du document, les principales définitions, propositions, etc. seront explicitement déclarées et numérotées. Nous n'avons pas suivi cette présentation dans ce chapitre car il en aurait été considérablement alourdi.

Si X est un ensemble discret - typiquement le domaine d'une image numérique - une topologie sur X sera définie à partir de voisinages élémentaires - tel que c'est fait pas exemple dans (Maître 2003, ch.3) - et non à partir de la définition d'ouverts. Une topologie sur X sera donc spécifiée par la donnée d'un graphe $G = (X, V)$ symétrique (ou non orienté) et réflexif (possédant toutes les boucles (x, x)). Deux nœuds x et y voisins dans G seront notés $x \leftrightarrow y$. La connexité sur X correspondra alors à la connexité de chemin dans G (un chemin est une suite de nœuds $(x_1, x_2, \ldots x_p)$ telle que $x_1 \leftrightarrow x_2 \leftrightarrow \ldots x_p$).

4.2 Treillis des parties et des partitions d'un ensemble

Soit X un ensemble et $A \subset \mathcal{P}(X)$ un ensemble de parties de X.

Nous disons que A est une **pré-partition** de X si tous ses ensembles sont deux-à-deux disjoints.

Une **partition** est alors une pré-partition *maximale*, i.e. une pré-partition à laquelle aucun ensemble ne peut être ajouté sans rompre sa propriété de pré-partitionner X. Une partition est donc un ensemble de parties qui recouvrent X sans s'intersecter (définition usuelle).

Nous notons $\mathbb{P}(X)$ l'ensemble des partitions de X et si X est doté d'une topologie, $\mathbb{P}_C(X)$ désigne l'ensemble de ses partitions constituées d'ensembles **connexes**. Les parties d'une partition sont aussi appelées *classes* si elles sont conçues comme les classes d'une relation d'équivalence sur X.

L'ensemble $\mathcal{P}(X)$ des parties de X et celui $\mathbb{P}(X)$ de ses partitions constituent deux structures de treillis.

4.2.1 Structures de treillis

Un **treillis** $\mathcal{T}$ est un couple $(X, \leq)$ où X est un ensemble et $\leq$ une *relation d'ordre partiel* sur X telle que pour tout couple $(x, y) \in X^2$, le plus petit majorant commun (**supremum**) de x et y

$$x \vee y \quad \triangleq \quad inf\{z \in X | z \geq x \quad \text{et} \quad z \geq y\}$$

et leur plus grand minorant commun (**infimum**)

$$x \wedge y \quad \triangleq \quad sup\{z \in X | z \leq x \quad \text{et} \quad z \leq y\}$$

existent et appartiennent à X.

Quand ces définitions s'étendent à toute famille F d'éléments, le treillis est déclaré **complet**. On note alors $\vee F$ et $\wedge F$, resp. le supremum et l'infimum d'une famille F. Un treillis complet possède donc deux bornes absolues, notées $\vee\mathcal{T}$ et $\wedge\mathcal{T}$ qui majorent et minorent tous ses éléments.

Si dans $(X, \leq)$, seul l'infimum (resp. supremum) de deux points existe systématiquement, on dit que $(X, \leq)$ est un **inf-demi-treillis** (resp. **sup-demi-treillis**).

Lorsqu'il n'y aura aucune ambiguité, nous omettrons de mentionner l'ordre considéré, et nous dirons simplement de X qu'il est le treillis.

Si X est fini, il suffit que deux élément quelconques possèdent toujours un minorant et un majorant commun pour que $(X, \leq)$ soit un treillis.

Un treillis sur un ensemble fini est automatiquement complet.

4.2.2 Parties et partitions

L'ensemble $\mathcal{P}(X)$ des parties d'un ensemble X est un treillis complet (même si X est infini) pour la relation d'inclusion large ($\subseteq$). $\vee$ est dans ce cas l'union ($\cup$), $\wedge$ l'intersection ($\cap$), $\vee\mathcal{P}(X) = X$ et $\wedge\mathcal{P}(X) = \emptyset$.

De son côté, l'ensemble $\mathbb{P}(X)$ des partitions de X est un treillis complet pour la relation de **finesse**. Cette relation, notée $\leq$, est définie à partir de l'ordre sur les parties par :

$$A \leq B \quad \Leftrightarrow \quad \forall a \in A \quad \exists b \in B \quad | \quad a \subseteq b.$$

Quand $A \leq B$, on dit que la partition A est plus fine que B, ou encore que A est une sur-partition de B. Inversement, B est dite moins fine que A, ou appelée une sous-partition de A. Quand les partitions correspondent à des segmentations des pixels d'une image, on parle de sur et de sous segmentations. Intuitivement, la relation de finesse $A \leq B$ exprime que A peut être obtenue à partir de B par redécoupage de certaines de ses parties[2]. Inversement, B peut être déduite de A par réunion de certaines de ses parties.

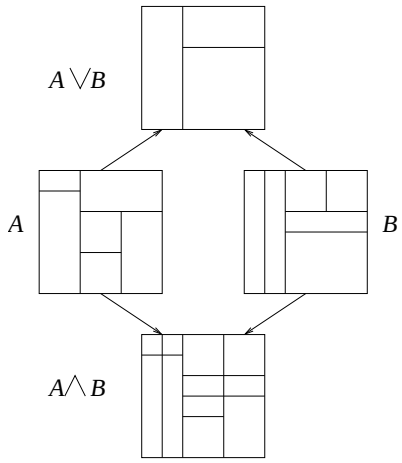


FIG. 4.1 – Infimum et supremum de deux partitions.

Supremum et infimum de deux partitions : Pour $(A, B) \in \mathbb{P}^2(X)$,

• Deux éléments de X appartiennent à une même classe de $A \vee B$ si et seulement si ils appartiennent à une même classe de A *ou* de B. Si X est muni d'une topologie, $A \vee B$ correspond à la partition qui ne comporte que les frontières communes à A et B (voir figure 4.1).

• De son côté, $A \wedge B$ est composé de toutes les intersections deux-à-deux des parties de A et de B. Deux éléments de X appartiennent à une même classe de $A \wedge B$ si et seulement si ils appartiennent à une même classe de A *et* de B. Si X est un espace topologique, $A \wedge B$ comporte toutes les frontières présentes dans A et dans B.

Une manière de concevoir le sup et l'inf de deux partitions est de considérer leur **graphe d'intersection**. Ce graphe bipartite $G = (A, B, \cap)$ possède un nœud pour chaque partie de A ou B et une arête pour chaque couple $(x, y) \in A \times B$ tel que $A \cap B \neq \emptyset$.

Dans le graphe d'intersection :

• Les arêtes correspondent aux parties de $A \wedge B$.

• Les composantes connexes correspondent aux parties de $A \vee B$.

Comme nous l'avons remarqué dans (Guigues 1999a) le graphe d'intersection de A et B permet de retrouver instantanément l'inégalité de Lerman (1981), à savoir que si A et B sont deux partitions alors

$$|A| + |B| \quad \leq \quad |A \vee B| + |A \wedge B|. \tag{4.1}$$

Il suffit en effet d'écrire que le nombre cyclomatique[3] $\gamma(G)$ d'un graphe G est toujours positif :

$$\gamma(G) \quad \triangleq \quad M - N + C \quad \geq 0 \tag{4.2}$$

où N est le nombre de nœuds de G, M son nombre d'arêtes et C son nombre de composantes connexes.

[2]Attention : dans le treillis des ensembles, si $x \subseteq y$, x est appelé un *sous*-ensemble de y. Par contre, dans le treillis des partitions, si $A \leq B$, A est notée «inférieure» à B mais est appelée une *sur*-partition de B, ce qui renvoit à l'idée que A correspond à sur-découpage de X par rapport à B.

[3]Ce nombre donne la dimension d'une base de cycles de G, cf. (Gondran et Minoux 1995, p. 22).

Quand G est le graphe d'intersection de deux partitions, on a alors $N = |A| + |B|$, $M = |A \wedge B|$ et $C = |A \vee B|$[4].

La borne inférieure du treillis des partitions est la partition composée de tous les singletons de X :

$$\wedge\mathbb{P}(X) = \big\{\{x\}\big\}_{x\in X}.$$

À l'autre extrémité du treillis $\mathbb{P}(X)$, sa borne supérieure est la partition constituée du seul ensemble X lui-même :

$$\vee\mathbb{P}(X) = \big\{X\big\}.$$

On voit que l'on peut obtenir toute partition intermédiaire soit par le haut, en divisant récursivement X, soit par le bas, en réunissant récursivement des classes. Lorsque l'on effectue de telles opérations, on parcourt alors des séquences ordonnées de partitions.

4.2.3 Séquences et chaînes - couverture

Par analogie avec les ensembles, nous disons de deux partitions A et B qui sont ordonnées pour la relation de finesse qu'elles sont *emboîtées*. Un ensemble $\{P_i\}_{i=1\ldots k}$ de parties ou de partitions strictement emboîtées $P_1 < P_2 < \ldots < P_k$ est alors appelé une **séquence**. En traitement d'image, une séquence de partitions est appelée une **pyramide** de segmentations[5]. Une *séquence complète* est une séquence $\{P_i\}_{i=1\ldots k}$ qui traverse tout le treillis, i.e. telle que $P_1 = \wedge\mathcal{T}$ et $P_k = \vee\mathcal{T}$.

Soit X un ensemble partiellement ordonné pour une relation d'ordre strict $<$. Pour $(x, y) \in X^2$ on dit alors que y *couvre* x, ce que nous notons $x \prec y$, quand

$$x < y \quad \text{et} \quad \nexists z \in X \quad | \quad x < z < y.$$

$\prec$ est appelée la **relation de couverture** associée à $<$. Notons que l'ordre partiel $<$ s'obtient comme la fermeture transitive de la relation de couverture.

Une *partie* y d'un ensemble en couvre une autre, x, ssi elle s'obtient en ajoutant un point à x.

[4]Dans (Guigues 1999a), nous proposons des méthodes s'appuyant sur le partitionnement du graphe d'intersection pour déterminer des appariements multivoques entre régions de deux partitions. Nous proposons en particulier de nous intéresser aux couples de partitions pour lesquelles il y a égalité dans l'expression 4.1 de Lerman, qui correspondent à des partitions pour lesquelles on sait définir sans ambiguité une notion de correspondance $n-m$ entre leurs parties. Le graphe d'intersection G de telles partitions - dont nous disons qu'elles sont *non-entrelacées* - est un *arbre*. Nous suggérons alors que le *diamètre* $D(G)$ de cet arbre (longueur de sa plus longue chaîne d'arêtes) caractérise la complexité structurelle de la correspondance entre A et B : si $D(G) = 1$, il y a couplage, les parties de A et B sont en bijection, $A = B$; si $D(G) = 2$, les parties de A et B sont en relations $1-n$ ou $n-1$, i.e. les partitions sont localement ordonnées ; etc. Nous proposons alors, pour obtenir des appariements de différentes complexités entre deux partitions A et B, de

i) construire leur graphe d'intersection G et d'attribuer ses arêtes par une mesure d'«importance» de l'intersection (surface, épaisseur...).

ii) calculer l'arbre recouvrant de poids minimal T de G.

iii) déterminer le «tronçonnement» (ensemble d'arêtes) de T de poids minimal qui permet de diviser l'arbre en composantes de diamètre borné par un entier k donné.

Pour $k = 1$ on obtient une approximation d'un couplage de poids maximal de G (Kuhn 1955) par une successions de deux étapes optimales. Pour $k > 1$, la procédure permet d'établir les «meilleures» correspondances multivoques avec contrainte sur la «complexité structurelle» des correspondances, généralisant les méthodes d'appariement 1-1 proposées par Vinet (1991) (voir aussi (Cocquerez et Philipp 1995) qui utilisent la méthode de Vinet pour comparer des résultats de segmentation). Nous donnons dans (Guigues 1998) un algorithme efficace pour résoudre le problème de tronçonnement optimal d'un arbre en composantes de diamètre borné. Voir aussi (Guigues 1997b) et (Guigues et Pierrot Deseilligny 1998) pour plus de détails.

[5]Remarque : Le terme «pyramide» possède deux sens bien différents en traitement d'image et en classification. En image, il désigne une séquence d'images de résolution décroissante, une séquence de partitions emboîtées ou - abusivement - une hiérarchie de régions. En classification, une pyramide est un ensemble de parties que l'on peut munir d'une relation d'ordre total particulière (Batbedat 1990), qui correspond à la notion d'hypergraphe d'intervalles de la théorie des graphes (Gondran et Minoux 1995). Cette notion englobe les hiérarchies de parties mais contient aussi des structures beaucoup plus complexes. Nous reparlerons des pyramides au sens de la classification en 7.1.3.

Une *partition* B d'un ensemble en couvre une autre, A, ssi elle s'obtient en réunissant deux classes de A.

Une **chaîne** de longueur k d'un treillis est un ensemble $\{P_i\}_{i=1\ldots k}$ d'éléments de $\mathcal{T}$ tels que $P_1 \prec P_2 \prec \ldots \prec P_k$. Une chaîne est donc une séquence qui ne peut pas être raffinée, i.e. telle qu'entre deux éléments successifs de la séquence, aucun élément intermédiaire ne peut être introduit. Une *chaîne complète* est une chaîne qui traverse tout le treillis.

La longueur d'une chaîne complète de partitions est systématiquement égale à $|X|$, celle d'une chaîne complète de parties est systématiquement égale à $|X| + 1$.

4.2.4 Représentation des treillis

On représente habituellement un treillis par son graphe de couverture $(X, \prec)$ (voir figures 4.2 et 4.3). Ce graphe est sans cycle. On peut le dessiner dans le plan (avec des intersections entre arêtes, il n'est généralement pas planaire), de telle manière que si $x < y$, x soit placé plus bas que y. Une chaîne est alors un chemin (au sens des graphes) à ordonnées strictement croissantes dans le graphe. Une chaîne complète est un chemin reliant $\wedge\mathcal{T}$ et $\vee\mathcal{T}$.

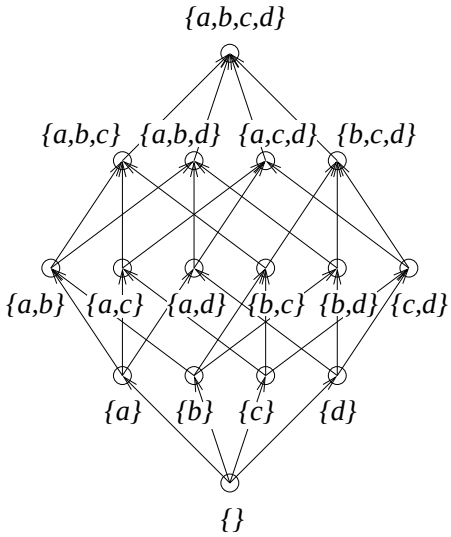


FIG. 4.2 – Treillis des parties d'un ensemble à quatre éléments.

Remarque : l'ensemble des images à un seul canal sur un domaine donné (fonctions d'un ensemble $\mathcal{D}$ quelconque dans un ensemble $\mathcal{V}$ totalement ordonné, e.g. $\mathbb{R}$ ou $\mathbb{Z}$) constitue également un treillis pour la relation d'ordre partiel

$$I \leq J \qquad \Leftrightarrow \qquad \forall x \in \mathcal{D} \quad I(x) \leq J(x).$$

Ce treillis sert de base pour la transposition aux images des opérations morphologiques sur les ensembles (Serra 1982). Il est complet quand $\mathcal{V}$ est compact, e.g. est un intervalle fermé (borné) de $\mathbb{R}$ ou $\mathbb{Z}$. Sa borne inférieure est l'image «noire», sa borne supérieure l'image «blanche».

4.3 Représentation et manipulation de partitions d'images

Intéressons-nous maintenant aux structures informatiques permettant de représenter et manipuler des partitions d'images.

Remarques préliminaires :

i) Nous n'entrerons pas dans certains détails de topologie ou de géométrie qui seraient nécessaires à une

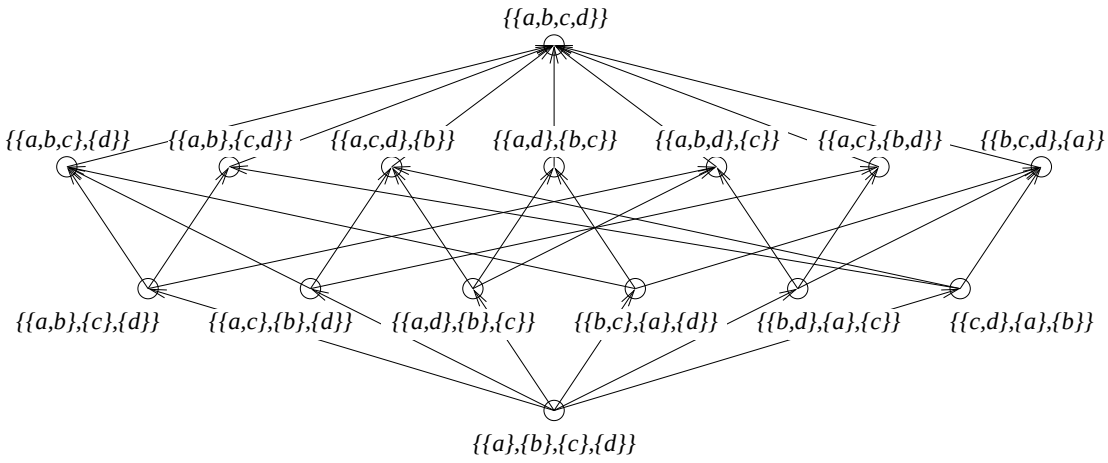


FIG. 4.3 – Treillis des partitions d'un ensemble à quatre éléments..

présentation mathématique rigoureuse.
ii) La présentation est axée sur les partitions d'images 2D, mais toutes les notions s'étendent, plus ou moins facilement, à des espaces de dimensions arbitraires.
iii) La condition de partitionnement n'est jamais nécessaire, seul importe le *non recouvrement* de deux régions : on peut toujours compléter une pré-partition en une partition, par adjonction du complémentaire des régions.

4.3.1 Généralités

Si $\mathcal{D}$ est un domaine d'un ensemble doté d'une topologie plane convenable, il y a essentiellement trois manières de décrire une partie R de $\mathcal{D}$:

i) Par la fonction indicatrice 1_R de R.

ii) Par l'ensemble explicite des points de R, i.e. l'image réciproque de 1 par 1_R.

iii) Par la description de la frontière δR de R par un ensemble de courbes de Jordan, correctement orientées de manière à discerner l'intérieur de l'extérieur de la région, R de $\mathcal{D}\backslash R$, via une convention préalable (p.ex. intérieur à gauche en suivant les courbes).

Ces trois modes de description se transmettent naturellement aux partitions.
Si $P = \{R_i\}_{i\in\langle 0,N\langle}$ est une partition (finie) de $\mathcal{D}$, on peut la représenter :

i) Par une fonction d'étiquettage $L : \mathcal{D} \to \langle 0, N\langle$ définie par

$$L \triangleq \sum_{i\in\langle 0,N\langle} i \times 1_{R_i}.$$

L associe à chaque point de $\mathcal{D}$ le numéro de la région à laquelle il appartient. Les R_i sont alors les classes d'équivalence de la relation «avoir la même étiquette». Pour une image numérique I, L est une image de même support que I à valeurs entières, appelée *image d'étiquettes* (*image de labels* en franglais).

ii) Par les ensembles R_i sous forme explicite.

iii) Par un **graphe des frontières de régions** de P.

Intéressons-nous à ce graphe et commençons par le décrire dans le cas continu.

4.3.2 «Partitions» d'images idéales

Sur des domaines continus, les partitions en régions n'en sont pas réellement...

Si le support $\mathcal{D}$ de l'image est une partie ouverte d'un espace continu, e.g. $\mathbb{R}^n$, on ne peut pas en général paver $\mathcal{D}$ par des ensembles de nature topologique homogène, c'est-à-dire ni constituer sur $\mathcal{D}$ une partition composée uniquement de fermés, ni une partition composé uniquement d'ouverts. Comme il n'y a aucune raison, au niveau d'une frontière entre deux régions, de privilégier l'une ou l'autre des régions voisines, on exclut finalement la frontière des deux ensembles.

Intéressons-nous au cas d'un domaine du plan. Si X est un ensemble, notons $\overline{X}$ sa fermeture topologique, $\overset{o}{X}$ son intérieur et $\delta X \triangleq \overline{X} \backslash \overset{o}{X}$ sa frontière.

Classiquement (Mumford et Shah 1989), une segmentation S de $\mathcal{D}$ est alors définie comme un couple $S = (P, \Gamma)$, où $P = \{R_1 \dots R_n\}$ est un ensemble de régions ouvertes disjointes telles que $\bigcup_{R \in P} \overline{R} = \overline{\mathcal{D}}$ et $\Gamma = \mathcal{D} \backslash \bigcup_{R \in P} R = \bigcup_{R \in P} \delta R$ est un continuum de dimension 1, un ensemble de courbes qui représente l'ensemble des frontières des régions.

L'*interface* entre deux régions R_i et R_j est alors donnée par $B_{ij} = \delta R_i \cap \delta R_j$[6]. Deux régions i et j sont déclarées voisines quand $B_{ij} \neq \emptyset$. Décomposons chaque interface B_{ij} en portions connexes. Modulo quelques hypothèses de régularité sur Γ, ces courbes sont des fermés 1D (sauf quand elles se connectent à la frontière du domaine $\mathcal{D}$). Elles s'intersectent aux lieux d'interface ponctuelle entre plus de deux régions, appelées *jonctions*. Comme à nouveau il n'y a pas de raison d'affecter ces jonctions à une frontière plutôt qu'à une autre, itérons le procédé qui a été appliqué aux régions. Ouvrons toutes les frontières au niveau des jonctions et collectons les points dans un ensemble J. Plaçons dans un ensemble B les portions connexes de frontières ouvertes restantes. Si une portion de frontière est une courbe sans bord (un cercle topologique), ouvrons-la en un point arbitraire et mettons ce point dans J. Si une portion de frontière $B_i \in B$ se connecte au bord de $\mathcal{D}$, i.e. si $\delta B_i \cap \delta \mathcal{D} \neq \emptyset$[7], plaçons également cette jonction dans J. La jonction entre deux portions de frontière B_i et B_j est alors donnée par $\delta B_i \cap \delta B_j$[8].

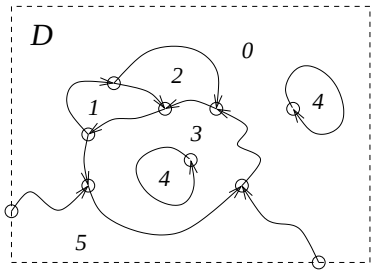


FIG. 4.4 – Graphe de frontières des régions d'une partition du plan. Flèches : frontières / arêtes du graphe. Cercles : jonctions / noeuds du graphe. Nombres : étiquettes des régions / faces du graphe.

Nous appelons alors le graphe $C = (J, B)$ le graphe de *connexions* de Γ (cf. figure 4.4). Par construction, C est *planaire*. Mais il y a généralement plusieurs façons de représenter un graphe planaire dans le plan. La topologie des connexions ne suffit pas à déterminer la topologie de P (ne connaissant par exemple que le graphe de connexions de la partition de la figure 4.4, on pourrait très bien imaginer que les deux composantes connexes de la région 4 soient toutes deux intérieures à la région 3, etc.).

La théorie des cartes combinatoires de Tutte (1984) propose une manière de décrire de façon non équivoque la topologie de P, i.e. de spécifier P à un homéomorphisme près (la théorie est développée de manière générale sur n'importe quelle variété topologique de dimension 2).

[6] On suppose qu'il n'existe aucune frontière interne aux régions, i.e. qui n'existe que dans B_{ii}.

[7] L'opérateur «frontière» agit ici en dimension 1, c'est-à-dire en dimension intrinsèque aux frontières.

[8] On imagine aisément l'extension en dimension quelconque de ce procédé de décomposition récursive des interfaces entre ouverts.

Pour notre part, nous aurons également besoin de la géométrie de Γ. Positionner les jonctions et décrire la géométrie des frontières suffit alors à décrire P si elle est uniquement constituée de parties *connexes*. Les régions R de P correspondent alors aux *faces* du graphe. Si $R \in P$, chacune des courbes de Jordan qui constitue δR correspond à un cycle élémentaire du graphe de connexions auquel on aurait adjoint la frontière de $\mathcal{D}$ (nous excluons toutefois $\delta\mathcal{D}$ de la description car il s'agit d'une frontière *arbitraire* qui ne correspond pas à une limite d'un objet de l'image). On peut alors reconstituer l'ensemble des régions de P mais avec une numérotation arbitraire. Pour maintenir une numérotation informative sur les faces - e.g. que l'on puisse mettre en relation avec d'autres données concernant les régions auxquelles elles correspondent - on peut alors orienter les arêtes de C et indiquer pour chacune d'elles les numéros des deux faces qu'elle voit à sa gauche et à sa droite (il suffit que chaque face soit désignée par une arête, mais il est utile de conserver l'ensemble de la labellisation). Une table d'équivalence de faces permet alors de gerer les partitions en régions non connexes (une partitions en régions non connexes n'est rien d'autre qu'une partition d'une partition en régions connexes).

Nous définissons donc le **graphe des frontières de regions** de P comme le graphe planaire *représenté* $G_B = (J, B, P)$, où P représente la donnée des *faces* de G_B, que nous supposons ici en bijection avec les régions de P. Nous utiliserons l'accronyme **RBG** (Regions Boundaries Graph) pour désigner ce graphe.

4.3.3 Partitions d'images numériques

Frontières pixellaires et inter-pixellaires

Pour les images numériques 2D s'appuyant sur un maillage carré, i.e. dont le domaine $\mathcal{D}$ est une partie bornée de $\mathbb{Z}^2$, on peut choisir d'utiliser des frontières *pixellaires* ou des frontières *inter-pixellaires* (voir la figure 4.5).

(a) (b)

FIG. 4.5 – Frontière discrète pixellaire et inter-pixellaire. (a) La frontière de la région gris clair passe par les pixels gris foncé. (b) Elle passe entre les pixels.

Le modèle de frontière pixellaire représente l'analogue discret du modèle continu que nous venons de voir.

D'un point de vue géométrique, ce modèle est certainement plus juste que le modèle inter-pixellaire : si l'image est supposée être le résultat d'une quantification d'une réalité continue alors les limites réelles des objets passent effectivement quelque part au sein des pixels. Les pixels frontaliers portent des mesures de mélange, desquelles, et au vu des mesures des pixels voisins, il est possible de déduire une estimation sub-pixellaire de la position de la limite réelle. Cette recherche de précision, qui peut revêtir de l'importance par ailleurs, est cependant secondaire pour nous.

En outre, le modèle de frontière pixellaire sur une maille carrée pose de nombreux problèmes de topologie discrète (difficulté avec le théorème de Jordan, nécessité d'employer deux notions de voisinage différentes pour un ensemble et son complémentaire, etc. cf. eg. (Maître 2003, ch. 3)).

Finalement, du point de vue du problème de partitionnement d'image, le modèle inter-pixellaire est plus naturel et plus facile à gérer : les régions forment effectivement une partition du domaine de l'image.

Nous nous contenterons donc de partitions discrètes au sens propre, aux limites inter-pixellaires.

Pour définir le graphe de frontières d'une partition, il faut également spécifier le type de topologie utilisée : 4 ou 8-voisinage. Le plus simple est d'opter pour une topologie de 4-voisinage. Le maillage des frontières est dans ce cas le maillage dual de celui des pixels (au sens de la dualité RBG/RAG, cf. ci-dessous). En outre, le suivi de frontière peut s'effectuer par inspection du seul voisinage immédiat d'un élément orienté de frontière inter-pixellaire.

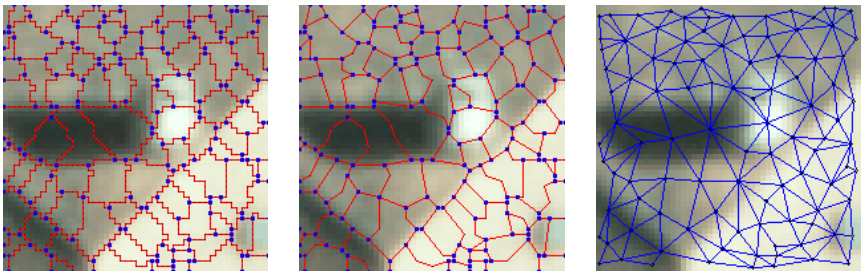
Dans ce qui suit nous supposerons par ailleurs que les régions de la partition sont connexes. Pour gérer les partitions en ensembles non connexes on utilisera une table d'équivalence sur les faces du graphe.

Graphe des frontières de régions (RBG)

Nous définissons donc le **graphe des frontières de régions (RBG)** associé à une partition $P = \{R_1, \ldots, R_n\}$ en régions connexes d'un domaine $\mathcal{D}$ de $\mathbb{Z}^2$ comme le graphe planaire représenté $G_B = (J, B, P)$ dont

- Chaque arête $b \in B$ modélise une portion connexe de frontière interpixellaire entre deux régions 4-adjacentes de P.
- Chaque noeud $j \in J$ modélise un point de jonction entre plus de deux frontières de régions, un point de jonction d'une frontière avec le bord du domaine ou une origine arbitraire d'une frontière fermée.
- Chaque face de $p \in P$ représente une région de la partition P.

La figure 4.6(a) représente un tel graphe.



(a) (b) (c)

FIG. 4.6 – Graphes associés à des partitions d'images numériques 2D. (a) Un graphe de frontières de régions (RBG) aux frontières inter-pixellaires. (b) Le même graphe dont les frontières ont été modélisées par des polylignes. (c) Le graphe d'adjacence géométrique de régions (RAG) dual du RBG vu en (a).

Des structures informatiques compactes et efficaces s'appuyant sur la théorie de Tutte et appelées *cartes combinatoires* ou *cartes topologiques* ont été proposées pour représenter et manipuler les graphes de frontières inter-pixellaires d'une image 2D (Braquelaire et Brun 1998). Les modèles sont théoriquement extensibles à des espaces de dimensions arbitraires, cependant les seules implémentations effectives que nous connaissons se bornent à l'espace 3D. Brun et Mokhtari (2001) présentent un panorama de ces modèles.

Graphes d'adjacence de régions (RAGs)

Le **Graphe d'Adjacence de Regions** (Region Adjacency Graph, **RAG**) est étroitement lié au RBG. Il y a en réalité deux sortes de RAGs que nous appelons *RAGs géométriques* et *RAGs topologiques*.

Le premier type de RAG est simplement défini par la relation de *dualité* qu'il entretient avec le RBG (Pavlidis 1977; Kropatsch 1995) : Si $G_B = (J, B, P)$ est le RBG associé à P, alors le **RAG géométrique** associé à P est le graphe $G_A = (P, B^*, J)$ dans lequel :

• Par rapport au RBG, le rôle des nœuds et des faces est échangé : les faces du RBG (régions) constituent les nœuds du RAG et les nœuds du RBG (jonctions) constituent les faces du RAG.

• Les arêtes du RAG sont en bijection avec celles du RBG : à chaque arête $b = (j_1, j_2)$ du RBG, ayant r_1 pour face droite et r_2 pour face gauche dans G_B, est associée l'arête $b^* = (r_1, r_2)$ de G_A qui a j_1 pour face gauche et j_2 pour face droite (cf. fig. 4.7).

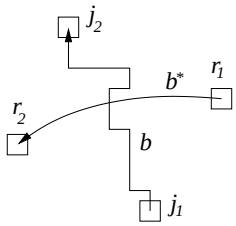
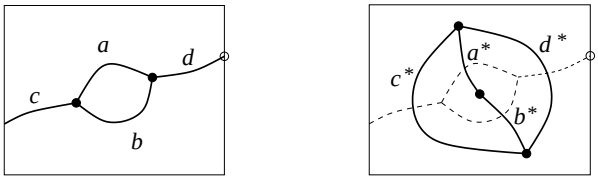


FIG. 4.7 – Dualité entre RBG et RAG géométrique.

La figure 4.6(c) représente le RAG géométrique associé au RBG vu en 4.6(a). Comme le RBG, le RAG géométrique est un *multi-graphe*, ce qu'illustre la figure 4.8.



(a) RBG (b) RAG

FIG. 4.8 – Une région enclavée. (a) Dans le RBG, les arêtes a et b sont des multi-arêtes. (b) Dans le RAG géométrique, ce sont c^* et d^* - les duales de c et d - qui sont multiples.

Le **RAG topologique** G_A^t peut alors être défini comme le *simplifié* du RAG géométrique G_A, c'est-à-dire comme le graphe simple obtenu à partir de G_A en remplaçant toutes ses arêtes multiples par des arêtes simples (transformation de la collection des arêtes en un ensemble). Alors que les arêtes de G_A détaillent la décomposition en éléments connexes des interfaces entre régions, les arêtes de G_A^t signalent simplement l'adjacence entre régions, i.e. indiquent si deux régions possèdent ou non une frontière commune. Évidemment, le RAG topologique ne permet plus de reconstruire la géométrie des frontières.

RBAG

Il est utile de synthétiser les données d'un RBG et de son dual géométrique en une structure unique - que nous appelons un **RBAG**[9] - sur laquelle on pourra adopter deux *points de vue*, un point de vue primal, celui du RBG, et un point de vue dual, celui du RAG géométrique. La structure d'RBAG instancie physiquement les trois types d'éléments, régions connexes P, jonctions J et frontières B. Chaque élément d'un RBAG peut alors être attribué. Chaque région peut ainsi porter l'échantillon complet des valeurs des pixels de la zone qu'elle représente, ou un modèle de ces valeurs (e.g. certaines statistiques, un modèle fonctionnel). De même, chaque frontière peut porter la description complète de la géométrie de la frontière qu'elle représente (la séquence des codes de Freeman de chaque pas inter-pixellaire), un modèle de cette géométrie (e.g. une approximation polygonale comme en figure 4.6(b)) ou simplement certains attributs de la frontière (e.g. sa longueur).

[9] Lire «Air Bag» :-)

Un RBAG *explicite* l'ensemble des données utiles pour manipuler des segmentations : la *géométrie* d'une image segmentée à travers le RBG, sa *topologie* d'adjacence à travers le RAG et les *mesures* image à travers les attributs de régions ou de frontières (les frontières peuvent par exemple porter des mesures de gradient).

Si $S \subset P$ est une région ou un groupe de régions, les algorithmes de segmentation qui seront proposés ci-dessous utiliserons trois opérations de base sur un RBAG :

• deux opérations de *requête* :

1) retrouver la géométrie complète de la frontière de S.

2) retrouver l'ensemble des régions voisines de S.

qui s'effectuent efficacement (les algorithmes sont faciles à imaginer).

• une opération de *modification*, à savoir fusionner l'ensemble des régions de S en une seule. La figure 4.9 illustre la manière dont on met à jour la structure lors d'une telle fusion.

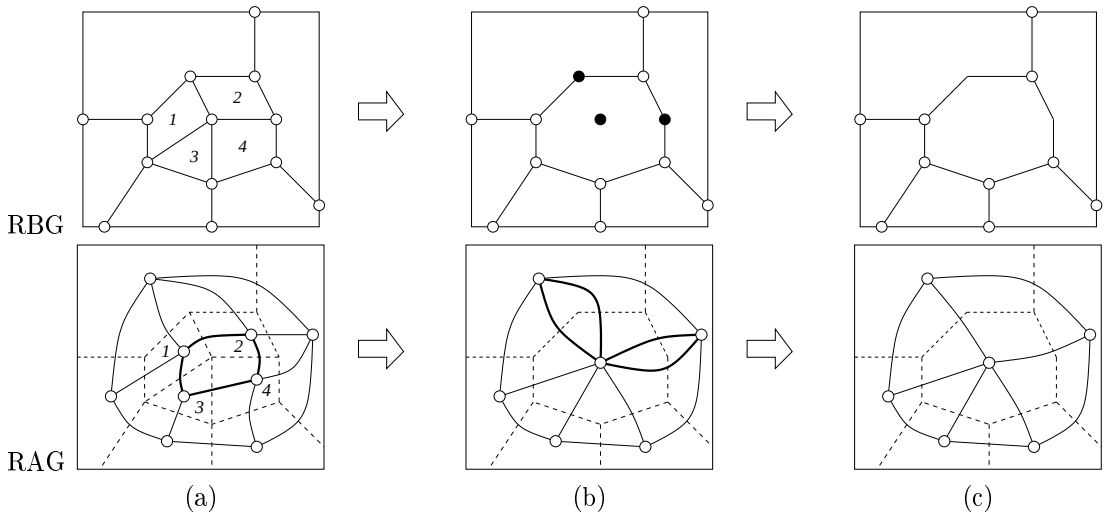


FIG. 4.9 – Fusion de régions dans un RBAG. La fusion des 4 régions 1, 2, 3 et 4 s'effectue en 2 étapes. *Du point de vue du RBG (1ère ligne)* : 4 frontières sont supprimés entre (a) et (b). 2 jonctions intermédiaires (nœuds de degré 2) et une jonction libre (degré nul) apparaissent alors. Elles sont supprimées entre (b) et (c). *Du point de vue du RAG (2ème ligne)* : entre (a) et (b), les 4 nœuds sont contractés en un seul et leurs relations d'adjacence (arêtes du sous-graphe induit par $\{1, 2, 3, 4\}$, marquées en gras en (a)) sont supprimées. Entre (b) et (c), les jonctions intermédiaires superflues apparaissent comme des multi-arêtes qui sont simplifiées. Les arêtes à supprimer entre (a) et (b) se détectent naturellement en utilisant le RAG. Les jonctions libres à supprimer entre (b) et (c) sont les nœuds de degré nul du RBG. Les jonctions intermédiaires à supprimer entre (b) et (c) correspondent aux arêtes devenues multiples dans le RAG après contraction de $\{1, 2, 3, 4\}$ et qui possèdent une jonction commune du point de vue du RBG.

4.4 Hiérarchies

4.4.1 Définitions générales

Nous disons de deux parties A et B d'un ensemble X qu'elles sont **hiérarchisées** si et seulement si elles sont soit emboîtées, soit disjointes, c'est-à-dire ssi $A \cap B \in \{A, B, \emptyset\}$ (on dit aussi que A et B n'ont pas d'intersection propre). Par extension, nous disons qu'un ensemble H de parties est hiérarchisé quand ses éléments le sont tous deux-à-deux.

Si X est un ensemble, une **hiérarchie** H sur X est alors un ensemble *hiérarchisé* de parties de X qui vérifie :

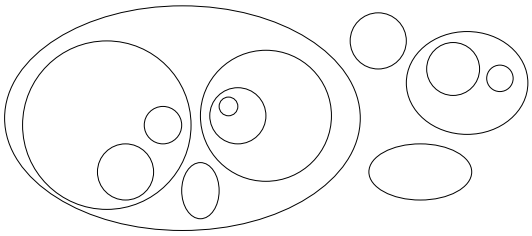


FIG. 4.10 – Parties hiérarchisées.

i) $\emptyset \notin H$
ii) $X \in H$.
iii) $\forall x \in X \quad : \quad \{x\} \in H$.

Nous notons $\mathbb{H}(X)$ l'ensemble des hiérarchies sur X et si X est muni d'une topologie, $\mathbb{H}_C(X)$ désigne l'ensemble des hiérarchies sur X formées exclusivement de parties connexes.

X est appelé le **sommet** de la hiérarchie. Nous parlerons également d'*Univers*. Nous le notons $\vee H$.

L'ensemble $\mathcal{B}(H) \triangleq \{\{x\}\}_{x \in X}$ est appelé la **base** de la hiérarchie. Nous dirons que ses éléments sont les *atomes* de la hiérarchie (ou de l'Univers). Notons que $\mathcal{B}(H)$ est la sur-partition absolue de X.

Les parties de H qui ne sont ni l'Univers entier, ni des atomes seront dites **intérieures** à la hiérarchie. Leur ensemble, que nous notons $\overset{\circ}{H}$, sera donc appelé l'intérieur de H[10]. La hiérarchie triviale d'intérieur vide sera aussi appelée la hiérarchie *creuse*.

$\mathbb{H}(X)$ est un inf-demi-treillis pour l'inclusion entre *hiérarchies*. Sa borne inférieure est la hiérarchie creuse et les hiérarchies maximales sont les hiérarchies binaires (voir ci-dessous).

Une hiérarchie H, si elle est finie, est un sup-demi-treillis pour l'inclusion entre *parties*, qui possède la particularité qu'*aucun* couple de parties ne possède de minorant commun. Par la suite, toutes les hiérarchies considérées seront supposées finies.

On représente alors habituellement H par le graphe $(H, \prec)$, où $\prec$ est la relation de couverture associée à l'inclusion entre parties. La figure 4.11 montre une représentation de ce graphe sous forme de *dendrogramme* (représentation non orientée avec des arêtes anguleuses). $(H, \prec)$ est une anti-arborescence : on obtient une arborescence de racine X en changeant l'orientation de toutes les arêtes (c'est-à-dire en considérant $(H, \succ)$). Sous forme non orientée, $(H, \prec$ ou $\succ)$ est un arbre.

De façon classique, nous assimilerons une hiérarchie H à l'arboresence $(H, \succ)$. Nous parlerons donc des nœuds et des arêtes d'une hiérarchie ; les atomes seront également appelés *feuilles* et le sommet la *racine*.

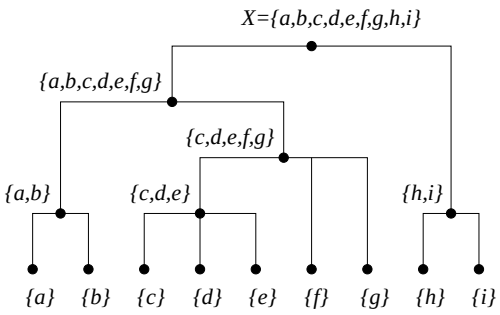


FIG. 4.11 – Représentation d'une hiérarchie sous forme de dendrogramme.

[10]Certains auteurs, e.g. Leclerc (1985), définissent l'intérieur de H par $H \backslash \mathcal{B}(H)$, incluant donc le sommet.

4.4.2 Père et fils, branches et infimum

Tout élément x d'une hiérarchie H finie, à l'exception du sommet, possède un unique élément couvrant que l'on appelle le **père** de x. Nous le notons $\mathsf{p}(x)$.

Inversement, tout élément x d'une hiérarchie H couvre un certain nombre d'éléments de H, appelés les **fils** de x. Nous les notons $\mathcal{F}(x) \triangleq \{y \in H, y \prec x\}$.

$|\mathcal{F}(x)|$ est appelé l'**arité** de x. Une hiérarchie dont tous les éléments hormis ceux de base sont d'arité n est une hiérarchie n-aire.

$\forall x \in H$, la chaîne $x \prec \mathsf{p}(x) \prec \mathsf{p}(\mathsf{p}(x)) \cdots \prec X$ est appelée la **branche** de x. Graphiquement, il s'agit du chemin joignant x à $\vee H$. Une branche issue d'une feuille de H - qui traverse toute la hiérarchie - est appelée une branche de H.

Dans le graphe $(H, \prec)$, le **supremum** $x \vee y$ de deux éléments x et y de H se trouve à la confluence de leurs branches (cf. fig. 4.12). De même, pour toute famille $Y \subset H$, le supremum $\vee Y$ est à la confluence de toutes les branches des $y \in Y$, ce qui justifie la notation $X = \vee H$ pour le sommet de H (on a aussi $\vee H = \vee \mathcal{B}(H)$).

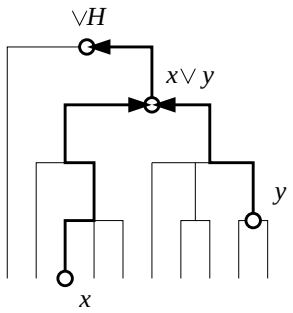


FIG. 4.12 – Branches et borne supérieure de deux nœuds d'une hiérarchie.

Le **niveau** d'un élément dans une hiérarchie est défini comme la longueur du plus long chemin qui le connecte à la base de la hiérarchie.

4.4.3 Hiérarchies et partitions - coupes

Construction d'une hiérarchie à partir de partitions

Une hiérarchie sur X peut également être vue comme l'ensemble des classes d'une séquence complète de partitions de $\mathbb{P}(X)$ (cf. fig. 4.13, c'est ainsi que Simon (1985) les définit).

En clair, toute hiérarchie H sur X peut se définir comme l'union

$$H \triangleq \bigcup_{i=1}^{k} P_i$$

des membres d'une famille $S = \{P_1, P_2 \ldots P_k\}$ de partitions monotones $P_1 < P_2 \cdots < P_k$ qui contient la sur et sous-partition absolue de X, i.e. telle que $P_1 = \wedge\mathbb{P}(X)$ et $P_k = \vee\mathbb{P}(X)$.

En particulier, une hiérarchie *binaire* sur X est l'union des éléments d'une *chaîne* complète $P_1 \prec P_2 \cdots \prec P_k$ de $\mathbb{P}(X)$.

Toutefois, ***une hiérarchie n'est pas équivalente à une séquence de partitions***. Plusieurs séquences produisent en effet la même hiérarchie. Autrement dit, une séquence de partitions contient plus d'information qu'une hiérarchie, c'est une structure plus contrainte. Formellement, il suffit de constater que la description d'une séquence utilise un niveau de parenthésage de plus que celle d'une hiérarchie :

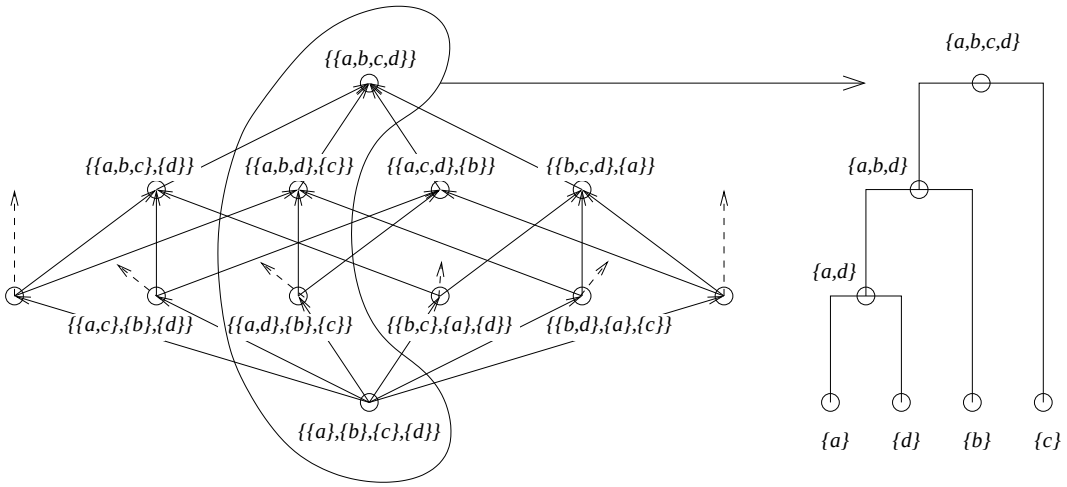


FIG. 4.13 – Construction d'une hiérarchie à partir d'une séquence de partitions.

c'est un ensemble d'ensembles d'ensembles. On passe à une hiérarchie en supprimant un niveau de parenthésage, par une union.

En traitement d'image, on fait cependant souvent l'amalgame entre pyramides de segmentations et hiérarchies de régions. La distinction claire entre ces deux types de données sera un élément central de notre approche multi-échelles de la segmentation.

Inversement, si l'on dispose d'une hiérarchie H sur X, on obtient une partition de X par *coupe* dans H.

Construction de partitions à partir d'une hiérarchie - coupes

Si H est une hiérarchie sur X, les deux définitions suivantes sont équivalentes :

i) Une **coupe** de H est une partition de X uniquement constituée de parties prises dans H (point de vue ensembliste).

ii) Une **coupe** de H est un ensemble de nœuds de H que toute branche de H intersecte une et une seule fois (point de vue graphique).

L'ensemble des coupes de H est donc l'ensemble des partitions de X que l'on peut construire avec des parties de H. Nous notons cet ensemble $\mathbb{C}ut(H)$.

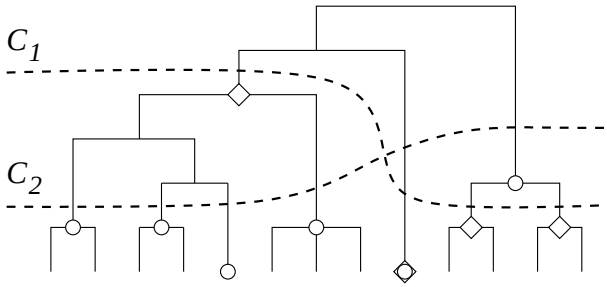


FIG. 4.14 – Coupes dans une hiérarchie. Les deux ensembles de nœuds marqués par des cercles et des losanges sont deux coupes. On les représente aussi par des courbes, C_1 et C_2, qui «coupent» les arêtes montantes des nœuds qui appartiennent à la coupe.

La définition graphique renvoie à la notion de coupe entre deux ensembles de nœuds d'un graphe. Si $G = (X, U)$ est un graphe et $A \subset X$ et $B \subset X$ alors une coupe entre A et B est un ensemble E d'arêtes qui, si elles sont supprimées du graphe, placent A et B dans deux composantes connexes distinctes. Autrement dit, E intersecte tous les chemins entre A et B. Une coupe est minimale si toute élimination d'un de ses élément lui retire la propriété d'être une coupe.

Dans le graphe $(H, \prec)$ associé à une hiérarchie H, tout nœud x - à l'exception du sommet - est en bijection avec l'arête $(x, \mathsf{p}(x))$ qui le connecte à son père. Ainsi, si l'on ajoute à H un nœud virtuel v et l'arête (X, v) qui fait de v le père de X, alors l'ensemble des coupes minimales - au sens graphique - entre la base de H et v est en bijection avec l'ensemble des coupes - au sens ensembliste - de H.

C'est en vertu de cette bijection que l'on représente une coupe dans une hiérarchie par un trait intersectant les arêtes de l'arbre (cf. fig. 4.14). Les éléments eux-mêmes de la coupe sont cependant des *nœuds* : les nœuds initiaux des arêtes coupées ou encore les plus hauts ensembles «sous» la coupe graphique, i.e. les ensembles maximaux parmi ceux qui restent connectés à la base après suppression des arêtes coupées.

Une hiérarchie admet généralement un très grand nombre de coupes (la taille des espaces $\mathbb{C}ut(H)$ sera évaluée ci-dessous pour certains types de hiérarchies). En outre, deux d'entres elles sont en général *non ordonnées* par finesse (rappelons que les coupes sont des partitions). C'est le cas par exemple des coupes C_1 et C_2 de la figure 4.14 (ce qui se constate visuellement parce qu'elles se «coupent» !).

Si deux coupes C_1 et C_2 ne sont généralement pas ordonnées, elles entretiennent tout de même une relation particulière, que nous avons appelée dans (Guigues 1999a) une relation d'ordre *local* : en certains lieux du domaine X, C_1 est une sur-partition de C_2 et en d'autres lieux, C_1 est une sous-partition de C_2. Précisons la nature de cette relation.

Si $P \in \mathbb{P}(X)$ et $Y \subset X$, notons $P \cap Y = \{R \cap Y\}_{R \in P}$ la restriction de la partition P au domaine Y.

Soit alors $(P, Q) \in \mathbb{P}^2(X)$. Nous disons que P et Q sont **localement ordonnées** si et seulement si :

$$\forall Y \in P \vee Q \qquad P \cap Y \leq Q \cap Y \quad \text{ou} \quad P \cap Y \geq Q \cap Y.$$

En français, les restrictions de P et Q aux parties du *supremum* de P et Q sont comparables par finesse. Sur certaines régions de $P \vee Q$, P est localement plus fine que Q, sur d'autres, c'est Q qui est localement plus fine. Évidemment, deux partitions ordonnées sont localement ordonnées. Notons que si l'ordre local est plus souple que l'ordre «global» induit par la relation de finesse, il est encore loin d'être un ordre total (par exemple, les partitions de la figure 4.1 ne sont pas localement ordonnées). Deux partitions sont localement ordonnées si et seulement si leur graphe d'intersection est de diamètre inférieur ou égal à 2 (Guigues 1999a).

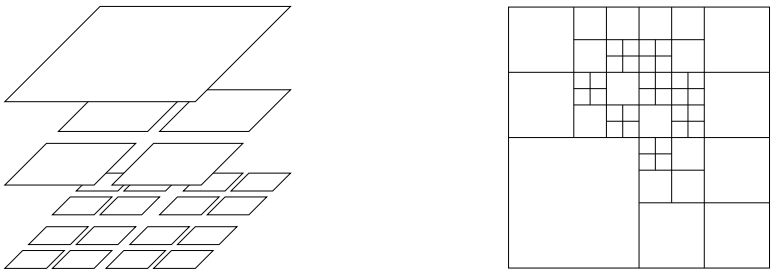
On vérifie alors aisément que le supremum $C_1 \vee C_2$ - au sens des partitions - de deux coupes C_1 et C_2 d'une hiérarchie est l'union des supremum - au sens des parties - des paires de parties comparables de C_1 et C_2, i.e. $C_1 \vee C_2 = \{x \cup y \mid (x, y) \in C_1 \times C_2, x \cap y \neq \emptyset\}$ ($C_1 \vee C_2$ est l'enveloppe supérieure des deux «coupes» graphiques). C_1 et C_2 sont donc localement ordonnées.

4.4.4 Hiérarchies et pyramides régulières - Quad trees

Les hiérarchies et les pyramides régulières ont été très largement utilisées en analyse d'image depuis qu'elles ont été introduite par Tanimoto et Pavlidis (1975).

L'espace X doit être un produit cartésien d'intervalles ouverts à droite, e.g. $X = [0, 1[^n$ ou $X = \langle 0, 256\langle^n$. La hiérarchie régulière $H_R(X)$ sur X est alors définie par division récursive de X en 2^n pavés similaires induits par dichotomie des intervalles sur lesquels est construit l'espace produit. La première étape de division de $[0, 1[^2$ donne ainsi les quatre régions $[0, 1/2[\times[0, 1/2[$, $[0, 1/2[\times[1/2, 1[$, $[1/2, 1[\times[0, 1/2[$, et $[1/2, 1[\times[1/2, 1[$. À la deuxième étape, chacune de ces régions est redivisée selon le même principe en 4 nouvelles régions, etc. Évidemment, pour les domaines discrets des images numériques, les intervalles doivent avoir des longueurs qui sont des puissances de 2 pour que la division récursive atteigne les pixels (on peut aussi le voir en sens inverse, comme une agglomération récursive).

Quand $n = 2$, l'arbre correspondant à $H_R(X)$ est un quad-tree (arbre quaternaire) ; si $n = 3$, c'est un oct-tree, etc.



(a) (b)

FIG. 4.15 – Une hiérarchie et une partition régulières.

Nous disons alors qu'une coupe d'une hiérarchie régulière est une *partition régulière*. Hiérarchies et partitions régulières sont à la base de l'algorithme de segmentation par Split and Merge (Division et Fusion) d'Horowitz et Pavlidis (1976).

Notons également que le support d'une *pyramide d'images* peut être conçu comme la chaîne des coupes *à niveau constant* d'une hiérarchie régulière.

4.5 Espaces ultramétriques - hiérarchies indicées

4.5.1 Distances et espaces ultramétriques

La notion de distance ultramétrique est plus forte que celle de distance métrique. Une ultramétrique δ sur un ensemble X vérifie toutes les propriétés d'une distance classique et une inégalité plus forte que l'inégalité triangulaire, qui s'exprime :

$$\forall (x, y, z) \in X^3 \qquad \delta(x, y) \leqslant \max\{\delta(x, z), \delta(z, y)\}.$$

Dans un espace ultramétrique $E = (X, \delta)$, cf. e.g. (Simon 1985) :
- Tous les triangles de E sont isocèles ou équilatéraux.
- Tous les points d'une sphère de E sont équidistants[11].
- Deux sphères qui s'intersectent sont nécessairement égales.
- Tout point intérieur à une boule est équidistant de tous les points de la surface de la boule, c'est-à-dire... en est le centre !
- Deux boules sont soit disjointes soit emboîtées. L'ensemble des boules de l'espace est donc hiérarchisé.

4.5.2 Boules d'un espace ultramétrique et hiérarchies indicées

Les singletons de X étant des boules de rayon nul et l'espace entier une boule de rayon $\max_{(x,y)\in X^2} \delta(x, y)$, les boules d'un espace ultramétrique (X, δ) constituent donc une hiérarchie sur X que nous appelons la **hiérarchie induite** par (X, δ).

Dans la lignée de la morphologie mathématique, une fonction $g : \mathbb{R}^+ \to \mathbb{R}^+$ strictement croissante sera appelée un **changement de contraste** (Matheron 1975).

[11] La sphère de centre x et de rayon r est l'ensemble des points à distance r de x. La boule (fermée) correspondante comporte tous les points à distance inférieure ou égale à r.

Soit δ une distance ultramétrique sur X et g un changement de contraste tel que $g(0) = 0$. $go\delta$ est alors également une ultramétrique et les hiérarchies induites par (X, δ) et $(X, go\delta)$ sont identiques.

Si δ est une ultramétrique sur X, notons alors $D_\delta : \mathcal{P}(X) \to \mathbb{R}^+$ la fonction qui à tout ensemble $Y \subset X$ associe son **diamètre** pour δ :

$$D_\delta(Y) \quad \triangleq \quad \max_{(x,y)\in Y^2} \delta(x, y).$$

Si H est la hiérarchie induite par (X, δ), alors le diamètre D_δ est une fonction nulle sur les feuilles de H et croissante pour l'ordre hiérarchique, i.e. vérifie

$$\forall (x, y) \in H^2 \qquad x \subseteq y \quad \Rightarrow \quad (D_\delta(x) \leq D_\delta(y)). \tag{4.3}$$

Une telle fonction est appelée un **critère croissant** en morphologie mathématique. Plus généralement, on peut définir de manière identique une notion de croissance (ou de décroissance, monotonie, etc.) pour n'importe quelle fonction d'un ensemble partiellement ordonné dans un autre (nous nous intéresserons par exemple ci-dessous au sens de variation de certaines énergies - i.e. fonctions à valeur réelle - définies sur le treillis des partitions).

Une **hiérarchie indicée** est un couple $I = (H, f)$, où H est une hiérarchie et f est un critère croissant sur H et nul sur ses feuilles, appelé son *indice*.

Si (H, f) est une hiérarchie indicée sur X, alors la fonction $\delta : X^2 \to \mathbb{R}^+$ définie par

$$\delta(x, y) = f(\{x\} \vee \{y\}) \tag{4.4}$$

est une distance ultramétrique sur X. H est alors la hiérarchie induite par (X, δ) et f coïncide avec la fonction diamètre D_δ sur les boules de (X, δ).

Espaces ultramétriques et hiérarchies indicées sont donc équivalents.

Nous les confondrons donc et nous noterons par une même lettre une fonction «distance ultramétrique» et la fonction «diamètre de boule», i.e. l'indice hiérarchique qu'elle induit.

Si cette lettre est δ alors, appliquée à un couple de points de X, $\delta(x, y)$ représentera leur distance.

Appliquée à une boule de (X, δ), i.e. à une partie $B \in \mathcal{P}(X)$ qui appartient à la hiérarchie H induite par (X, δ), $\delta(B)$ représentera son diamètre, c'est-à-dire l'indice de B dans H.

Par extension, si $Y \in \mathcal{P}(X)$ est une partie quelconque de X, $\delta(Y)$ représentera le diamètre de la plus petite boule contenant Y. D'un point de vue hiérarchique $\delta(Y)$ sera donc l'indice du supremum des singletons de Y, i.e.

$$\delta(Y) = \delta(\vee\{\{y\}\}_{y\in Y}).$$

Cette définition est bien consistante avec la relation 4.4 entre une ultramétrique et l'indice hiérarchique qu'elle induit.

Dans les approches multi-échelles de la segmentation qui seront développées ci-dessous, cette fonction ultramétrique / diamètre / indice sera appelée paramètre d'échelle ou simplement **échelle**.

4.5.3 Représentation graphique d'une hiérarchie

Comme nous venons de le voir, il existe une infinité d'ultramétriques associées à une hiérarchie H.

Il existe néanmoins une ultramétrique remarquable, à savoir celle qui correspond au *niveau* des ensembles dans H, qui est de toute évidence un critère croissant. L'ultramétrique du niveau est l'ultramétrique entière, i.e. à valeurs dans $\mathbb{N}$, qui est minimale : on ne peut réduire la distance entre deux éléments sans simultanément «réduire» la hiérarchie qu'elle induit, c'est-à-dire confondre certaines boules de l'espace.

Un autre critère croissant trivial est le cardinal des ensembles de H.

Lorsque l'on dessine une hiérarchie sous forme d'arbre, on utilise - souvent inconsciemment - un critère croissant sur H pour affecter aux éléments de H des ordonnées dans la représentation. La croissance du critère assure qu'un père sera plaçé au dessus de ses fils. On utilise habituellement le critère du niveau car c'est celui qui donne la représentation la plus «tassée» verticalement, en vertu de la minimalité de l'ultramétrique associée. Toutes les hiérarchies que nous avons représentées jusqu'à maintenant l'étaient avec des ordonnées proportionnelles aux niveaux.

Toujours pour le dessin d'une hiérarchie, la position horizontale des ensembles n'est pas indifférente si l'on souhaite obtenir une représentation *planaire*. Il faut en effet que l'abcisse des éléments respecte l'ordre partiel hiérarchique, ce qui peut se faire par descente recursive de H, comme suit.

Notons x la fonction «abcisse» sur H que l'on cherche à calculer de manière à obtenir une représentation planaire. Positionnons le sommet en $x(\vee H) = 0$ et appelons sur $\vee H$ la fonction recursive R suivante : $R(n)$ parcourt dans un ordre donné les fils $(s_1 \dots s_p)$ du nœud n et leur affecte l'abcisse $x(s_i) = x(n) + \sum_{j<i} |s_i|$ avant de s'appeler récursivement sur les s_i. À la fin de la récursion, les nœuds possèdent une abcisse compatible avec un affichage planaire. Pour ce qui est de l'ordre de parcours des fils d'un nœud, on peut par exemple considérer l'ordre des cardinaux ou celui des arités. Finalement, si l'on souhaite «recentrer» la représentation (après la procédure récursive qui précède elle est «penchée» sur l'axe $x = 0$) on effectuera une remontée ordonnée de H qui affecte la moyenne des abcisses de ses fils à chaque nœud, ou encore le milieu de l'intervalle d'abcisses des nœuds qui constituent sa base.

4.5.4 Représentation et coupes naturelles d'une hiérarchie indicée

Si une hiérarchie H est munie d'un *indice* δ, une **représentation naturelle** de H consiste bien entendu à affecter à chaque nœud $x \in H$ une ordonnée $y(x)$ qui correspond à son indice $\delta(x)$. Cette représentation est directement associée à une famille de coupes remarquables de (H, δ), que nous appelons la famille des **coupes naturelles** de (H, δ).

La coupe naturelle $\mathcal{C}_d$ de hauteur $d \geq 0$ de (H, δ) est définie comme l'ensemble des parties de H d'indice inférieur à d et maximales pour cette propriété :

$$\mathcal{C}_d = \{x \in H \quad | \quad \delta(x) \leq d \quad \text{et} \quad \nexists y \in H \quad \text{tel que} \quad x \subset y \quad \text{et} \quad \delta(y) \leq d\}\,.$$

Le fait que $\mathcal{C}_d$ soit une coupe, i.e. une partition de X, découle directement de la croissance de δ. Du point de vue de l'espace ultramétrique (X, δ), $\mathcal{C}_d$ représente l'ensemble des boules maximales de diamètre inférieur à d. La croissance de δ assure en outre que la famille des coupes naturelles de (H, δ) constitue une séquence de partitions monotones, autrement dit

$$d' \geq d \quad \Rightarrow \quad \mathcal{C}_{d'} \geq \mathcal{C}_d.$$

Le rapport intuitif entre la représentation naturelle et les coupes naturelles de (H, δ) est très simple : La coupe naturelle $\mathcal{C}_d$ correspond à la coupe graphique de la représentation naturelle selon la *droite horizontale* d'ordonnée d (voir la figure 4.16).

Autrement dit, la représentation naturelle correspond à la représentation dans laquelle les coupes naturelles apparaissent comme des coupes horizontales. Cette représentation sera la clef de voûte de l'approche de segmentation multi-échelle par minimisation d'énergie de la troisième partie du mémoire.

Notons que si deux ultramétriques reliées par un changement de contraste induisent la même hiérarchie, inversement les ultramétriques (critères croissants) compatibles avec une hiérarchie sont loin d'être toutes reliées par des changements de contraste. Ce point est lié à la différence fondamentale entre séquences de partitions monotones et hiérarchies : les ultramétriques contraste-équivalentes compatibles avec H sont celles qui induisent les mêmes familles de coupes naturelles. Graphiquement, on peut arbitrairement déformer l'axe vertical d'une représentation naturelle (moduler δ en $go\delta$, g strictement croissante), cela n'a aucune incidence sur l'ensemble - pris dans sa globalité - de ses coupes par des horizontales (la coupe particulière obtenue à une hauteur donnée change bien sûr). Par contre, transformer δ par une

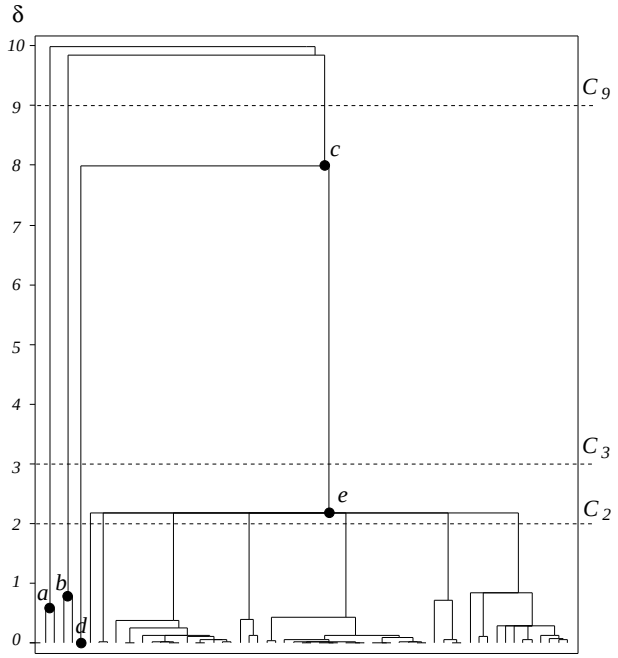


FIG. 4.16 – Représentation et coupes naturelles d'une hiérarchie indicée. $\mathcal{C}_9$ contient 3 ensembles qui correspondent aux trois nœuds a, b et c qui pendent des arêtes coupées par la droite d'ordonnée 9 (trait pointillé). $\mathcal{C}_3 = \{a, b, d, e\}$, etc.

fonction non croissante (ou de façon plus générale changer δ en δ' croissante pour H sans qu'il n'existe de relation fonctionnelle entre δ et δ') modifie la famille des coupes. Par exemple, sur la figure 4.16, faire passer b au dessus de c - e.g. en lui attribuant un indice de $8,5$ - supprimerait la coupe $\{a, b, d, e\}$ de la famille des coupes naturelles, la division en deux de b intervant avant celle de c.

Remarquons finalement qu'une pyramide régulière peut s'envisager comme la famille des coupes naturelles d'une hiérarchie régulière indicée par le niveau.

Hiérarchies de cocons

Chapitre 5

Hiérarchies de cocons

5.1 Introduction

Ce chapitre décrit une première approche de la segmentation qui s'appuie sur le formalisme de la théorie des graphes et aboutit à des ensembles, appelés *cocons*, qui s'organisent en hiérarchies.

Même s'il donne de moins bon résultats pratiques que le modèle énergétique que nous développerons dans la troisième partie du mémoire - nous verrons pourquoi - le modèle de cocon revêt plusieurs intérêts théoriques :

Il s'agit avant tout d'un «vrai» modèle multi-échelles au sens où l'aspect multi-échelles est une conséquence de la théorie et non un parti pris de modélisation.

Ce modèle conduit à une famille classique d'algorithmes de groupement hiérarchique (single, complete linkage et variantes). Nous retrouvons donc des méthodes déja connues. Cependant, encore une fois, ces méthodes se présentent comme les *solutions algorithmiques* à un problème de délimitation de régions clairement formalisé et non comme des heuristiques de groupement, la plupart du temps uniquement justifiées par des arguments intuitifs. En outre, le cadre des cocons permet de sélectionner certaines étapes «remarquables» lors du déroulement du groupement hiérarchique.

Le modèle de cocon a initialement été motivé par la remarque de certains défauts dans les formulations «classiques» de la segmentation d'image - par prédicats d'homogénéité. Commençons-donc par décrire ces approches classiques.

5.2 Méthodes de partitionnement en régions homogènes

5.2.1 Mesures d'hétérogénéité et prédicats d'homogénéité

La formulation «classique» de la segmentation s'appuie sur l'idée de partitionnement d'une image en régions homogènes. Cette définition remonte aux premières heures de l'analyse d'image. La définition informelle de la segmentation d'Haralick et Shapiro (1991) citée p. 42 renvoyait implicitement à ce type de formulation.

Notons $\mathcal{I}$ l'ensemble des images de $\mathcal{D}$ dans $\mathcal{V}$ et dotons-nous d'un **prédicat booléen** P sur les parties d'images[1] qui a pour rôle de statuer sur l'«homogénéité» d'une région pour certaines caractéristiques : niveau de gris, teinte, texture... Ce prédicat prend généralement la forme de l'imposition d'un *seuil* s sur une mesure m de l'*hétérogénéité* d'une région, qui traduit une notion intuitive de dispersion des valeurs internes à la région.

[1] Pour alléger les notations, la référence à l'image sera implicite, i.e. P sera noté comme un prédicat sur les parties de $\mathcal{D}$.

Définition 1 *Nous appelons* ***prédicats naturels*** *associés à la fonction* $m : \mathcal{P}(\mathcal{D}) \to \mathbb{R}$*, la famille* $\{P_m^s\}_{s\in\mathbb{R}}$ *des prédicats booléens sur* $\mathcal{P}(\mathcal{D})$ *définie par*

$$\forall R \in \mathcal{P}(\mathcal{D}) \qquad P_m^s(R) \quad \textit{est vrai} \quad \Leftrightarrow \quad m(R) \leq s.$$

Henri Maître (2003, p.293) a relevé «parmi les prédicats les plus utilisés en segmentation» (nous les donnons ici sous l'angle des mesures d'hétérogénéité dont ils dérivent) :

- Contraste global $\quad C(R) = \max\left\{I(x) - I(y) \mid (x,y) \in R^2\right\}$
- Différences limitées (ou contraste local) $\quad c(R) = \max\left\{I(x) - I(y) \mid (x,y) \in R^2, x \leftrightarrow y\right\}$
- Variance $\quad \sigma(R) = \dfrac{1}{|R|}\displaystyle\sum_{x\in R}\left(I(x) - \overline{I_R}\right)^2$
- Distance interquartiles $\quad \ldots$
- Entropie $\quad H(R) = -\displaystyle\sum_{x\in R} Pr(I(x))log(Pr(I(x))$
- ... $\quad$ où $Pr(v)$ désigne la fréquence de la valeur v dans la région.

Étant donné un prédicat P, une partition en régions d'une image est déclarée homogène si toutes ses régions le sont, i.e. vérifient P.

5.2.2 Consistance, croissance et maximalité

L'idée intuitive d'homogénéité est préservée par focalisation sur des sous-parties : si une région est jugée homogène, on s'attend à ce que ses sous-parties le soient également. Un prédicat qui se comporte selon cette intuition sera déclaré **consistant** (pour la focalisation) :

Définition 2 *(Horowitz et Pavlidis 1976) Un prédicat d'homogénéité* P *est* ***consistant*** *si et seulement si*

$$\forall (R,S) \in \mathcal{P}^2(\mathcal{D}) \qquad R \subset S \quad \Rightarrow \quad [P(S) \Rightarrow P(R)].$$

Dans le cas d'un prédicat naturel - défini par seuillage d'une mesure d'hétérogéneité m - on vérifie immédiatement que la consistance du prédicat P_m est équivalente à la **croissance** de m dans le treillis des parties de l'image.

On vérifie aisément que les prédicats naturels associés aux mesures de contraste (global ou local) sont consistants mais que ceux associés aux autres mesures citées ci-dessus ne le sont pas.

Si P est consistant, toute sur-partition d'une partition homogène l'est aussi. Et même si un prédicat n'est pas *systématiquement* consistant, i.e. s'il n'est pas consistant pour tout couple de régions emboîtées, on peut tout de même généralement trouver de nombreuses sur-partitions homogènes d'une partition homogène (il en existe la plupart du temps au moins une, à savoir la sur-partition absolue). Il est donc naturel de rechercher une partition **maximale** pour la propriété d'homogénéité, ce qui conduit à la définition :

Définition 3 ***Une segmentation en régions homogènes*** *d'une image* I *pour un prédicat booléen* P *sur* $\mathcal{P}(\mathcal{D})$ *est une partition* S *de* $\mathbb{P}(\mathcal{D})$

i) *homogène* : $\forall R \in S \qquad P(R)$ *est vrai*

ii) *maximale pour i)* : $\forall T \in \mathbb{P}(\mathcal{D}) \qquad S < T \quad \Rightarrow \quad T$ *n'est pas homogène.*

On utilise généralement (pour des raisons pratiques) une définition plus souple de la maximalité, qui exprime une idée de **maximalité locale**. La condition ii) est alors remplacée par

Définition 4 ***Maximalité locale ou 2-maximalité***

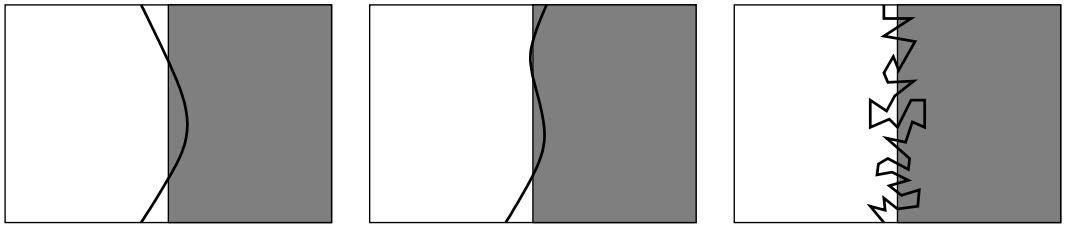
iii) localement maximale pour i) : $\forall T \in \mathbb{P}(\mathcal{D}) \quad S \prec T \quad \Rightarrow \quad T$ *n'est pas homogène.*

Où $\prec$ désigne la relation de couverture entre deux partitions. La maximalité locale restreint la condition d'hétérogénéité des sous-partitions aux partitions couvrant la partition testée, i.e. à celles que l'on peut obtenir par réunion de seulement deux régions. Nous l'appelons également 2-maximalité par parenté avec la notion de 2-normalité employée dans les méthodes par optimisation que nous verrons au chapitre suivant (p. 129).

On voit immédiatement que si P est consistant, la maximalité locale implique la maximalité.

5.2.3 Non unicité d'une solution

En général une segmentation en régions homogènes n'est pas unique sur $\mathbb{P}(\mathcal{D})$. Pour l'illustrer, considérons l'image idéale de la figure 5.1, formée de deux plages de valeurs strictement égales, une blanche et une grise. Si l'on utilise le prédicat naturel P_σ^s associé à la variance d'une région et un seuil s fixé non nul, les bipartitions indiquées par le trait gras 5.1(a),(b) et (c) peuvent être homogènes, ainsi que de nombreuses autres (toutes celles pour lesquelles dans chacune des deux régions la proportion entre points blancs et gris et assez inégale pour que le poids du deuxième mode de l'histogramme soit insuffisant pour que la variance dépasse s). Par ailleurs, si s est fixé de telle manière que l'image totale soit jugée hétérogène, toutes ces partitions sont maximales. Elles sont donc toutes solutions du problème de segmentation tel qu'il vient d'être défini.



(a) (b) (c)

FIG. 5.1 – Non unicité d'une segmentation en régions homogènes.

Comment choisir entre ces différentes solutions ?

5.2.4 Processus de construction de partitions

Comme l'a remarqué Pavlidis (1977), c'est alors le *processus* de construction d'une solution qui est déterminant. Il y deux stratégies principales pour construire une partition :

1. Partir d'une partition initiale et n'effectuer que des modifications de la solution qui conservent la propriété des régions de partitionner l'espace. Une manière naturelle de procéder est de partir d'une partition triviale - la sur-partition absolue ou la sous-partition absolue - et d'effectuer des mouvements, généralement locaux, dans le treillis des partitions :

• Déplacement vers une partition moins fine que la partition courante (**fusion de régions**), généralement vers une partition couvrante (fusion d'un couple de régions).

• Déplacement vers une partition plus fine que la partition courante (**division de régions**), généralement vers une partition couverte (division en deux) ou presque (division en quatre).

2. Remplir progressivement l'espace par des ensembles disjoints, i.e. partir d'une *pré-partition* initiale - un ensemble de parties disjointes mais ne recouvrant pas nécessairement l'espace - et la modifier ou la compléter en conservant la propriété de pré-partitionnement, et ce jusqu'à parvenir à une partition.

Les processus de **croissance de régions** appartiennent à cette seconde catégorie. Un singleton (*germe*) est introduit dans une pré-partition donnée. On fait alors croître ce germe par agglomération itérée de points voisins qui n'appartiennent pas déja à une région. La croissance peut s'effectuer en parallèle à partir d'une population de germes, ou séquentiellement, en ajoutant de nouveaux germes tant que l'ensemble ne partitionne pas l'espace.

Ces procédés sont généralement contraints à produire des partitions en régions connexes. La figure 5.1 ne présente par exemple que des bipartitions en ensembles connexes. Toutes les définitions qui précèdent se transposent immédiatement à ce cas. Dans le cadre d'un algorithme par fusion de régions, on garantit la connexité des régions produites en ne considèrant que les réunions de régions *voisines*.

On peut bien sûr alterner différent types de mouvements, par exemple des fusions et des divisions. Notons toutefois que si les opérations de fusion et division sont inverses l'une de l'autre d'un point de vue algébrique, elles ne le sont pas d'un point de vue informationnel, ce qui se traduit en pratique par des opérations de complexités différentes. En effet, fusionner deux régions voisines correspond à supprimer leur frontière commune, soit à supprimer de l'information. Diviser une région nécessite au contraire d'introduire une nouvelle limite, ce qui correspond à rechercher une information manquante (sauf si les frontières possibles sont prédéfinies comme dans le cas des fusions/divisions par quad trees). Cette assymétrie entre opérations de fusion et de division est bien entendu intimement liée à l'idée de *causalité* des formalismes multi-échelles.

5.2.5 Heuristiques de production d'une «bonne» solution

Comment guider un processus de construction de partition ?

Intuitivement, parmi les segmentations homogènes, celles qui le sont «le plus» sont préférables (ce qui n'est pas formalisé dans la définition ci-dessus). Une heuristique naturelle pour obtenir une «bonne» segmentation est alors de construire les régions les plus homogènes d'abord. Dans le cas d'un processus de fusion itérée de régions, une telle approche revient à parcourir une chaîne du treillis des partitions en sélectionnant à chaque étape le couple de régions dont l'union minimise la valeur m de la mesure d'hétérogénéité. Le processus s'arrête quand la partition est 2-maximale. De même, les processus de croissance agglomèrent en priorité les voisins les plus similaires à la région en croissance, et ce tant que la région reste homogène.

Gagalowicz et Monga (1986) ont formalisé ce type de procédure en considérant un ensemble de couples (prédicat d'homogénéité, fonction de coût). Il s'agit généralement de couples de la forme (P_m^s, m). Chaque couple est successivement utilisé pour fusionner l'image à partir d'une sur-segmentation, en réalisant itérativement l'union de la paire de régions voisines qui possède le plus faible coût et qui vérifie le prédicat (la segmentation trouvée est 2-maximale). Cet algorithme peut donc être considéré comme un algorithme «mixte», s'appuyant à la fois sur un critère à minimiser (approximativement, par une procédure locale) et sur un critère d'arrêt formulé par prédicats. Il est détaillé dans (Horaud et Monga 1995, Ch. 4).

5.2.6 Partitionnement par Quad Tree - Split and Merge

Nous venons d'insister sur la non unicité d'une segmentation en régions homogènes *dans le treillis des partitions d'une image*. Cependant, si l'on considère ce problème sur l'*ensemble des coupes d'une hiérarchie* H et que le prédicat P est *consistant*, alors la solution est unique.

C'est évident d'après ce que nous avons vu au chapitre précédent (section 4.5.4 p. 66) : dans H, une partition maximale pour P_m correspond à une coupe naturelle pour le critère m, qui est croissant quand P_m est consistant. En outre, comme nous l'avons vu, les coupes naturelles sont monotones.

Cette idée est à la base de l'algorithme de Split and Merge d'Horowitz et Pavlidis (1976), dont l'étape de division s'appuie sur un principe de coupe de la hiérarchie régulière associée à une image.

5.2.7 Facteurs géométriques - procédures complexes

La figure 5.1(c) montre une bipartition selon une frontière très irrégulière. Évidemment, toutes choses égales par ailleurs (homogénéité...), une telle partition est moins «agréable» qu'une partition plus régulière géométriquement. La formalisation par prédicat d'homogénéité ne permet cependant pas d'introduire naturellement ce type de facteur.

Dans le même ordre d'idées, le contrôle ne s'exerçant que sur l'homogénéité radiométrique des régions, le résultat est indépendant de leur taille ou de leur nature topologique (trou...). On obtient alors généralement des régions de dimensions très hétérogènes. Il est souvent nécessaire de filtrer les trop petites régions. En pratique, les algorithmes proposés enchaînent plusieurs étapes de pré/post-traitement et/ou combinent différents critères pour former des prédicats composites. On aboutit typiquement à des procédures complexes, dépendant de nombreux paramètres (plus ou moins critiques). Un bon exemple d'une telle procédure est donné par la méthode de Tamaki, Yamamura, et Ohnishi (1999). Ces auteurs proposent, à bon escient, d'introduire différents critères géométriques dans un algorithme de fusion de régions. Cependant, comme pour la méthode de Gagalowicz et Monga (1986), la manière d'enchaîner les critères, le choix des seuils, etc., sont totalement heuristiques. Ces choix relèvent bien entendu du bon sens. Il n'empêche que ce que vise à produire l'algorithme n'est plus clairement formalisé, que l'on n'a aucune garantie d'invariance ni de robustesse, ni aucune idée sur la sensibilité des différents paramètres.

5.2.8 Formulation orientée frontières

Felzenszwalb et Huttenlocher (2001) ont récemment proposé une autre formulation de la segmentation s'appuyant sur l'emploi de prédicats. La décision porte cependant sur les *frontières* de la partition plutôt que sur ses régions.

Définition 5 ***«Bonne» segmentation au sens de Felzenswalb et Hutterlocher.***
Si S est une partition d'une image I, on considère un prédicat booléen sur l'ensemble F des paires de régions voisines de S (frontières) qui possède le sens intuitif suivant :

$P(f \in F)$ est vrai $\Leftrightarrow$ «Il existe une preuve d'existence d'une frontière entre les deux régions adjacentes» («There is an evidence for a boundary»).

Une partition S est alors déclarée

- *trop fine s'il existe une frontière de F pour laquelle P est faux,*
- *trop grossière s'il existe une sur-partition de S qui n'est pas trop fine,*
- *«bonne» si S n'est ni trop fine ni trop grossière.*

Il s'agit d'une version complémentaire de la définition classique orientée régions, au sens de la complémentarité topologique entre régions et frontières. Toutefois, les deux définitions ne sont vraisemblablement pas équivalentes du point de vue des familles de «bonnes» partitions qu'elles déterminent[2]. Notons qu'encore une fois, une «bonne» segmentation pour Felzenswalb et Hutterlocher n'est pas nécessairement unique. Nous verrons ci-dessous la manière dont les auteurs définissent le prédicat «être évidente» sur les frontières de régions.

5.3 Idées à l'origine de l'approche par cocons

5.3.1 Principaux défauts des approches par prédicats

Nous relevons finalement trois défauts majeurs des approches par prédicats d'homogénéité :

[2]Nous avons tenté sans succès de transformer un prédicat régions consistant en un prédicat frontières consistant - notion qui est à définir - de telle manière que les «bonnes» segmentations coincident en régions et en frontières. Il semble que cela ne soit pas possible mais que globalement l'approche frontières soit plus riche : une partie de la réduction dans le sens régions vers frontières est possible. Nous ne détaillerons pas ces développements, qui sont légèrement en dehors de notre propos et qui n'ont pas totalement aboutit.

i) Elles ne formulent pas explicitement une partie de l'objectif poursuivi. Cette partie correspond implicitement à un objectif de minimisation (nous verrons au chapitre 8 que l'on peut formaliser les approches par prédicats comme des cas particuliers de méthodes de minimisation d'énergie). En pratique, c'est le *processus* de recherche d'une solution qui sélectionne la «meilleure» segmentation.

ii) Elles ne permettent pas de prendre en compte naturellement des mesures concernant la géométrie et la topologie des solutions (outre la connexité). Il est alors nécessaire d'introduire des pré/post-traitements ou d'enchaîner différents critères heuristiques pour aboutir à des résultats satisfaisants.

iii) Elles dépendent de manière critique d'au moins un paramètre (un seuil d'homogénéité) qui est directement lié à la dynamique de l'image. Le résultat n'est donc absolument pas robuste à des changements de contraste.

En outre, les approches par prédicats posent a priori l'extraction de régions comme un problème de partitionnement.

5.3.2 Un problème d'ensembles

Le point de départ du modèle de cocon est une interrogation sur la manière d'exprimer la pertinence visuelle d'*une* région individuelle d'image.

Les cocons tentent de formaliser une intuition courante : un certain nombre d'objets visuellement pertinents se caractérisent par le fait qu'il *se distinguent* de leur environnement du point de vue d'une qualité donnée (intensité lumineuse, couleur, texture, ...). Par exemple, une orange posée sur une table - même si l'éclairage provoque des dégradés de lumière sur sa surface bombée - se distingue de son environnement par sa couleur caractéristique. Ou encore une feuille de papier qui serait posée sur la table à côté de l'orange apparaîtrait globalement plus claire que son environnement.

Nous avons alors cherché à caractériser ces zones «remarquables» par des propriétés de bas niveau, i.e. ne faisant pas intervenir de modèle, et «objectives», c'est-à-dire ne dépendant d'aucun paramètre extérieur mais uniquement de caractéristiques intrinsèques à l'image.

Pour fixer une terminologie, disons d'une région qu'elle est *contrastée* quand elle possède la propriété - encore informelle - de se distinguer de son environnement pour une caractéristique visuelle donnée.

5.3.3 Critères relatifs - contraste simultané

Comme nous l'avons vu, l'approche classique de la segmentation repose sur l'emploi de *prédicats absolus* sur les caractéristiques d'une région, prédicats tels que «être bleue», «être claire» ou «être homogène en intensité». Une zone pertinente pour un prédicat P est alors définie comme une région R *maximale* pour P, c'est-à-dire vérifiant P mais telle que tout élargissement de R viole P. Or, par exemple, la propriété d'«être bleue» est éminement subjective. Elle repose nécessairement sur une décision, sur la position de la frontière entre le «bleu» et le «non bleu», soit sur l'emploi de paramètres subjectifs.

L'idée qui nous a guidée dans la mise en place du critère de cocon est que pour espérer pouvoir aboutir à une caractérisation *objective* de zones pertinentes, il est nécessaire de s'en remettre à un critère **relatif**.

C'est à nouveau Henri Poincaré qui nous a influencé sur ce point. En effet, d'après Poincaré «les sensations sont intransmissibles, ou plutôt tout ce qui est qualité pure en elles est intransmissible et à jamais impénétrable. Mais il n'en est pas de même des relations entre ces sensations. (...) Rien n'est objectif que ce qui est identique pour tous[3]. Or on ne peut parler d'une pareille identité que si une comparaison est possible» (Poincaré 1905, pp.179-180). L'intransmissibilité des sensations rappelle que l'existence d'un monde réel extérieur est totalement indécidable. Par contre, on voit que les *relations* entre sensations peuvent vérifier un principe d'invariance entre observateurs : n'est objectif (c'est-à-dire - d'un point de vue réaliste - n'est susceptible de correspondre à un phénomène réel autonome) que ce qui est indépendant d'un observateur spécifique, et donc comparable entre les différents observateurs (par une communication utilisant un langage commun). On peut donc conclure que, comme seules les *relations*

[3]Cette notion d'objectivité correspond à ce qu'Einstein appelait inter-subjectivité.

entre sensations sont transmissibles, elles seules sont susceptibles d'être objectives : on peut se mettre d'accord sur le fait qu'une certaine zone d'image est *plus bleue* que son environnement mais pas sur le fait qu'elle soit *bleue* dans l'absolu.

On aboutit au même genre d'idée quand on s'intéresse non pas à l'invariance par changement d'observateur mais à l'invariance de notre capacité à reconnaître les objets face à des changements de conditions d'observation : intensité lumineuse, point de vue, etc. Le principe, que David Marr (1982, p. 259) appelle «contraste simultané», semble à nouveau remonter à Ernst Mach (1886, note p. 76) : «un *objet visible*, sous des conditions de luminosité d'intensité variable, peut être reconnu *comme étant le même* à la condition que la sensation excitée dépende du *ratio* entre l'intensité illuminant l'objet et celle illuminant son environnement». L'expression sous forme de ratio de la «sensation excitée», i.e. de la variable sur laquelle s'appuie le fait que nous discernions un objet, provient d'un soucis d'invariance par transformation linéaire de l'intensité lumineuse globale. Mais ce n'est pas tant cette expression sous forme de ratio qui nous intéresse pour l'instant[4] que le fait que la décision doive s'appuyer sur une comparaison entre le stimulus émanant de l'objet et celui émanant de son environnement. Appuyons nous encore une fois sur les propos de Marr (1982, p.259) : «Imaginez que vous vous promeniez le long d'un parterre sur lequel une fleur jaune ou bleue s'épanouit au milieu d'une étendue d'herbe verte. Bien que les caractéristiques spectrales *absolues* de la lumière provenant de la fleur ne fournissent aucune information fiable quant aux caractéristiques de sa surface, tant du point de vue de l'intensité que des propriétés colorimétriques, ses caractéristiques *relatives* aux surfaces environnantes sont, elles, probablement fiables. Si la fleur parait plus claire que l'herbe, c'est probablement dû à des caractéristiques de la fleur et non à son éclairage (bien que le soleil puisse à cet instant n'éclairer que la corolle de la fleur). Si la fleur semble plus bleue ou plus jaune que l'herbe, c'est probablement vrai»[5] (l'emphase mise sur les mots *absolues* et *relatives* est de nous).

Malgré cela, les critères classiques de segmentation utilisent des prédicats portant sur l'aspect absolu des régions. La décision est en quelque sorte trop hâtive pour être objective, ou robuste.

Voyons comment l'idée de zone contrastée par rapport à son environnement peut être formalisée dans le cadre de la théorie des graphes, et que cette formalisation conduit - dans un cas strict - à des solutions multi-échelles (hiérarchisées) qui sont effectivement robustes à une grande classe de transformations des données d'entrée.

5.4 Sous-graphes contrastés connexes (cocons)

5.4.1 Définition

Groupement sur un graphe attribué par une dissemblance

La notion de cocon est définie sur un triplet (X, U, d), où $G = (X, U)$ est un graphe non orienté et d une mesure de dissemblance portée par les arêtes de G.

• X représente les objets élémentaires que l'on cherche à regrouper en ensembles «pertinents» (e.g. les pixels d'une image ou les régions d'une segmentation fine de l'image).

• $U \subset \{\{x, y\} | (x, y) \in X^2\}$ est un ensemble d'arêtes qui définissent une topologie sur X.

[4] Elle conduit à la célèbre loi psychophysique de Fechner reliant la «sensation excitée» au logarithme de l'intensité du stimulus. Nous verrons ci-dessous l'importance de ce type de ratio dans un cadre multi-échelles.

[5] «Suppose you pass an embankment where a yellow or blue flower happens to be growing amid a background of green grass and clover. Although the absolute spectral characteristics of the light coming from the flower cannot at all be relied upon as a clue to its surface characteristics, either in the matter of its lightness or of its spectral properties, nevertheless its characteristics relative to other nearby surfaces probably are reliable. If the flower appears lighter than the grass, than this is probably due to a characteristic of the flower and not of the illumination (though the head of the flower could be just catching the sun). If the flower looks bluer than the grass, than it probably really is. If it looks yellower, then, again, it probably really is»(Marr 1982, p.259).

• d est une fonction de U dans un intervalle réel ouvert $\Delta \triangleq]\Delta_0, \Delta_1[\subset \mathbb{R}$ non vide (nous notons comme précédemment $\overline{\Delta} \triangleq [\Delta_0, \Delta_1]$ la fermeture de Δ).

U doit être vue comme une relation de voisinage qui définit la topologie de X (bien que U ne sera que rarement choisie réflexive). Cette topologie est extrêmement importante pour notre modèle : nous verrons que la notion de cocon - qui repose sur les ensembles *connexes* induits par le graphe G - n'a que peu d'intéret si G est un graphe dense.

G est un graphe non orienté[6]. Nous aurons cependant occasionnellement besoin de considérer une arête prise dans un sens précis. Si e est une arête vue dans un sens donné, i.e. si $e = (i, f)$ avec $\{i, f\} \in U$, e^- désignera sont nœud initial, i, et e^+ son nœud final, f.

Le rôle de la fonction de dissemblance d est de qualifier le degré de différence accordé aux objets voisins dans G. Selon l'origine du graphe et la nature du problème de groupement, d pourra traduire une notion d'éloignement des objets ou de certains de leurs attributs, de réticence à leur groupement, de probabilité que les régions élémentaires appartiennent à des formes distinctes, etc.

Il est important de noter que la définition de d est restreinte aux seules arêtes présentes dans G. En effet, les mesures de dissemblance sont habituellement définies sur *toutes les paires* d'objets élémentaires, et on peut généralement simuler l'inexistence d'une arête en lui affectant une dissemblance extrême (Δ_1). Dans notre modèle, les dissemblances extrêmes Δ_0 et Δ_1 ont un statut particulier et l'absence d'arête traduit réellement la non comparabilité de deux objets[7].

Pour fixer les idées, pour la segmentation d'image, G sera un graphe d'adjacence de régions (RAG) construit sur une segmentation fine de l'image, modélisant ainsi la topologie du support de l'image. d mesurera alors la différence de contenu image entre deux régions élémentaires voisines, typiquement l'éloignement de leurs distributions radiométriques ou le gradient le long de leur frontière commune. Les mesures spécifiques que nous avons utilisé n'ont cependant pas d'importance pour le moment : toutes les définitions et propriétés relatives aux cocons qui vont être étudiées sont valables pour des graphes et des mesures de dissemblance quelconques et s'appliquent donc à tout type de problème de groupement sur un triplet (X, U, d).

Connexions, sous-graphes induits et cocycles

Définition 6 *Soit $G = (X, U)$ un graphe.*

• *Pour deux ensembles de noeuds $(A, B) \in \mathcal{P}^2(X)$, l'ensemble des arêtes défini par*

$$A \sim B \quad \triangleq \quad \{u \in U \mid u \subset A \cup B, \ u \cap A \neq \emptyset \ \text{ et } \ u \cap B \neq \emptyset\}$$

est appelé l'ensemble des ***connexions*** *entre A et B.*

• *Pour alléger l'écriture, $A \sim A$ sera noté $\widetilde{A}$, et sera appelé l'ensemble des* ***connexions internes*** *au groupe A. Quand le graphe de référence devra être explicite, nous noterons également $U_A \triangleq \widetilde{A}$ l'ensemble des connexions internes à A pour $G = (X, U)$.*

• *Pour tout $A \subset X$, le sous-graphe $G_A \triangleq (A, U_A)$ est appelé le sous-graphe de $G = (X, U)$* ***induit*** *par A.*

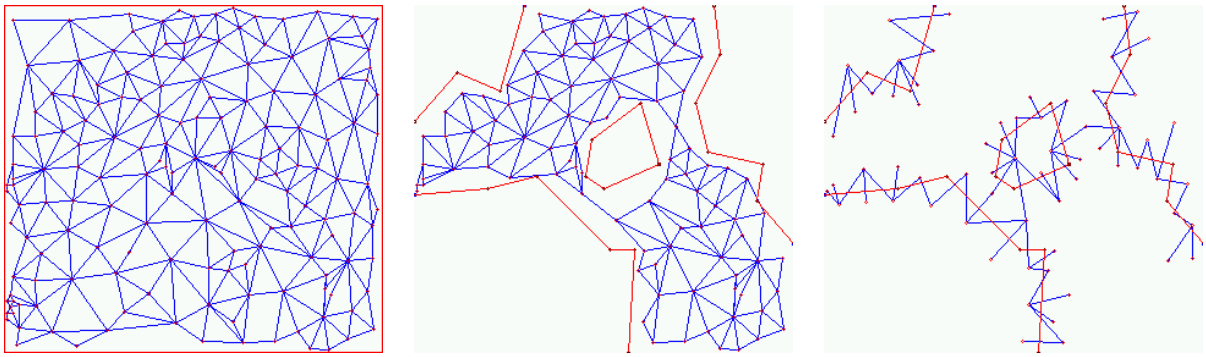
Notons que si $A \cap B = \emptyset$, la définition de $A \sim B$ se réduit à

$$A \sim B \quad = \quad \{u \in U \mid u \cap A \neq \emptyset \text{ et } u \cap B \neq \emptyset\}.$$

Nous nous intéresserons essentiellement dans ce chapitre à des sous-graphes induits. Nous ne ferons alors fréquemment pas la différence entre un ensemble de noeuds et le sous-graphe qu'il induit. Cette définition est illustrée par la figure 5.2(b).

[6] Nous aurions pu définir G comme un graphe orienté - i.e. dont les arêtes sont des couples de nœuds- et *symétrique*, et imposer que d soit également symétrique. La définition non-orientée simplifie toutefois la présentation.

[7] Notons qu'utiliser une dissemblance est la façon la plus simple d'introduire les cocons. Nous en verrons ci-dessous une définition plus générale, utilisant des fonctions d'ensemble sur U (ou énergies sur les parties de U).



(a) (b) (c)

FIG. 5.2 – Sous graphe induit et cocycle. (a) Un graphe d'adjacence de régions. (b) Un sous-graphe induit. (c) Cocycle correspondant à (b).

Définition 7 *L'ensemble $Y \sim X \backslash Y$ est appelé le* ***cocycle*** *de Y en théorie des graphes.*

Si G est un RAG, le cocycle de Y représente l'ensemble des arêtes traversant la frontière de Y, ce qui est illustré par la figure 5.2(c).

En tant que concept orienté, nous utiliserons la notation habituelle $\omega(Y)$ pour désigner le cocycle de Y dans lequel chaque arête e de $\omega(Y)$ est orientée dans le sens sortant de Y, i.e. de telle sorte que $e^- \in Y$:

Définition 8 $\omega(Y) \triangleq \{(x,y) \in X^2 \mid \{x,y\} \in U, x \in Y, y \notin Y\}.$

Avec cette notation,

Définition 9 *Le* ***voisinage*** *$\mathcal{V}(Y)$ de $Y \subset X$ est défini par $\mathcal{V}(Y) \triangleq \{e^+ | e \in \omega(Y)\}$.*

Statistiques internes et externes

La définition d'une famille de cocons sur (X, U, d) dépend de deux fonctions statistiques f et g sur des échantillons de nombres réels. Ces statistiques ont pour rôle de synthétiser les collections[8] de dissemblances portées par des ensembles d'arêtes. Pour fixer les idées, les différentes statistiques qui seront envisagées par la suite sont :

- la valeur minimale et maximale (resp. notées min et max),
- la valeur moyenne (notée moy)
- les q-quantiles, avec $q \in [0,1]$, (notés $[q]$) et en particulier pour $q = 1/2$, la médiane (notée med).

f et g opèrent sur des échantillons de réels, cependant, pour alléger les notations, quand E sera un ensemble d'arêtes du graphe, alors $f(E)$ (resp. $g(E)$) désignera la valeur de f (resp. g) appliquée à la collection des dissemblances portées par les arêtes de E. Autrement dit

$$f(E) \triangleq f(\ [d(e), e \in E]\ ).$$

[8] Plusieurs arêtes d'un ensemble peuvent porter les mêmes valeurs de dissemblance, c'est pourquoi nous parlons de *collections* ou d'*échantillons* de dissemblances, qui sont des «paquets» d'objets de même nature qui peuvent comporter plusieurs occurrences d'un même objet, contrairement aux ensembles classiques. Nous notons les collections entre crochets. La collection vide est également notée $\emptyset$.

Pour une raison qui apparaîtra clairement par la suite, nous supposons que f et g vérifient

$$\text{i)} \quad \forall\, C = [x_i], x_i \in \Delta \quad C \neq \emptyset \;\Rightarrow\; f(C) \in \Delta \text{ et } g(C) \in \Delta \quad : f \text{ et } g \text{ sont internes à } \Delta \tag{5.1}$$

$$\text{ii)} \quad f(\emptyset) = \Delta_0 \quad : f \text{ est minimale sur le vide} \tag{5.2}$$

$$\text{iii)} \quad g(\emptyset) = \Delta_1 \quad : g \text{ est maximale sur le vide} \tag{5.3}$$

Sous-graphes contrastés et cocons

Avec ces notations nous pouvons définir les objets principaux de ce chapitre :

Définition 10 *Soit $G = (X, U, d)$ un graphe attribué par une dissemblance d à valeurs dans Δ et f et g deux statistiques vérifiant 5.1, 5.2 et 5.3. Pour tout ensemble $Y \subset X$ de noeuds de G*

i. Le ***contraste interne*** I_f *de Y (ou indifféremment de G_Y) est défini comme la valeur de la statistique f des dissemblances des arêtes du sous-graphe induit par Y, i.e. par*

$$\mathrm{I}_f(Y) \quad \triangleq \quad f\big(Y \sim Y\big) \quad = \quad f\big(\widetilde{Y}\big). \tag{5.4}$$

ii. Le ***contraste externe*** E_g *de Y est défini comme la valeur de la statistique g des dissemblances du cocycle de Y, i.e. par*

$$\mathrm{E}_g(Y) \quad \triangleq \quad g\big(Y \sim X \backslash Y\big). \tag{5.5}$$

iii. Y est ***contrasté*** *pour f et g si et seulement si*

$$\mathrm{I}_f(Y) \quad \leq \quad \mathrm{E}_g(Y). \tag{5.6}$$

(quand l'inégalité est stricte, Y est dit strictement contrasté).

iv. Y est un ***f-g-cocon*** *(strict) de G si et seulement si*

a) Y est (strictement) ***contrasté*** *pour f et g*

b) G_Y est un sous-graphe ***connexe****.*

v. $\mathcal{C}_f^g(G)$ désigne l'ensemble des f-g-cocons stricts de G privé de l'ensemble vide.

Ces définitions sont illustrées par la figure 5.3.

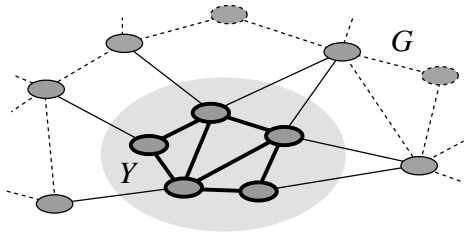


FIG. 5.3 – Illustration de la définition de cocon dans le cas d'un graphe attribué par une distance. Un f-g-cocon Y d'un graphe G construit sur des points du plan et attribué par la distance entre points voisins est un sous-ensemble connexe de points pour lequel la statistique f des distances entre points de l'ensemble (arêtes grasses) est inférieure à la statistique g des distances avec l'extérieur (arêtes fines). Le reste du graphe (en pointillé) n'intervient pas dans la définition. L'ensemble Y représenté ici est visiblement un *max-min*-cocon strict et donc également un *moy-moy* ou un *med-med*-cocon strict.

Par la suite, lorqu'il n'y aura aucune ambiguité sur les fonctions f et g utilisées nous écrirons simplement «contraste», «cocon», etc. sans préciser quelles statistiques sont employées.

5.4.2 Interprétation

Les cocons modélisent des objets dont les parties internes voisines se ressemblent globalement plus (au sens de f) qu'elle ne ressemblent à l'environnement de l'objet (au sens de g). Cette définition modélise bien des sous-parties qui se démarquent de leur environnement : leur frontière externe est globalement plus marquée que leurs frontières internes[9]. Remarquons que les différences entre parties élémentaires (nœuds) sont toutes considérées localement, à travers la relation de voisinage.

Dans le cas de l'image, le graphe correspond à une relation de voisinage spatial entre parties élémentaires insécables (atomes : pixels ou régions d'une partition fine). La façon la plus naturelle de construire des dissemblances consiste dans ce cas à refléter la différence de contenu radiométrique entre atomes voisins. De telles dissemblances correspondent alors à mesurer la variation radiométrique encourue quand on effectue un déplacement spatial élémentaire. Elles correspondent donc à la notion intuitive de dérivée directionnelle. Par exemple, si le graphe considéré est le graphe de 4 ou 8-voisinage sur les pixels d'une image I, la dissemblance entre deux pixels voisins x et y peut être calculée par $|I(x) - I(y)|$ (ou par une forme plus robuste de dérivée discrète). Un cocon est alors une région telle que les variations de l'image sont globalement plus rapides (au sens des statistiques f et g) quand on traverse la frontière de la région que quand on reste à l'intérieur[10]. Cette interprétation des cocons comme zones plus «lisses» à l'intérieur que sur leur frontière vaut pour des dissemblances traduisant des différences élémentaires. Nous verrons ci-dessous d'autres interprétations pour d'autres types de dissemblances.

5.4.3 Cocons triviaux

Pour tout graphe attribué $G = (X, U, d)$, et quelles que soient les fonctions f et g considérées (qui vérifient les propriétés 5.1 à 5.3) on constate que :

1. L'ensemble vide est un cocon strict.
2. Tout sommet de G est un cocon strict, sauf éventuellement s'il est le support d'une boucle.
3. Toute composante connexe de G est un cocon strict.
4. Le graphe G entier est un cocon strict si et seulement si il est connexe.

En effet, $\forall f, g$

1. $f(\emptyset) = \Delta_0 < g(\emptyset) = \Delta_1$.
2. Si x ne porte pas de boucle : $f(\widetilde{\{x\}}) = f(\emptyset) = \Delta_0 < \Delta_1 < g(\{x\} \sim X \backslash \{x\})$.
3. Si Y est une composante connexe de G : $f(\widetilde{Y}) < g(Y \sim X \backslash Y) = g(\emptyset) = \Delta_1$.
4. Si G n'est pas connexe, par définition il ne peut pas être un cocon, et s'il est connexe, la propriété 3 assure que c'est un cocon.

L'ensemble vide n'est évidemment pas d'un grand intérêt, c'est pourquoi l'ensemble des cocons en est privé. En outre, sauf mention explicite du contraire, les graphes considérés seront supposés *sans boucle*.

[9] Pour la petite histoire, c'est ma femme qui - sans le vouloir - m'a suggéré le terme de «cocon». En effet, à l'époque des premiers «cocons», elle était enceinte et portait des vêtements de grossesse de marque «cocoon»... «cocon» m'est alors apparu comme une abbréviation éloquente de «connexe contrasté». Le sens courant du mot cocon (enveloppe soyeuse de certaines chrysalides dixit le Petit Larousse) traduit en effet relativement bien la nature de ces objets mathématiques : il s'agit de sous-ensembles confinés, séparés de leur environnement par une enveloppe plus «résistante» que les cloisons internes.

[10] On peut imaginer une transposition de cette idée au continu, en transformant la différence finie selon une arête (x, y) en dérivée en x dans une direction xy donnée de l'espace. Le contraste «interne» d'un ensemble Y peut alors se transposer en la statistique f des valeurs absolues des dérivées directionnelles internes à Y, et son contraste «externe» en la statistique g des valeurs absolues des dérivées le long de sa frontière, mesurés dans la direction normale (il ne s'agit que d'une suggestion intuitive : nous n'avons pas rigoureusement étudié cette transposition).

5.4.4 Variantes : cocons complets et complets-complets

Les cocons complets sont une variante des cocons.

On considère toujours un graphe $G = (X, U, d)$ comme le graphe d'adjacence des éléments de X, c'est-à-dire le graphe à travers lequel on envisage la relation de voisinage spatial. Par ailleurs, on définit le graphe complet $K_X = (X, X^2, d')$ sur X, dont les arêtes sont attribuées par une fonction d' qui coincide avec d sur U[11]. Comme nous définissons ici G puis K_X, d' est vu comme un prolongement de d. Dans la plupart des cas, c'est en fait d qui provient d'une restriction d'une dissemblance d' au graphe partiel G de K_X.

Définition 11 *On définit le* ***contraste interne complet*** *d'un sous-graphe induit $G_Y = (Y, U_Y)$ par :*

$$\mathrm{I}_f^K(G_Y) \quad \triangleq \quad \mathrm{I}_f(K_Y) \quad = \quad f(Y^2). \tag{5.7}$$

Un ***cocon complet*** *de $G' = (X, U, d')$ est alors un sous-graphe G_Y connexe du point de vue de l'adjacence U et qui vérifie la relation de contraste complet :*

$$\mathrm{I}_f^K(Y) \quad \leq \quad \mathrm{E}_g(Y). \tag{5.8}$$

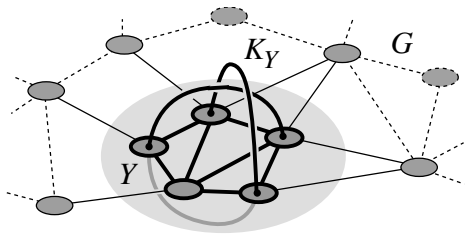


FIG. 5.4 – Un cocon complet. Son contraste interne est calculé sur le graphe complet K_Y induit par Y (en gras). Par ailleurs le contraste externe est toujours restreint au graphe d'adjacence spatiale (trait fin).

Dans un cocon complet, la mesure de contraste interne n'est pas restreinte au graphe d'adjacence «spatiale» mais la comparaison est étendue à tous les couples possibles du sous-graphe. On impose ainsi que la ressemblance interne soit indépendante de la proximité des éléments dans G. Par contre, la connexité s'entend toujours au sens de l'adjacence spatiale et l'on effectue toujours la mesure du contraste externe sur la frontière de l'ensemble (voir figure 5.4).

Évidemment, si G est déja le graphe complet, cocons «simples» et «complets» coïncident. Comme nous le verrons ci-dessous, la notion de cocon n'a toutefois que peu d'intérêt sur un graphe complet. Notons néanmoins une façon naturelle d'étendre également la notion de contraste *externe* en une notion «complète», sans pour autant transformer G en K_X :

Définition 12 *On définit le* ***contraste externe complet*** *d'un sous-graphe induit $G_Y = (Y, U_Y)$ par :*

$$\mathrm{E}_g^K(G_Y) \quad \triangleq \quad \mathrm{I}_g^K(G_{Y \cup \mathcal{V}(Y)}) \quad = \quad g\left((Y \cup \mathcal{V}(Y))^2\right). \tag{5.9}$$

Un ***cocon complet-complet*** *de $G' = (X, U, d')$ est alors un sous-graphe G_Y connexe du point de vue de l'adjacence U et qui vérifie la relation de contraste complet-complet :*

$$\begin{aligned} \mathrm{I}_f^K(Y) &\leq \mathrm{E}_g^K(Y) \\ \Leftrightarrow \quad f(Y^2) &\leq g\left((Y \cup \mathcal{V}(Y))^2\right) \end{aligned} \tag{5.10}$$

[11] K_X est un graphe non orienté. Par abus de notation, nous notons X^2 l'ensemble des *paires* d'éléments de X.

En français, dilater de 1 un cocon complet-complet, par ajout de tous ses voisins, augmente globalement la dissemblance entre ses éléments internes, c'est-à-dire introduit dans l'ensemble des éléments globalement moins ressemblants (entre eux ou avec les éléments de l'ensemble) que ne le sont les éléments de l'ensemble. Malgré le fait que l'on compare tous les éléments deux à deux, la topologie induite par G joue un rôle extrêmement important : c'est elle qui détermine la notion de voisinage et donc la connexité et la manière dont croît un ensemble connexe.

Dans le cadre de l'image, nous avons vu que pour certaines dissemblances sur un RAG les cocons simples pouvaient s'interpréter comme des zones «lisses» localement maximales d'une image. De leur côté, les cocons complets-complets modélisent des zones «homogènes» localement maximales. Les cocons complets correspondent à des objets intermédiaires.

5.5 La hiérarchie des cocons durs (max-min)

Cette section s'intéresse à une première famille de cocons, appelés **cocons durs**[12]. Ces cocons sont définis par les statistiques $f = max$ et $g = min$. C'est sous cette forme que nous avons initialement imaginé la notion de contraste. Ce sont alors les propriétés remarquables de ces cocons durs et leur relation avec des méthodes classiques de groupement hiérarchique qui sont à l'origine de l'élargissement ultérieur de la définition de cocon à d'autres fonctions f et g.

5.5.1 Définition et interprétation

Définition 13 *Un max-min-cocon strict d'un graphe pondéré $G = (X, U, d)$ est appelé un* ***cocon dur*** *de G. Un cocon dur Y est donc un ensemble connexe qui vérifie :*

$$\max\{d(e) \mid e \in \widetilde{Y}\} \quad < \quad \min\{d(e) \mid e \in Y \sim X \backslash Y\}. \tag{5.11}$$

On constate immédiatement que les fonctions max et min vérifient les propriétés 5.1 à 5.3 requises sur f et g : le max et le min appartiennent à l'intervalle de variation de leurs arguments, et si $\Delta \triangleq]\Delta_0, \Delta_1[$ alors il est naturel de poser $\max(\emptyset) \triangleq \Delta_0$ et $\min(\emptyset) \triangleq \Delta_1$.

Les cocons durs modélisent des ensembles dont la plus forte dissemblance interne est *strictement* plus faible que la plus faible dissemblance externe. Ainsi, si la dissemblance traduit les variations locales d'une image, alors les cocons durs sont les groupes connexes de pixels tels que la plus forte variation interne au groupe est plus faible que la plus faible variation sur la frontière du groupe. Autrement dit, un cocon dur possède une frontière extérieure nette, partout plus nette que n'importe laquelle de ses frontières internes.

Les cocons durs peuvent strictement s'envisager comme des portions d'image «lisses» localement maximales, i.e. telle que tout élargissement les rend moins lisses. En effet, toute sur-partie connexe d'un cocon Y contient un des voisins de Y et donc une des arêtes de $Y \sim X \backslash Y$. Adjoindre l'une quelconque d'entre elles à l'intérieur de l'ensemble augmente son contraste interne.

Notons que quel que soit le graphe $G = (X, U, d)$, et quel que soit $Y \subset X$, on a $U_Y \subset Y^2$ et donc $\mathrm{I}_{\max}(Y) \leq \mathrm{I}^K_{\max}(Y)$. **Les cocons complets durs constituent donc un sous-ensemble des cocons durs**.

Quand G est déjà complètement connecté, cocons «simples» et «complets» se confondent. Il est toutefois important de remarquer que la notion de cocon dur n'a d'intéret réel que pour les graphes *peu denses*, et c'est là toute l'importance de la restriction de la dissemblance d aux seules arêtes de U. Prenons en effet l'exemple d'un graphe complètement connecté dont les points sont des points d'un espace métrique et où d mesure la distance entre deux points. Un cocon dur est alors un ensemble dont le diamètre est plus faible que sa distance aux points qui ne lui appartiennent pas. Autrement dit, il s'agit d'un ensemble isolé du reste du monde par un vide au moins égal à sa taille. C'est évidemment une condition très restrictive, qui ne tolère que les groupes remarquablement compacts et isolés. Par contre, si les points du graphe ne

[12] Dans (Guigues et al. 2001a) et (Guigues et al. 2003c) les cocons durs sont simplement appelés cocons.

sont que localement connectés, i.e. si le graphe est un graphe de proximité (un graphe de k plus proches voisins, une triangulation de Delaunay, un graphe de Gabriel, etc.), alors les cocons durs sont pertinents : ils représentent les ensembles dont les «vides» internes sont plus faibles que les vides qui les séparent du reste du monde. Autrement dit, il s'agit de lieux d'agglutination, correspondant à des maxima locaux pour la densité de points.

Notons également que **l'ensemble des cocons complet-complet durs est toujours vide**. En effet, pour tout $Y \subset X$ on a $Y^2 \subset (Y \cup \mathcal{V}(Y))^2$ et donc $\mathrm{I}^K_{\max}(Y) \geq \mathrm{E}^K_{\min}(Y)$. La notion de cocon complet-complet n'a donc d'intérêt que pour des couples de statistiques (f, g) telles que si $A \subset B$ alors $g(B^2)$ peut être supérieur à $f(A^2)$. Il ne faut cependant pas que cette propriété soit systématique si l'on ne veut pas que tous les ensembles soient des cocons. $f = g = max$ n'est donc pas satisfaisant. Par contre utiliser $f = g = moy$ ou $f = g = med$ peut être intéressant.

5.5.2 Propriétés

Hiérarchisation

Théorème 1 *Pour tout graphe attribué $G = (X, U, d)$ connexe et sans boucle, l'ensemble des max-min-cocons stricts de G est une hiérarchie sur X.*

Preuve : Les singletons et le graphe entier sont toujours des cocons qui forment la base et le sommet de la hiérarchie. La preuve consiste alors à établir que deux sous-graphes connexes non hiérarchisés ne peuvent pas simultanément être deux cocons.

Soit $(A, B) \in \mathcal{P}^2(X)$ deux sous-ensembles connexes de noeuds tels que ni $A \cap B = \emptyset$, ni $A \subset B$, ni $B \subset A$. Comme $A \cap B \neq \emptyset$ et $A \backslash B \neq \emptyset$ lui est connecté, il existe une arête $e_1 \in \widetilde{A}$ telle que (voir fig. 5.5)

$$e_1 \cap (A \cap B) \neq \emptyset \quad \text{et} \quad e_1 \cap (A \backslash B) \neq \emptyset.$$

De la même manière, on trouve toujours une arête $e_2 \in \widetilde{B}$ telle que

$$e_2 \cap (A \cap B) \neq \emptyset \quad \text{et} \quad e_2 \cap (B \backslash A) \neq \emptyset.$$

On a donc $e_1 \in B \sim X \backslash B$ et $e_2 \in A \sim X \backslash A$.

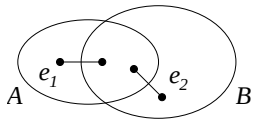


FIG. 5.5 – Arêtes considérées dans la démonstration de hiérarchisation des cocons durs.

Des définitions des contrastes internes et externes de A et B, on déduit alors que

$$\begin{array}{lll} \mathrm{I}_{\max}(A) \geqslant d(e_1) & \text{et} \quad \mathrm{E}_{\min}(A) \leqslant d(e_2) & \text{d'une part, et} \\ \mathrm{I}_{\max}(B) \geqslant d(e_2) & \text{et} \quad \mathrm{E}_{\min}(B) \leqslant d(e_1) & \text{d'autre part.} \end{array}$$

Ce qui fournit $\mathrm{E}_{\min}(B) \leqslant \mathrm{I}_{\max}(A)$ et $\mathrm{E}_{\min}(A) \leqslant \mathrm{I}_{\max}(B)$. Si l'on suppose finalement que A est *max-min*-cocon *strict*, i.e. que $\mathrm{I}_{\max}(A) < \mathrm{E}_{\min}(A)$, on obtient alors l'inégalité

$$\mathrm{E}_{\min}(B) \leqslant \mathrm{I}_{\max}(A) < \mathrm{E}_{\min}(A) \leqslant \mathrm{I}_{\max}(B)$$

qui interdit que B soit lui aussi un cocon. ■

Dans la démonstration qui précède, la condition de **connexité** des cocons est essentielle pour assurer leur hiérarchisation. En effet, on construit aisément des ensembles strictement max-min-contrastés non

connexes et non hiérarchisés : Considérons trois sous-graphes constrastés Y_1,Y_2 et Y_3, non vides et non *voisins* (tels que $Y_1 \sim Y_2 = \emptyset$ et $Y_2 \sim Y_3 = \emptyset$). Dans ce cas, si $\mathrm{I}_{\max}(Y_1) < \mathrm{E}_{\min}(Y_2)$ et $\mathrm{I}_{\max}(Y_2) < \mathrm{E}_{\min}(Y_1)$, alors $C_1 = Y_1 \cup Y_2$ est contrasté. Par symétrie, $C_2 = Y_2 \cup Y_3$ est également contrasté et C_1 et C_2 ne sont pas hiérarchisés.

La condition de connexité de G lui-même n'est pas essentielle, elle, mais simplifie l'énoncé du théorème. En effet, si G est non connexe, on peut toujours en faire l'analyse composante connexe par composante connexe. L'ensemble des cocons d'un graphe non connexe n'est pas une hiérarchie au sens strict, mais comme les cocons sont connexes, c'est une union de hiérarchies disjointes (une forêt si l'on représente les hiérarchies par des arbres).

L'importance de la condition de connexité pour obtenir une hiérarchie a peut-être un lien avec la théorie morphologique des *opérateurs connexes*, qui élargissent systématiquement les zones plates d'une image et permettent donc de définir des hiérarchies (Serra et Salembier 1993; Salembier et Serra 1995). Cette relation éventuelle serait à creuser. Nous verrons également ci-dessous que la hiérarchisation des *composantes connexes* des ensembles de niveau d'une image mono-canal (Matheron 1975) s'obtient comme un cas particulier de hiérarchie dure.

Notons également que seul l'ensemble des cocons **stricts** est hierarchisé, ce qui peut être mis en évidence en considérant un graphe sur lequel toutes les valeurs de dissimilarités sont égales. Tous ses sous-graphes connexes sont alors contrastés *au sens large*, et ne forment donc pas une hiérarchie (si le graphe possède un minimum d'arêtes...).

Distance ultramétrique

Comme nous l'avons vu, les structures hiérarchiques sont intimement liées à celles d'espaces ultramétriques. Une distance ultramétrique associée à la hiérarchie des cocons durs est directement donnée par le contraste interne des cocons.

Proposition 1 *Le contraste interne du plus petit cocon contenant deux noeuds est une distance ultramétrique associée à $C^{min}_{max}(G)$. Formellement,*

$$\forall (x,y) \in X^2 \qquad \delta(x,y) = \mathrm{I}_{\max}\left(\bigcap \left\{C \in \mathcal{C}^{min}_{max}(G) | \{x,y\} \subset C\right\}\right) \tag{5.12}$$

est une distance sur X qui vérifie

$$\forall (x,y,z) \in X^3 \qquad \delta(x,y) \leqslant \max\{\delta(x,z), \delta(z,y)\}. \tag{5.13}$$

et $E = (X, \delta)$ est un espace ultramétrique dont les boules sont les cocons durs de G.

Preuve : $\delta(x,y)$ est bien définie pour tout (x,y) car X est un cocon. Elle est positive, symétrique et nulle si $x = y$ car $\widetilde{\{x\}} = \emptyset$. Concentrons-nous sur la vérification de l'inégalité «ultratriangulaire» (5.13).

Pour tout $Y \subset X$, notons C_Y le plus petit cocon contenant Y. Soit $(x,y,z) \in X^3$. Posons $C_1 = C_{\{x,y\}}$, $C_2 = C_{\{y,z\}}$, $C_3 = C_{\{x,z\}}$ et $C = C_{\{x,y,z\}}$. Deux cas différents se présentent alors :

a) Soit $C_1 = C_2 = C_3 = C$.

Alors $\delta(x,y) = \delta(x,z) = \delta(y,z)$ et (5.13) est immédiatement vérifiée.

b) Soit $\exists p : C_p \neq C$.

Comme les cocons sont hiérarchisés et comme $\forall i \neq p : C_i \cap C_p \neq \emptyset$, il vient

$$\forall i \neq p : \quad C_i \subset C_p \quad \text{ou} \quad C_p \subset C_i.$$

Or

$$\forall i \neq j : C_i \subset C_j \Rightarrow \{x,y,z\} \subset C_j.$$

Comme par ailleurs $\forall j : C_j \subset C$, on obtient

$$\forall i \neq j : C_i \subset C_j \Rightarrow C_j = C.$$

Or $C_p \neq C$, soit finalement

$$\forall i \neq p : C_p \subset C_i = C.$$

Donc deux cocons minimaux parmi les trois considérés sont égaux, et possèdent ainsi le même contraste interne. L'inégalité (5.13) est donc vérifiée pour toute permutation de (x, y, z).

Par ailleurs, la correspondance entre boules de E et cocons de G découle directement de la définition d'une ultramétrique. ■

Le contraste interne est donc un critère croissant associé à la hiérarchie dure. Une autre manière que 5.12 d'écrire $\delta(x, y)$ est en effet :

$$\delta(x, y) = \mathrm{I}_{\max}(\{x\} \vee \{y\})$$

où $\{x\}$ et $\{y\}$ sont les cocons triviaux associés à x et y et $\{x\} \vee \{y\}$ est leur sup dans la hiérarchie dure, soit en termes d'arbres le nœud auquel les branches de $\{x\}$ et $\{y\}$ convergent.

Le contraste interne joue donc le rôle du paramètre d'échelle dans cette hiérarchie.

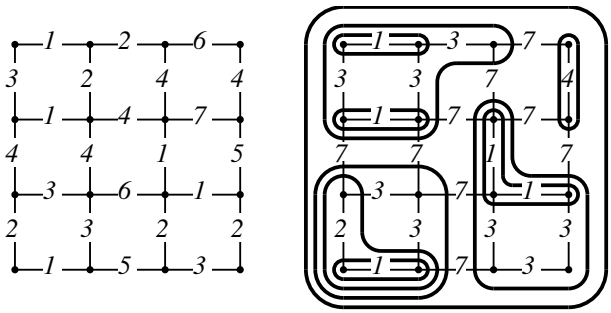


FIG. 5.6 – Exemple de cocons durs et d'ultramétrique associée. Gauche : graphe pondéré d'origine. Droite : ultramétrique et cocons associés. Remarques : i) nous n'avons pas représenté les singletons, qui sont tous des cocons. ii) l'ultramétrique n'a été reportée que sur les arêtes du graphe d'origine mais elle est définie pour tout couple de nœuds.

Robustesse de la hiérarchie dure

La hiérarchie des cocons durs est robuste dans le sens suivant :

Proposition 2 *Pour tout graphe $G = (X, U, d)$, la hiérarchie des cocons durs est invariante par toute transformation strictement croissante de la fonction de dissemblance.*

Preuve : Pour toute fonction réelle t strictement croissante, et $\forall Y \subset X$, de manière évidente

$$\max_{e \in \widetilde{Y}} d(e) < \min_{e \in Y \sim X \setminus Y} d(e) \quad \Leftrightarrow \quad \max_{e \in \widetilde{Y}} t \circ d\,(e) < \min_{e \in Y \sim X \setminus Y} t \circ d\,(e).$$

■

Avec cette invariance les hiérarchies de cocons durs vérifient les trois qualités requises en 1967 par Johnson pour les algorithmes de groupement de points s'appuyant sur une dissemblance. La table 5.1 les énonce telles qu'elles sont données par Zahn (1971).

Notons que cette propriété d'invariance s'étend à tous les couples de fonctions statistiques (f, g) pour lesquelles l'*ordre* qu'elle induisent sur les échantillons de réels est inchangé par transformation monotone. Il s'agit essentiellement des statistiques «robustes», i.e. des statistiques de rang tels que les quantiles. La propriété est par exemple fausse pour les moyennes.

1. Les données d'entrée doivent consister en un ensemble de points et une matrice de similarités seulement («Input data should consist solely of a point set and a matrix of similarities»).
2. Ce qu'accomplit la méthode de groupement doit pouvoir être décrit de manière claire, explicite et intuitive («The method should be such that a clear, explicit, and intuitive description of what the clustering accomplishes is possible»).
3. Le résultat doit être invariant par transformation monotone de la mesure de similarité («The method should be invariant under monotone transformations of similarity measure»).

TAB. 5.1 – Critères de qualité de Johnson pour les algorithmes de groupement par similarité.

Un cas particulier : composantes connexes des sections d'une image

Le principe d'invariance par changement de contraste est fondamental en morphologie mathématique. Plus particulièrement, il est à la base de la représentation par «ensembles de niveau», «coupes» ou «sections» d'une image en niveaux de gris (Matheron 1975). Voyons ici que cette représentation entretient un rapport étroit avec les hiérarchies dures.

Soit X un ensemble et $I : X \to \mathbb{R}$ une image en niveaux de gris de support X.

La section inférieure (resp. supérieure) de niveau λ de I est définie par

$$\begin{aligned} L^\lambda &= \{x \in X | I(x) \leq \lambda\} \\ (\text{resp.} \quad L_\lambda &= \{x \in X | I(x) \geq \lambda\}). \end{aligned}$$

Ces définitions sont illustrées par la figure 5.7[13].

Chacune de ces chaînes de sections emboîtées (indexées par λ) permet de reconstruire complètement l'image. En effet

$$I(x) = \inf\{\lambda | x \in L^\lambda\} = \sup\{\lambda | x \in L_\lambda\}.$$

Plaçons-nous dans le cas où X est un ensemble discrêt et dotons-le d'une topologie de voisinage sous la forme d'un graphe symétrique $G = (X, U)$. Notons alors C^λ (resp. C_λ) l'ensemble des composantes connexes de L^λ (resp. L_λ) au sens de G. Définissons finalement $C^* = \bigcup_{\lambda \in \mathbb{R}} C^\lambda$ et $C_* = \bigcup_{\lambda \in \mathbb{R}} C_\lambda$.

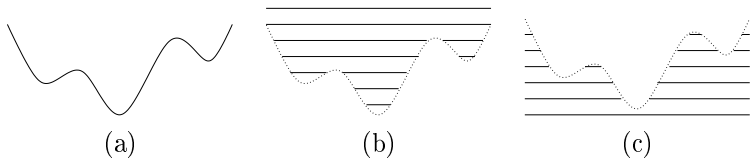


FIG. 5.7 – Ensembles de niveau ou sections d'une fonction. (a) Une fonction réelle I. (b) sections inférieures de I. (c) sections supérieures de I.

Les ensembles C^* et C_*, constitués des composantes connexes des sections de I, sont tous deux hiérarchisés. Ils ne forment cependant pas des hiérarchies au sens strict sur X, car ils ne contiennent pas tous les singletons de X. Les sections supérieures ne comportent comme singletons que les minima locaux ponctuels de I et celles inférieures ses maxima locaux ponctuels. Nous parlerons tout de même ici de hiérarchie, par abus de langage. La *structure* de ces représentations multi-échelles est de toute évidence invariante par changement de contraste, l'axe des «échelles» (le niveau de section) étant directement celui des niveaux de gris.

Notons que la topologie des sections de I est intimement liée à la ligne des partage des eaux de I. Plus précisément, la notion importante est l'échelle à partir de laquelle deux points se retrouvent connectés,

[13] Si l'on veut que $\forall \lambda \quad L^\lambda \cup L_\lambda = X$, il faut rendre une des deux inégalités stricte. Avec une définition symétrique, comme celle donnée ici, $L_\lambda(I) = L^{-\lambda}(-I)$.

c'est-à-dire l'ultramétrique associée à la hiérarchies des sections, qui représente le niveau à partir duquel deux bassins versants se mélangent.

Plusieurs auteurs se sont récemment (ré-)intéressés à ces représentations, en privilégiant toutefois les *frontières* des composantes connexes des ensembles de niveau, soit les *courbes de niveau* de l'image envisagée comme une surface topographique. Remarquons alors que du point de vue des frontières de leurs ensembles, C^λ et C_λ coïncident.

Caselles, Coll, et Morel (1998) abordent le problème en continu, dans l'espace des fonctions $\mathbb{R}^2 \to \mathbb{R}$ à variations bornées (BV). Ils proposent une représentation des fonctions BV appelée *carte topographique* et montrent d'importants résultats théoriques concernant la structure de ces cartes : régularité, simplicité, fermeture, etc. des courbes de niveau qui ne touchent pas le bord de l'image. Ils utilisent ces cartes topographiques pour traiter le problème de désocclusion. Au niveau d'une occlusion (à la frontière du domaine de l'image : trou ou bord), les courbes de niveaux s'interrompent brutalement et donnent lieu aux célèbres jonctions en T de Kanisza. Les auteurs proposent alors de compléter les parties manquantes d'une image par la surface de variations bornées qui donne lieu à la reconnection la plus lisse possible des courbes interrompues, i.e. par des portions de courbure minimale. L'algorithme qui dérive de la résolution de ce problème variationnel donne des résultats très impressionants (Caselles, Masnou, et Morel 2001).

Monasse et Guichard (2000) proposent de ne retenir d'une image que la structure des composantes connexes des ensembles de niveau (supérieures ou inférieures), indépendament de l'échelle à laquelle elles apparaissent, i.e. uniquement la hiérarchie C^* ou C_*, sans l'indice associé. Étant invariante par changement de contraste, on peut en effet argumenter qu'il s'agit de la structure «forte» de l'image. Ils remarquent alors que cette structure est totalement déterminée par la donnée des courbes de niveau et d'une *orientation* de ces courbes, qui permet de discriminer l'intérieur de l'ensemble de son extérieur. Les auteurs dégagent un algorithme rapide pour construire une représentation arborescente des courbes de niveau, suivant l'inclusion des portions connexes. L'orientation des courbes permet de savoir si une courbe est de niveau inférieur ou supérieur à sa courbe «mère» dans la hiérarchie. Ils montrent alors l'intérêt d'une telle structure pour des problèmes de filtrage, de segmentation ou de détection de changements et d'appariement d'images.

Examinons maintenant la relation entre cocons durs et éléments connexes des sections d'une image discrète en niveaux de gris.

Notons $G^o = (X, U^o)$ le graphe obtenu à partir de $G = (X, U)$ en ajoutant à U toutes les boucles : $U^o \triangleq U \cup \{(x,x) | x \in X\}$. On a alors

Proposition 3 *Soit $G = (X, U)$ un graphe symétrique et I une image en niveaux de gris sur X, i.e. une valuation réelle des nœuds de G. Les composantes connexes des sections supérieures de (G, I) sont les cocons durs de (G^o, d), où d est la fonction de dissemblance définie par*

$$\forall u \in U^o \quad d(u) = \max\{I(u^-), I(u^+)\}. \tag{5.14}$$

Pour tout λ, C^λ s'obtient alors comme coupe naturelle de niveau λ de la hiérarchie dure indicée par le contraste interne (changer I en $-I$ conduit aux ensembles inférieurs).

Preuve : $C \in C^\lambda$ si et seulement si il est connexe et vérifie

$$\begin{array}{rlcl} & \forall x \in C & I(x) & \leq & \lambda \\ \text{et} & \forall x \in \mathcal{V}(C) & I(x) & > & \lambda. \end{array}$$

(Rappelons que $\mathcal{V}(C) = \{e^+ | e \in \omega(C)\}$ désigne le voisinage de C).

Par ailleurs, pour la dissemblance d, le contraste interne dur d'un ensemble $Y \subset X$ est donné par

$$\mathrm{I}_{\max}(Y) = \max_{u \in \tilde{Y}} \left\{\max\{I(u^-), I(u^+)\}\right\} = \max_{y \in Y} I(y)$$

(cette expression n'est valable pour un singleton $Y = \{y\}$ qu'à condition que la boucle (y, y) existe).

Le contraste externe de Y vaut de son côté

$$\mathrm{E}_{\min}(Y) = \min_{u \in Y \sim X \setminus Y} \left\{ \max\{I(u^-), I(u^+)\} \right\}.$$

Soit alors $C \in C^\lambda$. D'une part

$$(\forall x \in C \quad I(x) \leq \lambda) \quad \Rightarrow \quad \mathrm{I}_{\max}(C) \leq \lambda.$$

D'autre part,

$$(\forall x \in \mathcal{V}(C) \quad I(x) > \lambda) \quad \Rightarrow \quad (\forall u \in Y \sim X \setminus Y \quad d(u) > \lambda) \quad \Rightarrow \quad \mathrm{E}_{\min}(C) > \lambda.$$

C est donc un cocon dur.
Réciproquement, si Y est un cocon dur,

$$\mathrm{I}_{\max}(Y) < \mathrm{E}_{\min}(Y) \quad \Rightarrow \quad (\forall x \in \mathcal{V}(Y) \quad I(x) > \mathrm{I}_{\max}(Y)).$$

et donc

$$\forall \lambda \in [\mathrm{I}_{\max}(Y), \mathrm{E}_{\min}(Y)[\qquad Y \in C^\lambda.$$

Les cocons durs de (G^o, d) et les éléments de C^* coïncident donc.
La dernière relation assure en outre que pour tout λ, les cocons maximaux pour l'inclusion et de contraste interne inférieur à λ (i.e. les éléments de la coupe naturelle à λ de la hiérarchie dure) sont les éléments de C^λ. ■

Notons la nécessité d'adjoindre les boucles (x, x) au graphe, afin de ne conserver comme cocons à un élément que les minima locaux de I et non pas tous les singletons.

La hiérarchie des éléments connexes des sections d'une image discrète mono-canal apparaît donc comme un cas particulier de hiérarchie de cocons durs. Notons toutefois que dans l'approche par cocons, la notion de voisinage peut être choisie de manière arbitraire. En utilisant des voisinages d'ordre supérieur à 2, on étend ainsi la notion de connexité spatiale stricte (4 ou 8 connexité) à une notion de proximité : des composantes de L^λ séparées de plusieurs pixels peuvent être considérées comme des unités potentielles. Cependant, toutes ces composantes ne peuvent pas être incluses sans briser la structure hiérarchique de l'ensemble. Le principe de contraste permet d'en sélectionner un sous-ensemble hiérarchisé. Plus la connexité est forte, plus l'ensemble retenu s'apauvrit. À l'extrême, si G est totalement connecté, i.e. si $G = (X, X^2)$, le seul ensemble «supérieur» restant, outre X lui-même, est l'ensemble - pris comme un tout - des minima absolus de l'image. Notons que comme une transformation monotone des niveaux de l'image I entraine une transformation monotone de la dissemblance $d(u) = \max\{I(u^-), I(u^+)\}$, on retrouve l'invariance par changement de contraste de la structure des ensembles de niveau.

La dissemblance d, qui conduit aux sections supérieures, n'est bien entendu pas la seule envisageable sur une image en niveaux de gris. Nous avons déja évoqué l'expression différentielle $d'(u) = |I(u^-) - I(u^+)|$ et donné une interprétation des cocons durs qu'elle produit (composantes lisses localement maximales). Comme toutes les hiérarchies dures, celles qui s'appuient sur d' sont invariantes par transformation monotone des *dissemblances*. La question qui vient alors naturellement est : à quelles transformations des *valeurs de l'image* sont-elles robustes ? Autrement dit, quelles sont les transformations sur les valeurs de I qui conduisent systématiquement (pour toute image I) à des transformations monotones de d', garantissant ainsi l'invariance structurelle de la hiérarchie dure ? On montre aisément que seules les transformations *affines* strictement croissantes garantissent cette invariance[14] pour tout f.

[14] C'est d'ailleurs également vrai en continu : si $f : \mathbb{R}^n \to \mathbb{R}$ est une «image» lisse (partout différentiable), notons $\mathcal{O}(f)$ l'*ordre* sur l'ensemble des couples $(point, direction)$ induit par les valeurs absolues des dérivées directionnelles de f. On peut alors montrer que les fonctions affines strictement croissantes sont les seules fonctions $g : \mathbb{R} \to \mathbb{R}$ qui confondent $\mathcal{O}(f)$ et $\mathcal{O}(gof)$.

En conclusion, les cocons durs qui sont directement associés aux valeurs de l'image I - via la dissemblance d - donnent ses sections, et sont invariants par changement de contraste sur I ; ceux qui sont associés aux variations de I - via d' - fournissent ses composantes lisses, qui ne sont invariantes que par transformation affine des valeurs de I.

Notons enfin à titre de curiosité que les zones plates d'une image $I : X \to \mathbb{R}$ pour la topologie $G = (X, U)$ s'obtiennent également en tant que cocons durs de (G, d''), où d'' est la dissemblance :

$$d''(u) = \delta(I(u^+) - I(u^-))$$

avec δ le delta de Kronecker.

5.5.3 Construction de la hiérarchie dure

Voyons maintenant comment construire les hierarchies de cocons durs (et complets). L'algorithme proposé découle de la proposition suivante :

Proposition 4 *Soit $G = (X, U, d)$ un graphe attribué par une dissemblance d. Soit $Y \subset X$ un ensemble connexe de noeuds, $Y \neq X$. Soit C le plus petit cocon contenant Y. Pour tout $v \in \mathcal{V}(Y)$, soit*

$$M(v) \quad \triangleq \quad \max\{d(e), e^- \in Y, e^+ = v\}. \tag{5.15}$$

Alors le voisin v^* de Y qui minimise M est inclus dans C.

Preuve par l'absurde : Remarquons que comme $Y \neq X$, C existe toujours. Supposons que $v^* \notin C$. Dans ce cas, il existe $v \in \mathcal{V}(Y), v \neq v^*$ tel que $v \in C$. Donc le contraste interne de C vérifie $\mathrm{I}_{\max}(C) \geqslant M(v)$. Comme par ailleurs $v^* \in \mathcal{V}(Y)$ et $v^* \notin C$, on a $v^* \in \mathcal{V}(C)$. Donc le contraste externe de C vérifie $\mathrm{E}_{\min}(C) \leqslant M(v^*)$. Par définition de v^* on a enfin $M(v^*) \leqslant M(v)$, ce qui conduit à l'inégalité $\mathrm{E}_{\min}(C) \leqslant M(v^*) \leqslant M(v) \leqslant \mathrm{I}_{\max}(C)$ qui contredit le fait que C est un cocon strict pour *max* et *min*. ■

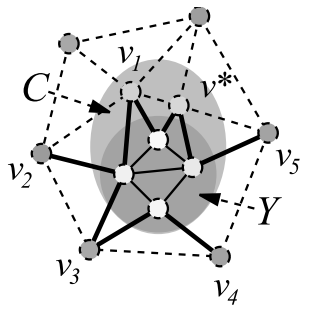


FIG. 5.8 – Un pas de croissance d'un cocon dur (max-min). Les voisins v_i sont triés selon le *max* des distances pour les atteindre depuis l'ensemble courant Y. Le voisin v^* qui minimise cette valeur est dans le cocon C contenant Y.

Chaque noeud isolé étant lui-même un cocon (aux boucles près...), on voit que l'on peut construire l'ensemble des cocons contenant un noeud donné par un processus itératif d'agglomération du voisin qui minimise le critère M. L'algorithme qui en découle est une variante de type croissance de région d'un algorithme de **complete linkage** (lien complet).

Le complete linkage appartient à une famille d'algorithmes de groupement hiérarchique dont la forme générale a été proposée par Lance et Williams (1967) (voir aussi la synthèse de Simon (1985, p.130)). Ces algorithmes procèdent par réunion itérée de paires de groupes en suivant l'ordre donné par une fonction de priorité calculée à partir des dissemblances. Chaque élément de X est initialement placé dans un groupe

différent et les priorités de fusion initiales sont directement données par la fonction de dissemblance d. À chaque création d'un nouveau groupe $A \cup B$ à partir de deux anciens groupes A et B, la priorité de fusion de $A \cup B$ avec un autre groupe est calculée par une fonction qui ne dépend que des priorités et des cardinaux des ensembles à l'étape précédente.

Avec nos notations, ce calcul correspond à affecter une priorité de fusion de $A \cup B$ avec un autre groupe C en fonction de la valeur d'une statistique f donnée sur les connexions $A \cup B \sim C$ entre $A \cup B$ et C. Notons que le groupement est traditionnellement envisagé sur des graphes complets contrairement au cadre des cocons, dans lequel le graphe de voisinage peut être quelconque.

Si l'on choisit $f = min$, on obtient l'algorithme de **single linkage** (lien simple). Exprimé au niveau d'un seul groupe en croissance, le critère d'agglomération du minimum est donné par

$$m(v) \quad \triangleq \quad \min\{d(e), e^- \in Y, e^+ = v\}. \tag{5.16}$$

Il réunit donc itérativement les plus proches voisins. Ce critère conduit à un *arbre recouvrant de poids minimal* (Minimum Spanning Tree, MST) supporté par le graphe, i.e. à un arbre $T = (X, V), V \subset U$ recouvrant $G = (X, U)$ (ayant les mêmes composantes connnexes que G) et qui minimise

$$\sum_{v \in V} d(v)$$

(notons que cet arbre n'est nécessairement unique que si toutes les dissemblances sont différentes).

L'algorithme de contruction d'un MST par agglomération itérative a été initalement trouvé par Prim (1957). Un autre algorithme classique, qui procède par fusions en parallèle, est dû à Kruskal (1956). L'ultramétrique associée au single linkage a été appelée ultramétrique sous-dominante par Roux en 1968. Si la dissemblance est une distance, cette ultramétrique correspond au sup des ultramétriques bornées par la distance (deux (ultra)métriques δ et δ' sont ordonnées pour $\leq$ ssi $\forall x, y \quad \delta(x, y) \leq \delta'(x, y)$), voir (Simon 1985, p.129). Depuis les travaux de Zahn (1971), le MST a souvent été utilisé en clustering, puis en segmentation, et particulièrement récemment dans les approches multi-échelles de la segmentation morphologique (Meyer 1999a; Meyer 1999b; Zanoguera 2001).

À l'opposé, le **complete linkage** correspond à choisir $f = max$. Au contraire du single linkage, ce critère est très précautionneux : il évalue la proximité de deux groupes à la distance de leurs éléments les plus lointains (diamètre de l'union si le graphe est complet).

Entre le single et le complete linkage on trouve l'**average linkage** (lien moyen), qui utilise $f = moy$.

En résumé, nous venons de montrer qu'au cours d'une croissance de type complete linkage à partir d'un noeud isolé, l'ensemble aggloméré passe par certaines étapes où il est un cocon dur, entrecoupées de phases intermédiaires durant lesquelles la propriété de cocon n'est pas vérifiée. Notons que comme les cocons complets et complets-complets sont des cas particuliers de cocons durs, le même algorithme permet de les trouver. La détermination des cocons durs nécessite donc de tester le contraste de l'ensemble à chaque pas de l'agglomération. On doit pour cela avoir accès aux valeurs du contraste interne et externe. Il faut donc soit les recalculer à chaque pas, soit les tenir à jour dynamiquement.

• Le **contraste interne des cocons durs** ne pose aucun problème car il peut être mis à jour directement grace à la valeur $M(v^*)$ de chaque nouveau noeud v^* ajouté. En effet, on a immédiatement :

$$\mathrm{I}_{\max}(Y \cup \{v^*\}) = \max\{\mathrm{I}_{\max}(Y), M(v^*)\}. \tag{5.17}$$

En revanche, pour la mise à jour du **contraste interne complet**, on doit balayer l'ensemble des connexions du nouveau venu vers l'ensemble courant, par

$$\mathrm{I}^K_{\max}(Y \cup \{v^*\}) = \max\left\{\mathrm{I}_{\max}(Y), \max\{d(y, v^*), y \in Y\}\right\}. \tag{5.18}$$

• De son côté, le calcul du **contraste externe** nécessite à première vue de parcourir à chaque pas le cocycle de l'ensemble courant. Une solution plus efficace consiste à tenir à jour une deuxième queue de priorité sur les voisins de l'ensemble en cours de croissance, triée selon le critère 5.16 du minimum. On dispose alors directement du contraste externe en examinant la valeur m du premier élément de la queue.

Le critère de contraste dur emploie donc une aggrégation de type complete linkage mais teste la validité d'une étape en la comparant à un critère de type single linkage. La notion de cocon dur repose donc fondamentalement sur la coopération entre les deux types de critères.

Un algorithme de calcul de la chaîne des cocons emboîtés sur un nœud donné est décrit en annexe B, ainsi qu'un algorithme de calcul de l'ultramétrique associée à la hiérarchie dure.

5.5.4 Résultats expérimentaux

Pour la segmentation d'images, nous avons envisagé des croissances de cocons à partir de sur-segmentations obtenues par un algorithme de ligne de partage des eaux sur le module d'un gradient de l'image (cf. annexe A). Le graphe considéré est le graphe d'adjacence des régions de cette sur-segmentation.

Mesures de dissemblances

Nous avons expérimenté différentes mesures de dissemblance entre régions élémentaires.

La première idée qui vient à l'esprit est de faire reposer les dissemblances sur le gradient présent sur les frontières entre ensembles. Ce type d'attribution des arêtes du graphe pose cependant des problèmes techniques. Pour que la mesure soit fiable, il faut que le gradient sur la frontière soit mesuré dans la direction normale. En effet, si l'on ne s'appuie que sur le module du gradient, un élément d'une frontière peu marquée se connectant à une ligne de fort contraste obtient une dissemblance biaisée par la présence de la ligne de fort contraste. Le biais n'est pas du tout négligeable sur les frontière d'une LPE, qui sont généralement courtes. En outre, l'algorithme de LPE que nous avons considéré produit des frontières discrètes inter-pixellaires, ayant seulement deux orientations possibles $N - S$ ou $E - O$. L'attribution de mesures pertinentes de gradient directionnel nécessiterait une modélisation plus fine de la géométrie de la LPE (du type des modélisation fines des courbes de niveau proposées par Caselles et al. (1998) ou Monasse et Guichard (2000)), ce que nous n'avons pas cherché à faire.

Nous avons préféré nous orienter vers des dissemblances s'appuyant sur des tests statistiques entre les échantillons radiométriques portés par les régions (qui s'apparentent, tel que nous le voyons a posteriori, à une variante statistique de l'idée de «gradient mosaïque» de Beucher (1990)).

Notons X_1 et X_2 deux échantillons de respectivement n_1 et n_2 observations d'un espace $\mathcal{X}$, éventuellement nD (e.g. $\mathbb{R}^n$).

• **Statistique de Kolmogorov-Smirnov (KS)**.

$\mathcal{X}$ est ici supposé discret et doté d'un ordre total. En imagerie numérique, ce cas correspond typiquement à l'échelle $\langle 0, 255\langle$ des niveaux de gris d'une image en noir et blanc. Pour $i \in \{1, 2\}$, notons $F_i : \mathcal{X} \to \mathbb{R}$ la fréquence des observations (leur histogramme normalisé) et R_i la fonction de répartition associée à X_i, donnée par :

$$R_i(j) = \sum_{k \leq j} F_i(k).$$

Le test non paramétrique de Kolmogorov-Smirnov s'appuie alors sur la statistique

$$d_{KS}(X_1, X_2) = \sqrt{\frac{n_1 \times n_2}{n_1 + n_2}} \max_{i \in \mathcal{X}} |R_1(i) - R_2(i)|.$$

Supposons que les échantillons X_1 et X_2 aient été produits par tirages indépendants selon deux lois P_1 et P_2 (i.i.d.). On montre alors que si $P_1 = P_2 = P$ alors $d_{KS}(X_1, X_2)$ suit une loi donnée, notons-la

$\mathcal{KS}$, qui est *indépendante* de P (Smirnov 1948). Connaissant la loi $\mathcal{KS}$, on peut donc tester, à un niveau de confiance donné, l'hypothèse $P_1 \neq P_2$ contre l'hypothèse nulle $P_1 = P_2$.

En outre, $\mathcal{KS}$ converge vite vers une loi indépendante de n_1 et n_2. Pour des tailles d'échantillons raisonnables (> 40), le seuil de décision sur $d_{KS}(X_1, X_2)$ ne dépend pratiquement que du niveau de confiance (ce seuil vaut par exemple $1,36$ pour une confiance de 95%, $1,63$ pour 99% et $1,95$ pour $99,9\%$). Les seuils pour les petits échantillons sont tabulés dans (Siegel 1956).

Cette distance possède cependant certains inconvénients.

Avant tout, elle utilise des fonctions de répartition, qui s'appuient sur un ordre total sur $\mathcal{X}$. Elle est donc difficilement extensible à des espaces sur lesquels il n'existe pas d'ordre naturel, comme les espaces colorimétriques. Cette limitation de fond est difficile à palier. Une extension du test de KS à des variables qualitives est proposée par Asseraf (1998). Elle n'est toutefois pas utilisable sur des espaces colorimétriques car elle s'appuie sur une probabilisation explicite de l'ensemble des parties de l'espace des réalisations !

Une idée simple pour étendre le test de KS à des données multi-dimensionnelles est alors d'effectuer le test sur un certain nombre de projections mono-dimensionnelles de la distribution (loi marginales). Si l'on note $D = \{D_i\}_{i=1\ldots p}$ l'ensemble des droites de projection considérées, on définit une nouvelle mesure de dissemblance par :

$$d_{KS}(X_1, X_2, D) = \max_{i=1\ldots p} d_{KS}(\mathcal{P}_{D_i}(X_1), \mathcal{P}_{D_i}(X_2)). \tag{5.19}$$

où $\mathcal{P}_{D_i}(.)$ désigne l'opérateur de projection sur la droite D_i[15]. Dans ce qui suit, nous considérons des images couleur, à trois canaux, et les projections seront tout simplement celles sur chacun des axes rouge, vert et bleu[16].

Le deuxième inconvénient est que ce test ne prend pas en compte la métrique pouvant exister sur l'espace des réalisations. En effet, si l'on a affaire à deux échantillons dont les événements de probabilité non nulle ont des supports disjoints, la mesure de KS est constante, et ce quelle que soit la distance entre les différents événements. Intuitivement, d_{KS} ne dépend que de «l'entrelacement» des masses de probabilité. Prenons par exemple le cas extrême de deux échantillons $X_1 = [x_1]$ et $X_2 = [x_2]$ de $\langle 0, 255\langle$ constitués chacun d'une seule observation. Alors, $\forall x_1 \neq x_2$, $d_{KS}(X_1, X_2) = 1/\sqrt{2}$: la statistique ne dépend pas de la distance entre les observations. Ce type de comportement s'observe également pour des échantillons de grande taille.

Évidemment, ce comportement est parfaitement naturel du point de vue de l'hypothèse nulle que se propose d'éprouver le test de KS, à savoir l'égalité des lois. Toutefois, dans notre cadre, la statistique de KS n'est pas employée pour décider avec une confiance suffisante si oui ou non deux régions présentent des contenus *identiques* (et par exemple les fusionner si le test est positif) mais pour évaluer un degré de ressemblance. Plus particulièrement, dans le cadre des cocons, les mesures de dissemblance ont uniquement pour rôle d'*ordonner* les couples de régions selon la «proximité» de leurs contenus radiométriques. Pour l'exemple des deux échantillons à une seule observation donné ci-dessus, il est alors souhaitable que la dissemblance croisse avec la distance entre les observations.

Une manière d'introduire la métrique existant sur l'espace radiométrique dans le test de KS consiste alors à considérer que l'image I est bruitée, i.e. que $I = J + \varepsilon$, où ε est un bruit gaussien i.i.d., d'écart-type σ, centré sur une image «vraie», J. En chaque point p de I, la probabilité de la valeur «vraie» $J(p)$ de la radiométrie est alors distribuée selon une gaussienne centrée sur $I(p)$ et d'écart type σ. Sous ces hypothèses, pour effectuer un test sur J, il faut convoluer par une gaussienne les histogrammes

[15] On peut se demander si en multipliant les tests sur des lois marginales, on ne converge pas vers un test suffisant : si les lois diffèrent, il faut bien que cela se ressente quelque part (on peut voir le problème comme une inversion d'une transformée de Radon)...

[16] On pourrait aussi utiliser un axe de luminance et deux axes de chrominance orthogonaux, ou effectuer le test après changement de mode de représentation colorimétrique. Différents essais dans ce sens - en particulier utilisant les espaces colorimétriques «perceptuellement homogènes» *Lab* ou *Luv* (cf. e.g. Luong (1990)) - n'ont cependant pas permis de tirer de conclusion nette. Les résultats sont différents mais aucun n'est franchement meilleur. Nous préférons finalement adopter un point de vue «physique» plus que «psychophysique», en utilisant les mesures RVB brutes, directement issues des capteurs.

obtenus sur I avant d'effectuer la mesure de KS. En procédant ainsi pour l'exemple des deux échantillons ponctuels $[x_1]$ et $[x_2]$, d_{KS} diminue quand x_1 et x_2 se rapprochent. On modélise ainsi d'une part un bruit d'observation et d'autre part l'existence d'une distance sur l'axe des radiométries. Le deuxième point tient au fait que les probabilités gaussiennes sont strictement décroissantes avec l'écart à la moyenne, et ne s'annulent jamais. Pour la convolution à proprement parler, on ne peut donc pas tronquer les queues des gaussiennes. La solution la plus efficace est alors d'utiliser un filtrage auto-régressif qui nécessite seulement deux balayages d'un histogramme (Deriche 1993).

Une autre approche consisterait à remplacer l'opérateur max de d_{KS} par une moyenne. On perd cependant les bonnes propriétés théoriques du critère. Nous avons retenu la technique de convolution, avec $\sigma = 1$.

• **T-test de Fisher** et sa généralisation à $\mathbb{R}^n$ due à **Hotelling** (citée dans Zhu et Yuille (1996)).

Le test de Fisher-Hotelling (FH) fait une hypothèse gaussienne sur les échantillons observés et teste l'égalité de leur moyenne. Soit X_1 et X_2 deux échantillons de $\mathbb{R}^n$. Pour $i \in \{1, 2\}$., notons $\overline{X_i}$ la moyenne empirique de X_i et Σ_{X_i} sa matrice de variance/covariance.

Le test de FH s'appuie sur la statistique

$$d_{FH}(X_1, X_2) = \frac{n_1 \times n_2}{n_1 + n_2}(\overline{X_1} - \overline{X_2})^t \left(\frac{n_1\Sigma_{X_1} + n_2\Sigma_{X_2}}{n_1 + n_2 - 2}\right)^{-1} (\overline{X_1} - \overline{X_2})$$

qui suit une loi connue quand X_1 et X_2 ont été produits par deux gaussiennes de même moyenne, voir e.g. (Johnson 1982). Encore une fois, les seuils de décision associés à un niveau de confiance donné convergent très rapidement. Nous utilisons donc le résultat «brut» de la statistique d_{FH} pour attribuer les dissemblances.

Cette mesure a le bon goût d'être invariante par rotation du repère colorimétrique et est réputée robuste, même pour les petits échantillons. En outre, elle ne nécessite que le stockage des moments d'ordre inférieur à 2, contrairement à KS qui emploie des histogrammes.

• **«Distance» de Kullback-Leibler** (**KL**, encore appelée entropie croisée ou entropie relative)

Si $\mathcal{X}$ est discret, l'entropie de X_1 relativement à X_2 s'écrit

$$H(X_1|X_2) = \sum_{i \in \mathcal{X}} F_1(i) \log \frac{F_1(i)}{F_2(i)}$$

où les F_i sont les fréquences des observations (Cover et Thomas 1991).

Cette quantité issue de la théorie de l'information traduit l'inefficacité du code de Shanon-Fano s'appuyant sur la loi F_2 quand les échantillons apparaissent selon F_1. Si l'on fait une telle erreur - utiliser le code optimal associé à F_2 au lieu de F_1 - on aura besoin en moyenne de $H(X_1|X_2)$ bits supplémentaires pour coder une réalisation. Cette notion apparaît par exemple dans la solution d'un problème de segmentation en k classes formulé en terme de maximum de vraisemblance (Puzicha et Buhmann 1999).

L'entropie relative n'est cependant pas une distance. Elle est positive, ou nulle quand les échantillons sont égaux, mais n'est pas symétrique et ne vérifie pas l'inégalité triangulaire. Nous définissons donc une mesure de dissemblance par symétrisation :

$$d_{KL}(X_1, X_2) = \min\{H(X_1|X_2), H(X_2|X_1)\}. \tag{5.20}$$

Pour les images à plusieurs canaux, on utilisera ici aussi la technique de projection 5.19.

Contrairement à la distance KS, cette mesure est applicable à n'importe quel type de variable aléatoire à valeurs dans un espace discret. En contrepartie, pour les variables sur des espaces structurés (e.g. un intervalle de $\mathbb{N}$), elle ne prend pas en compte la structure de l'espace et en particulier ni l'ordre, ni la distance qui peut exister sur cet espace. Ce point peut se résoudre par la technique de convolution introduite pour la distance KS.

d_{KL} pose en outre un problème technique quand les fréquences des échantillons ont des supports différents, i.e. quand il existe i tel que $F_1(i) \neq 0$ et $F_2(i) = 0$, ou inversement. En effet, si $F_2(i) = 0$ le

code de Shanon associé à la loi F_2 propose une longueur de code $-\log F_2(i)$ qui est infinie pour l'événement i, et si $F_1(i) \neq 0$ alors $d_{KL}(X_1, X_2) = +\infty$. Ce n'est par contre pas le cas du code de Huffman, qui associe des mots-code finis à tous les événements. Nous avons donc utilisé dans (Guigues 2000b) une variante de la distance de KL s'appuyant sur un code de Huffman :

$$d_{KLH}(X_1|X_2) = \sum_{i \in \mathcal{X}} F_1(i)(\mathcal{L}_2(i) - \mathcal{L}_1(i))$$

où $\mathcal{L}_k(i)$ est la longueur du mot-code de Huffman associé à l'évèment i pour l'échantillon k.

Différents tests avec cette mesure de dissemblance ont cependant montré qu'elle se comportait nettement moins bien que les deux autres mesures envisagées. Nous n'en reproduirons pas ici, voir (Guigues 2000b).

Exemples de cocons durs

Les figures 5.9 et 5.10 présentent des résultats obtenus respectivement avec les dissemblances d_{FH} et d_{KS}.

La figure 5.9 montre que le critère de cocon parvient, à partir du singleton 5.9(e), à discriminer deux ensembles emboîtés pertinents : un pan de toît ou deux pans de toît adjacents de même exposition. Les étapes intermédiaires ne sont pas des cocons. Une entité globale comme la maison viole bien entendu le critère : elle possède d'importantes variations internes d'intensité, plus fortes que certaines variations situées sur sa frontière.

La figure 5.10 présente une chaîne de cocons durs emboîtés pour la mesure de KS. On constate ici que comme la frontière de l'ombre n'est pas franchement marquée, de nombreux cocons intermédiaires sont présents, qui correspondent à différentes hypothèses possibles pour la position de la limite. L'image du bas de la figure 5.10 montre les plus grands cocons qui ne sont pas le graphe entier. Elle illustre la «dureté» du critère *max-min*. On ne trouve en effet pas de cocons durs de grande taille.

Partitionnement par coupes itérées de hiérarchies dures

Comme nous l'avons expliqué au chapitre 4, les coupes naturelles d'une hiérarchie indicée produisent des partitions. Pour obtenir des segmentations au sens classique du terme, nous avons donc considéré des coupes de la hiérarchie dure selon le contraste interne des cocons, qui est le critère croissant naturellement associé à cette hiérarchie.

L'annexe B décrit un algorithme qui permet de calculer la partition maximale en cocons s-homogènes, c'est-à-dire la coupe à s de la hiérarchie dure indicée par le contraste interne, sans toutefois construire l'ensemble de la hiérarchie.

Comme vient de le montrer la figure 5.10, du fait de l'extrême sensibilité du contraste dur, on trouve très peu de grands cocons durs. Une seule coupe d'une hiérarchie dure conduit donc généralement à une importante sur-segmentation par rapport au niveau visé. Nous avons donc proposé dans (Guigues et al. 2001a) et (Guigues et al. 2003c) d'itérer construction de hiérarchie dure et s-coupe, chaque nouvelle hiérarchie prenant pour base la coupe effectuée à l'étape précédente et utilisant des dissemblances réévaluées.

L'algorithme consiste précisément en :

Algorithme 1 **Segmentation par coupe itérée d'une hiérarchie dure**

Paramètre : Seuil s de contraste interne.

1. Réaliser une sur-segmentation S (image de labels) de l'image I par LPE.
2. Construire le graphe d'adjacence $G = (X, U)$ de régions de S.
3. Itérer :

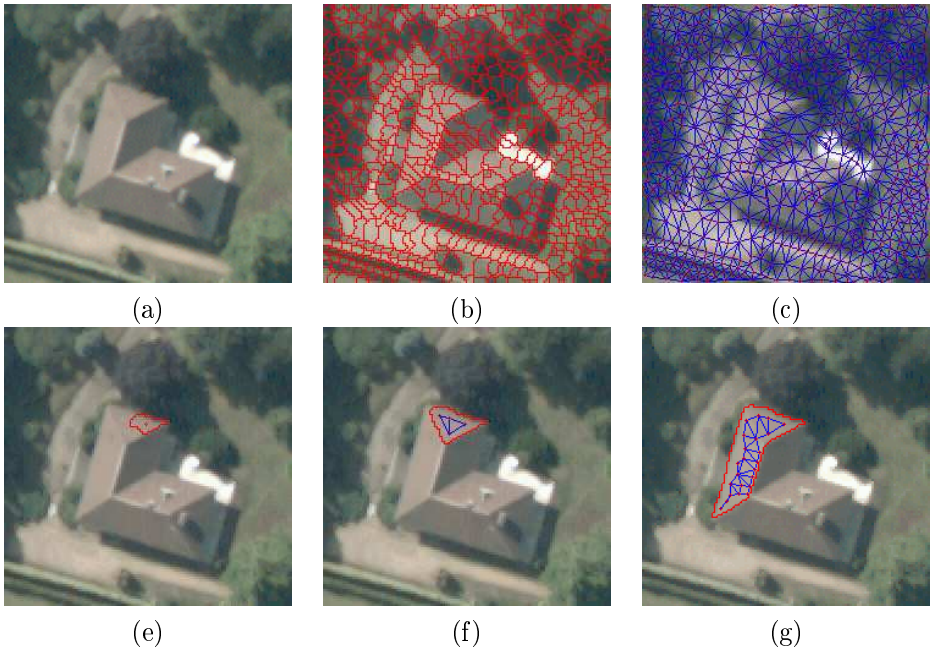


FIG. 5.9 – Une chaîne de cocons durs emboîtés. (a) Image aérienne. (b) Sur-segmentation par ligne de partage des eaux. (c) Graphe d'adjacence de régions correspondant. (e) Un cocon trivial à un élément. (f)(g) Les deux cocons de la branche de la hiérarchie issue de (e), sommet mis à part (dissemblance d_{FH}).

(a) Calculer les dissemblances d sur U.

(b) Calculer les cocons durs s-homogènes maximaux pour le critère de contraste interne. Si les cocons sont tous des singletons aller en 4.

(c) Contracter les cocons trouvés et mettre à jour la segmentation S (fusionner les labels).

4. Renvoyer S.

Fin de l'algorithme

Remarque : itérer le partitionnement selon les plus grands cocons qui ne sont pas le graphe entier converge en général vers une bipartition de l'image. En effet, sauf cas particulier, il existe toujours au moins un cocon non trivial : l'ensemble des nœuds connectés par les arêtes qui portent la plus faible dissemblance. En pratique, cette bipartition s'avère peu pertinente.

Il est important de voir que l'algorithme 5.5.4 n'est pas un simple algorithme de segmentation par prédicat, contrôlé par un seuil sur une mesure d'hétérogénéité. En effet, le critère de contraste sélectionne les étapes auxquelles l'algorithme peut s'arêter et, comme nous l'avons vu, cette sélection est relativement drastique (sur l'exemple de la maison de la figure 5.9, seules 4 positions de coupe sont possibles : au niveau des atomes, du premier pan de toît, des deux pans fusionnés ou de l'image entière). Le critère de cocon - qui est lui sans paramètre - réalise donc un filtrage sur les régions qu'il est possible d'obtenir dans une segmentation.

Les figures 5.11 et 5.12 détaillent sur un exemple les itérations successives de l'algorithme 5.5.4. Cet exemple utilise d_{KS} et un seuil $s = 1,5$.

La figure 5.13 présente quelques résultats de segmentation. Leur obtention nécessite de 5 à 10 secondes de calcul pour des images 512×512 avec un ordinateur cadencé à 500 MHz. Dans les exemples, le nombre de nœuds dans les graphes initiaux varie de 5000 à 20000 et autour de 20 itérations sont requises. Remarquons que le nombre d'itérations est indépendant de la taille de l'image, mais dépend seulement de son contenu (voir l'analyse de l'annexe B).

FIG. 5.10 – Exemples de cocons durs. En haut : une chaîne de cocons emboîtés. Les nœuds du cocon sont en rouge, ses arêtes en bleu et ses voisins en vert. En bas : plus grands cocons sous le graphe entier (dissemblance d_{KS}).

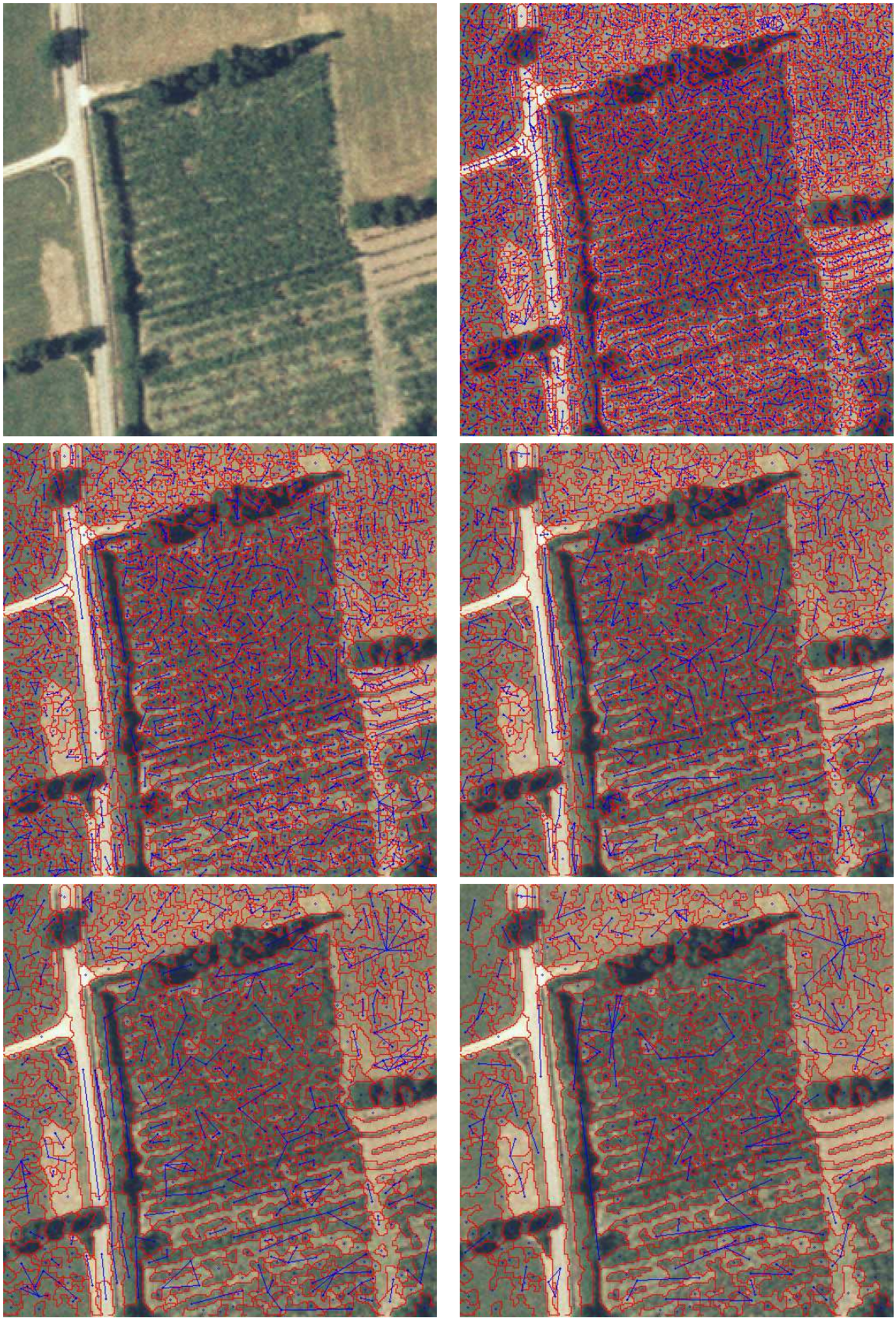


FIG. 5.11 – Cocons homogènes itérés (1/2). La dissemblance est d_{KS} et $s = 1,5$.

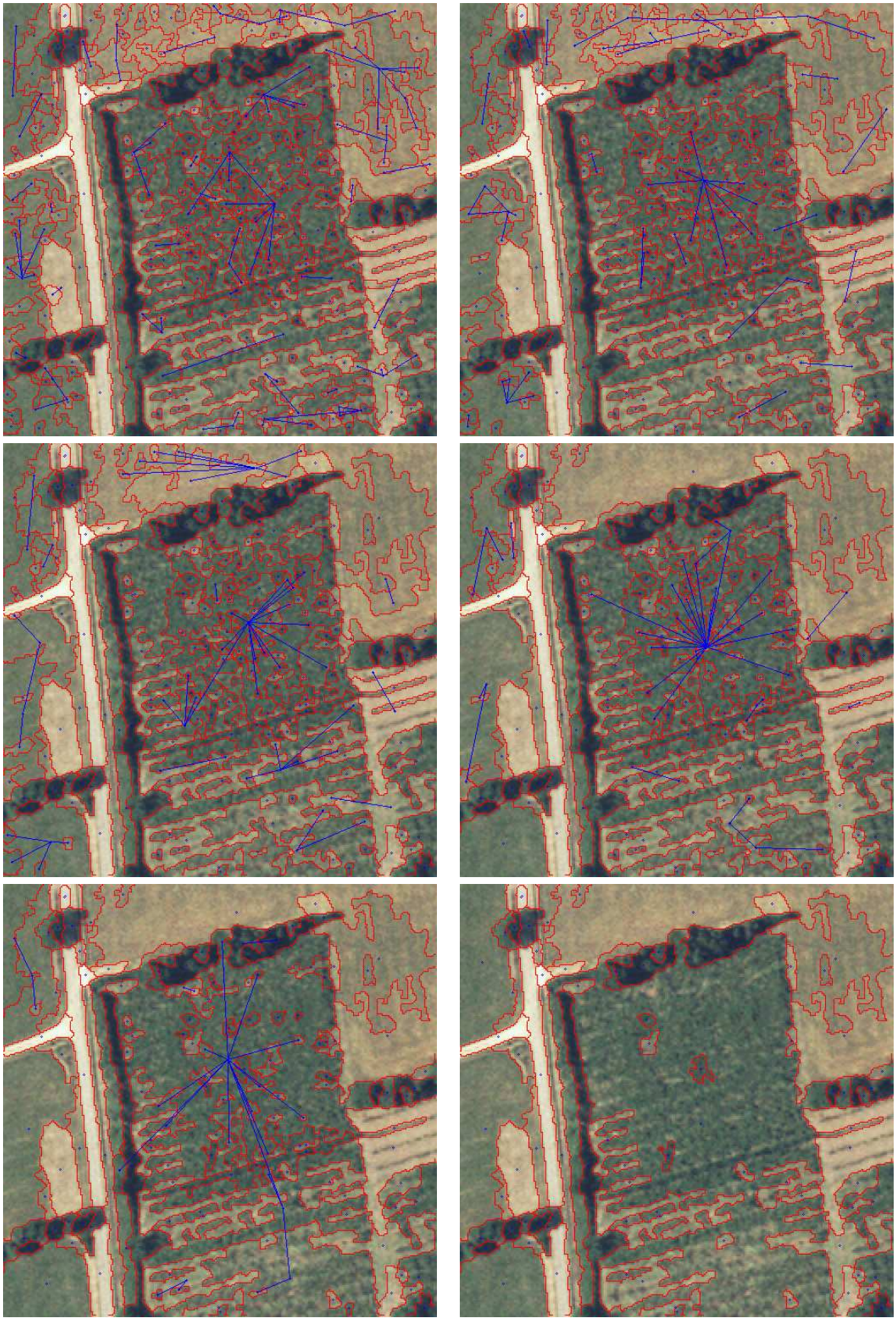
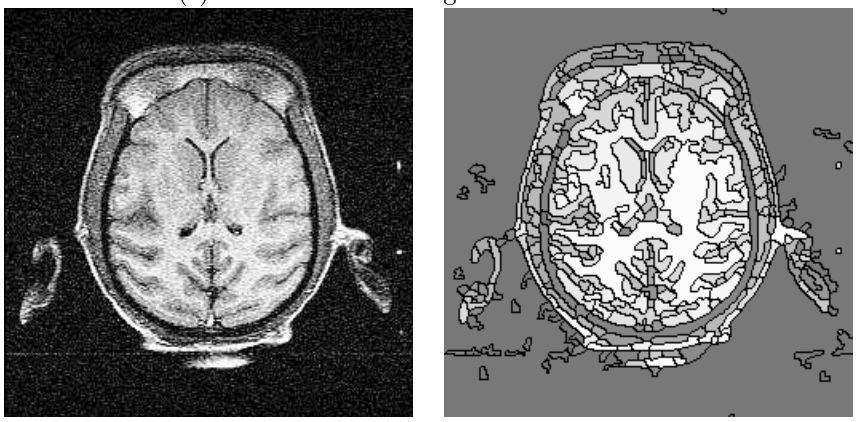
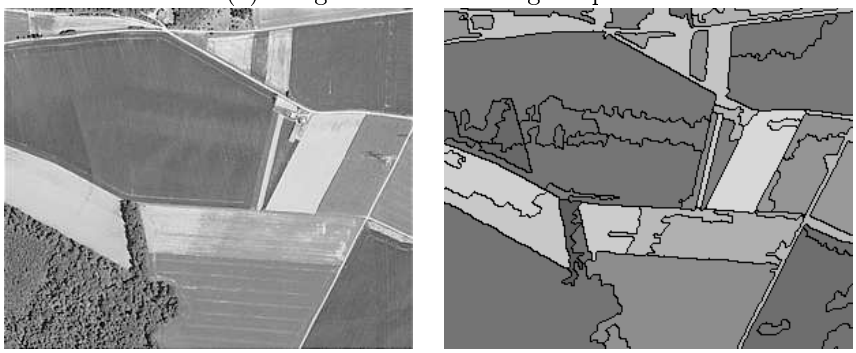


FIG. 5.12 – Cocons homogènes itérés (2/2).

(a) Trois niveaux de segmentation de Lena.



(b) Image à résonance magnétique.



(c) Image aérienne.

FIG. 5.13 – Exemples de segmentation par cocons maximaux itérés.

5.6 Hiérarchies de cocons équilibrés

5.6.1 Familles de cocons non hiérarchisés

Considérons maintenant des statistiques f et g plus robustes que max et min.

Les quantiles représentent l'assouplissement naturel de ces statistiques extrêmes. Pour conserver une certaine symétrie dans la définition des contrastes internes et externes, il est naturel de considérer $f = [1-q]$ et $g = [q]$, avec q variant entre 0 et 1/2 (rappelons que $[q]$ désigne la statistique de q-quantile, $q \in [0,1]$). $q > 1/2$ n'a pas d'intérêt car il définit une notion trop souple de contraste (vote sous-majoritaire). Quand $q = 0$ on retrouve $f = max$ et $g = min$. À l'opposé, quand $q = 1/2$, f et g sont deux médianes.

Un deuxième choix naturel est d'utiliser deux moyennes, i.e. $f = g = moy$.

Malheureusement, **en général les ensembles $\mathcal{C}_{moy}^{moy}$ et $\mathcal{C}_{[1-q]}^{[q]}$ ne sont pas hiérarchisés**.

La figure 5.14(a) l'illustre pour la moyenne. Elle présente un graphe qui contient deux *moy-moy*-cocons stricts $C_1 = \{A, B\}$ et $C_2 = \{B, C\}$ qui ne sont pas hiérarchisés. En effet,

$$\begin{aligned} & \mathrm{I}_{moy}(C_1) = 0.1 \quad < \quad \mathrm{E}_{moy}(C_1) = (0.2 + 0.9)/2 \\ \text{et} \quad & \mathrm{I}_{moy}(C_2) = 0.2 \quad < \quad \mathrm{E}_{moy}(C_2) = (0.1 + 0.9)/2. \end{aligned}$$

Pour ce qui est des quantiles, la figure 5.14(b) illustre comment pour tout $q \in]0, 1/2]$ on peut construire un graphe contenant deux $[1-q]$-$[q]$-cocons non hiérarchisés. $C_1 = \{A, B\}$ est contrasté pour tout quantile et $C_2 = \{B, C\}$ peut être rendu contrasté pour tout q en ajoutant un nombre suffisant d'arêtes incidentes à C portant toutes une dissimilarité plus forte que la dissimilarité de l'arête (B, C).

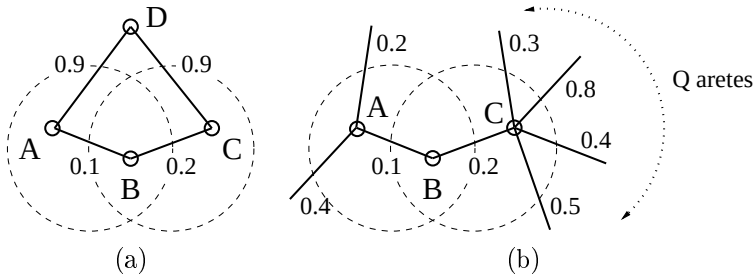


FIG. 5.14 – Non hiérarchisation a priori des familles de cocons moyens et s'appuyant sur des quantiles. Dans le graphe attribué (a), les groupes $\{A, B\}$ et $\{B, C\}$ sont deux *moy-moy*-cocons non hiérarchisés. (b) Montre que pour tout $q \in]0, 1/2]$, on peut construire un graphe qui contient deux $[1-q]$-$[q]$-cocons non hiérarchisés, en augmentant suffisamment le degré d'un des nœuds.

5.6.2 f-Linkage et sous-ensembles hiérarchisés de f-cocons

Malgré ce résultat négatif, intéressons-nous de plus près aux familles de cocons tels que $f = g$. Appelons de tels cocons des cocons **équilibrés**.

Bien entendu, choisir $f = g$ interdit que f et g vérifient simultanément $f(\emptyset) = \Delta_0$ et $g(\emptyset) = \Delta_1$.

Avant de remédier à ce problème, élargissons la définition de cocon, en nous passant de la notion de dissemblance et en définissant f et g comme des *fonction d'ensemble* sur les arêtes de $G = (X, U)$, c'est-à-dire comme des fonctions sur les parties de U et à valeur dans $\overline{\Delta}$. Évidemment, les statistiques de dissemblances - agrégeant des valeurs élémentaires portées par les arêtes - sont des cas particuliers de fonctions d'ensemble.

Si f est une fonction d'ensemble sur U, définissons alors les deux fonctions d'ensemble f^- et f^+ «extrêmes sur le vide» associées à f par

$$\begin{aligned} \forall A \subset U, A \neq \emptyset \quad & f^-(A) = f^+(A) = f(A) \\ & f^-(\emptyset) = \Delta_0 \\ & f^+(\emptyset) = \Delta_1 \end{aligned}$$

Définition 14 *Si $G = (X, U)$ est un graphe et f une fonction d'ensemble sur U, les f^--f^+-cocons de G sont appelés ses cocons équilibrés pour f, ou f-cocons.*

Comme nous l'avons vu, les cocons durs (*max-min*) sont obtenus par une procédure de type complete linkage, qui regroupe en priorité les deux ensembles dont la valeur de la plus forte connexion est minimale

(min du max des dissemblances des connexions). Nous nous sommes alors naturellement intéressés aux autres procédures de groupement de la même famille, utilisant des fonctions de priorité f autres que le max. Le résultat obtenu est que, sous une hypothèse faible sur f, une fusion selon la fonction de priorité f produit exclusivement des f-cocons. Détaillons maintenant ce résultat.

Définition 15 ***Hiérarchie de f-linkage***
Soit $G = (X, U)$ un graphe et f une fonction d'ensemble sur U. Si $Y \in \mathbb{P}(X)$ est une partition de X, une opération de f-linkage sur Y consiste à reunir les deux classes A et B de Y qui minimisent $f(A \sim B)$. La hiérarchie de f-linkage sur G est l'ensemble des parties obtenues par itération d'opérations de f-linkage à partir de la sur-partition absolue sur X. Nous la notons $\mathcal{H}_f(G)$.

Par exemple, un pas de *moy-linkage* (resp. *med-linkage*) fusionne les deux classes telles que la dissemblance moyenne (resp. médiane) sur leur frontière commune est minimale.

Définition 16 *Nous disons d'une fonction d'ensemble sur U qu'elle est «modérée»*[17] *quand elle vérifie* $\forall (A, B) \in \mathcal{P}^2(U)$

$$min\{f(A), f(B)\} \quad \leq \quad f(A \cup B) \quad \leq \quad max\{f(A), f(B)\} \tag{5.21}$$

On vérifie immédiatement que les moyennes et les quantiles de dissemblances sont des fonctions modérées. On a alors le

Théorème 2 *Si $G = (X, U)$ est un graphe sans boucle et f une fonction d'ensemble modérée sur U, alors*

$$\mathcal{H}_{f^+}(G) \ \subset \ \mathcal{C}_{f^-}^{f^+}(G). \tag{5.22}$$

Lemme 1 *Si L_p désigne la partition de X obtenue après p iterations de f^+-Linkage sur G, alors* $\forall p \in \langle 0, |X| \langle \quad \forall (A, B) \in L_p^2$

$$f^-(\widetilde{A}) \leq f^+(A \sim B). \tag{5.23}$$

Preuve du lemme :

La preuve est faite par récurrence sur les iterations de f^+-Linkage :

• À l'initialisation du f^+-Linkage, on a $L_0 = \{\{x\} : x \in X\}$ et donc comme G ne possède pas de boucle $\forall A \in L_0 : \widetilde{A} = \emptyset$. En conséquence

$$\forall (A, B) \in L_0^2 \qquad f^-(\widetilde{A}) = f^-(\emptyset) \leq f^+(A \sim B). \tag{5.24}$$

• Supposons maintenant que le lemme soit vérifié pour toutes les étapes $0 \ldots p$ de f^+-Linkage.

Soit A et B les ensembles de L_p qui minimisent $f^+(A \sim B)$.

Considérons tout d'abord le cas des ensembles autres que A ou B. Ces ensembles sont inchangés à l'étape $p + 1$ i.e. $\forall C \in L_p : C \notin \{A, B\} \Rightarrow C \in L_{p+1}$. Comme le lemme est vérifié pour l'étape p, ces ensembles vérifient

$$\begin{array}{rrcl} & f^-(\widetilde{C}) & \leq & f^+(C \sim A) \\ \text{et} & f^-(\widetilde{C}) & \leq & f^+(C \sim B) \\ \text{donc} & f^-(\widetilde{C}) & \leq & min\{f^+(C \sim A), f^+(C \sim B)\}. \end{array}$$

On vérifie aisément que si f est modérée (vérifie la propriété 5.21) alors f^- et f^+ le sont également. En appliquant alors la partie gauche de l'equation 5.21, on obtient

$$\begin{array}{rrcl} min\{f^+(C \sim A), f^+(C \sim B)\} & \leq & f^+\left((C \sim A) \cup (C \sim B)\right) & = \quad f^+(C \sim A \cup B) \\ \text{et donc} \quad f^-(\widetilde{C}) & \leq & f^+(C \sim A \cup B). & \end{array}$$

[17] Nous ne savons pas si cette propriété porte déja un nom...

Le lemme est donc vérifié à l'étape $p+1$ pour tous les ensembles inchangés entre p et $p+1$.

Considérons maintenant le cas de A et B. Après l'iteration $p+1$, ils sont fusionnés en $A \cup B$. Comme A et B sont disjoints

$$\widetilde{A \cup B} = \widetilde{A} \cup \widetilde{B} \cup (A \sim B).$$

Donc, d'après la partie droite de la propriété 5.21

$$f^-(\widetilde{A \cup B}) \quad \leq \quad max\{f^-(\widetilde{A}), f^-(\widetilde{B}), f^-(A \sim B)\}.$$

En appliquant le lemme pour l'étape p, on trouve alors que A et B vérifient

$$\begin{array}{rrcl} & f^-(\widetilde{A}) & \leq & f^+(A \sim B) \\ \text{et} & f^-(\widetilde{B}) & \leq & f^+(B \sim A) = f^+(A \sim B) \\ \text{donc} & f^-(\widetilde{A \cup B}) & \leq & f^+(A \sim B). \end{array} \tag{5.25}$$

Par ailleurs, (A, B) est le couple qui minimise $f^+(A \sim B)$, et donc $\forall C \in L_p$

$$\begin{array}{rrcl} & f^+(A \sim B) & \leq & f^+(A \sim C) \\ \text{et} & f^+(A \sim B) & \leq & f^+(B \sim C) \\ \text{donc} & f^+(A \sim B) & \leq & min\{f^+(A \sim C), f^+(B \sim C)\}. \end{array} \tag{5.26}$$

Comme $\forall C \in L_p : (A \cup B \sim C) = (A \sim C) \cup (B \sim C)$, la partie gauche de la propriété 5.21 donne $\forall C \in L_p$

$$min\{f^+(A \sim C), f^+(B \sim C)\} \quad \leq \quad f^+(A \cup B \sim C). \tag{5.27}$$

Finalement, comme $\forall C \in L_p : C \notin \{A, B\} \Rightarrow C \in L_{p+1}$, en regroupant les trois équations 5.25,5.26 et 5.27

on obtient

$$\forall C \in L_{p+1} \qquad f^-(\widetilde{A \cup B}) \quad \leq \quad f^+(A \cup B \sim C), \tag{5.28}$$

ce qui termine la preuve du lemme. ■

Preuve du théorème :

Par définition

$$\forall A \in L_p \qquad A \sim X \backslash A = \bigcup_{B \in L_p, B \neq A} A \sim B.$$

La partie gauche de 5.21 fournit alors $\forall p \in \{0, \ldots |X| - 1\}, \forall A \in L_p$

$$f^+(A \sim X \backslash A) \quad = \quad f^+\left(\bigcup_{B \in L_p, B \neq A} A \sim B\right) \quad \geq \quad \min_{B \in L_p, B \neq A} f^+(A \sim B).$$

Par application du lemme 5.23, on obtient finalement $\forall p \in \{0, \ldots |X| - 1\}, \forall A \in L_p$

$$\min_{B \in L_p, B \neq A} f^+(A \sim B) \quad \geq \quad f^-(\widetilde{A}).$$

et donc $\forall p \in \{0, \ldots |X| - 1\}, \forall A \in L_p$

$$f^+(A \sim X \backslash A) \quad \geq \quad f^-(\widetilde{A}).$$

qui est la définition du fait que A est un f^--f^+-cocon de G. ■

D'après ce théorème, même si par *moy* ou *med* linkage, on ne trouve pas *tous* les cocons équilibrés pour *moy* et *med* (puisqu'ils ne sont pas hiérarchisés), on ne trouve cependant *que* de tels cocons. De plus,

comme l'aggrégation se fait par ordre croissant sur $f(A \sim B)$, le f-linkage forme en priorité les f-cocons de plus faible contraste interne. Le f-linkage peut donc s'interpréter comme une procédure heurististique de calcul d'un ensemble de f-cocons hiérarchisés et de contraste interne aussi faible que possible.

Donc, en relâchant le critère dur de contraste, qui conduit à des ensembles hiérarchisés mais qui est trop sensible et produit peu d'ensembles, on aboutit à des critères trop souples, qui valident toutes les étapes d'un algorithme de groupement hiérarchique. Une coupe sur l'ultramétrique revient alors à un algorithme classique de segmentation seuillant sur un critère d'homogénéité.

Nous nous sommes alors engagés dans une autre voie pour calculer des coupes dans une hiérarchie, s'appuyant sur un principe de minimisation d'énergie. Différentes expériences ont été réalisées à partir d'un algorithme de groupement par lien moyen et la sélection d'une coupe de la hiérarchie sur un critère de codage minimal (Guigues, Le Men et Cocquerez, 2001b). Ce principe de coupe minimale nous a par la suite conduit à une méthode purement énergétique pour obtenir la hiérarchie elle-même, qui est décrite dans la troisième partie du document et qui s'avère finalement bien plus pertinente qu'une méthode de f-linkage. Nous ne reproduirons donc pas ici d'exemple de segmentations obtenues par la stratégie «hybride», f-linkage puis coupe minimale, voir (Guigues, Le Men et Cocquerez, 2001b).

Nous n'avons par contre pas fait d'expérience sur les familles de cocons médians ni avec des procédures de $[1-q]$-Linkage avec $q < 1/2$ et ce pour une raison pratique : la statistique de quantile n'admet pas de règle de mise à jour dynamique à coût constant : il faut systématiquement balayer l'ensemble des connexions et trier leurs valeurs pour calculer des quantiles, ce qui est relativement prohibitif en termes de temps de calcul. Les hiérarchies de $[1-q]$-linkage avec q faible sont pourtant certainement les plus pertinentes : les $[1-q]$-$[q]$-cocons avec q faible sont «presque» hiérarchisés, tout en étant plus robustes que les cocons durs.

5.7 Relations avec d'autres méthodes de groupement

Avant de conclure ce chapitre, nous examinons les relations entre le modèle de cocon et deux autres principes de groupement s'appuyant sur un graphe attribué par une dissemblance qui ont été développés pour la segmentation d'images.

5.7.1 Critère de contraste de Felzenswalb et Huttenlocher

Les cocons sont étroitement liés à la notion de «frontière évidente»[18] définie par Felzenszwalb et Huttenlocher (2001). Avec nos notations, ces auteurs définissent la «différence interne» («internal difference») d'un sous-graphe G_Y par

$$Int(Y) \triangleq \max_{e \in MST(G_Y)} \{d(e)\} \tag{5.29}$$

où $MST(G_Y)$ est un *arbre de poids minimal* recouvrant le sous-graphe induit G_Y[19]. L'unique différence avec notre «contraste interne» est que l'aggrégation des dissemblances est restreinte au MST de G_Y. L'aggrégation elle même est faite par un opérateur max, ce qui correspond au contraste interne des cocons «durs».

Felzenszwalb et Huttenlocher définissent également une différence entre deux sous-graphes Y et Z par

$$Dif(Y,Z) \triangleq \min(Y \sim Z) \tag{5.30}$$

Cette seconde mesure s'apparente au «contraste externe» des cocons durs mais porte sur deux ensembles voisins plutôt que sur un ensemble et son complémentaire.

[18] «evidence for a boundary» : littéralement «preuve qu'il existe une frontière».

[19] Remarquons que l'induit d'un MST du graphe entier est un MST du sous-graphe induit. C'est ce qui rend utilisable en pratique cette définition.

Finallement, Felzenszwalb et Huttenlocher considèrent qu'il existe une «frontière évidente» entre Y et Z si et seulement si

$$Dif(Y,Z) \quad > \quad \min\{Int(Y)+\tau(Y), Int(Z)+\tau(Z)\} \tag{5.31}$$

avec $\tau(Y) = k/|Y|$, où k est un paramètre et $|Y|$ la taille de l'ensemble Y.

Ils proposent alors un algorithme efficace - s'appuyant sur l'algorithme de base de croissance d'un MST - qui permet de calculer une «bonne segmentation» orientée frontières pour le prédicat «être évidente», au sens de la définition que nous avons donnée en section 5.2.8 p. 75.

5.7.2 Critère de coupe normalisée de Shi et Malik

Les cocons possèdent également une parenté avec le critère de *coupe normalisée* (normalized cut) de Shi et Malik (2000).

Pour mettre en évidence cette parenté, notons que si les dissemblances sont positives, on peut définir une valeur du **contraste** d'un sous-graphe induit Y par

$$\mathbf{C}_{f,g} \quad \triangleq \quad \frac{\mathrm{E}_g(Y)}{\mathrm{I}_f(Y)}. \tag{5.32}$$

Y est alors un cocon (strict) si et seulement si son contraste est (strictement) supérieur à 1.

Shi et Malik s'intéressent au problème de *division* d'une image en deux régions pertinentes. Ils modélisent le problème comme un problème de *coupe* dans un graphe : trouver un ensemble d'arêtes qui divise le graphe en deux composantes connexes (les graphes considérés s'appuient sur les pixels de l'image mais contiennent des voisinages d'ordre supérieur à 1 pour que la décision de découpage soit plus robuste). Les arêtes sont attribuées par une dissemblance d.

Wu et Leahy (1993) avaient précédemment envisagé ce problème et l'avaient modélisé comme un problème de coupe de poids total minimal (ce qui se résout par un algorithme de flot maximal en vertu du théorème de Ford-Fulkerson). Cependant, comme le remarquent Shi et Malik, ce critère de coupe possède plusieurs défauts. D'une part, il ne s'intéresse qu'aux dissemblances sur la frontière entre les deux ensembles. D'autre part, il n'est pas «normalisé» en fonction de la longueur de la frontière : le critère divise généralement le graphe en un singleton et son complémentaire car la coupe est alors très courte, même si les dissemblances sont fortes.

Avec nos notations, Shi et Malik proposent alors de diviser un graphe $G = (X, U, d)$ en deux, un ensemble Y et son complémentaire, en choisissant le Y qui minimise :

$$\begin{aligned} NCut(Y) \quad &\triangleq \quad \frac{f(\widetilde{Y})}{f(Y \sim X)} + \frac{f(\widetilde{X\backslash Y})}{f(X\backslash Y \sim X)} \\ &= \quad \frac{f(\widetilde{Y})}{f(\widetilde{Y}) + f(Y \sim X\backslash Y)} + \frac{f(\widetilde{X\backslash Y})}{f(\widetilde{X\backslash Y}) + f(Y \sim X\backslash Y)} \\ &= \quad \frac{\mathrm{I}_f(Y)}{\mathrm{I}_f(Y) + \mathrm{E}_f(Y)} + \frac{\mathrm{I}_f(X\backslash Y)}{\mathrm{I}_f(X\backslash Y) + \mathrm{E}_f(X\backslash Y)} \end{aligned} \tag{5.33}$$

où f est la fonction *somme* des dissemblances.

La «coupe» qui minimise $NCut$ est appelée la **coupe normalisée** (on voit que ce n'est plus seulement la coupe $Y \sim X\backslash Y$ entre Y et son complémentaire qui intervient, mais également les connections internes aux deux ensembles).

Les deux termes de la dernière équation sont totalement symétriques en Y et $X\backslash Y$ (c'est heureux !). Définissons alors pour un sous graphe Y :

$$\mathbf{C}'_f(Y) \quad \triangleq \quad \frac{\mathrm{I}_f(Y) + \mathrm{E}_f(Y)}{\mathrm{I}_f(Y)} \quad = \quad 1 + \frac{\mathrm{E}_f(Y)}{\mathrm{I}_f(Y)} \quad = \quad 1 + \mathbf{C}_{f,f}(Y).$$

$\mathbf{C}'_f(Y)$ correspond donc à une variante de la définition 5.32 du contraste du sous-graphe Y.

La bipartition qui minimise $NCut(Y)$ est alors celle qui maximise

$$\frac{1}{NCut(Y)} = \frac{1}{\frac{1}{\mathbf{C}'_f(Y)} + \frac{1}{\mathbf{C}'_f(X \setminus Y)}}.$$

$NCut$ maximise donc la moyenne harmonique des contrastes des deux ensembles auxquels on a ajouté 1.

Notons que la fonction f est ici une somme, c'est-à-dire n'est pas normalisée en fonction de la taille de l'ensemble. Cela ne pose pas de problème ici car l'ensemble et son complémentaire interviennent de manière symétrique.

Papadimitriou a malheureusement montré que l'optimisation de $NCut$ est NP-difficile, y compris restreinte à des graphes planaires ou même des maillages. Shi et Malik l'approximent par une technique s'appuyant sur la théorie spectrale des graphes (transposition au continu du problème de labellisation binaire, qui conduit à un problème de valeurs propres généralisées, puis binarisation de la solution continue par seuillage).

Le critère de coupe normalisée a été largement utilisé par Shi et Malik puis par des collaborateurs ultérieurs pour proposer des solutions à de nombreux problèmes de segmentation (niveaux de gris, texture, mouvement...).

5.7.3 Le problème de «comparaison avec le vide»

Tous les modèles de qualification d'ensembles qui s'appuient sur une comparaison entre intérieur et extérieur des ensembles font face au problème de «comparaison avec le vide».

En effet, les singletons, dont l'intérieur est vide, comme l'ensemble complet, dont l'extérieur est vide, ont un statut particulier. Ces «vides» sont modélisés mathématiquement comme des ensembles vides, cependant il correspondent en réalité à une absence d'information concernant les sous ou super-structures.

Pour la bonne définition des cocons, quelles que soient les fonctions f et g utilisées, nous devons poser $f(\emptyset) = \Delta_0$ et $g(\emptyset) = \Delta_1$. Tous les singletons et toutes les composantes connexes sont alors toujours des cocons. Si $\Delta =]0, +\infty[$, cela se retrouve par le fait que le contraste (equ. 5.32) de ces cocons triviaux est infini.

Pour leur critère de «frontière évidente», Felzenszwalb et Huttenlocher doivent ajouter un terme non nul à la différence interne des sous-graphes, pour éviter la division par zéro dans le cas de singletons. Le critère d'«évidence» dépend de manière critique de l'introduction de cette fonction τ. Sans elle, toutes les frontières entre deux singletons voisins seraient «évidentes» ! Felzenszwalb et Huttenlocher ne mentionnent pas ce problème, si celui de l'indétermination du contraste externe pour le graphe entier.

De son côté le critère de coupe normalisée *divise toujours* un graphe, i.e. le couple $(X, \emptyset)$ n'est jamais un minimum de $NCut$ sauf dans le cas pathologique d'un graphe sans arête, auquel cas $NCut$ est indécis entre toutes les solutions (c'est heureux). Donc $NCut$ ne peut pas être utilisé comme critère de décision quand à la division ou non d'un graphe.

$NCut$ n'est défini que pour la bipartition d'un graphe. Le critère se généralise cependant de façon naturelle à des partitions quelconques :

$$NCut(P) = \sum_{Y \in P} \frac{f(\widetilde{Y})}{f(Y \sim X)} \tag{5.34}$$

où P est une partition des nœuds du graphe.

Cependant, cette expression ne peut pas être utilisée pour sélectionner le cardinal de la partition : elle est toujours minimale (nulle) pour la sur-partition absolue. Encore une fois, ce problème provient de

l'effet de bord de «comparaison avec le vide»[20].

Ce problème intervenant aux limites du domaine perceptuel, au niveau des «atomes» et de l'«Univers», est extrêmement profond. Il réapparaîtra avec le modèle énergétique de la prochaine partie.

5.8 Conclusion

Les structures hiérarchiques ont été depuis longtemps étudiées et utilisées en classification et analyse de données, puis en segmentation d'image. Néanmoins, la plupart du temps la structuration hiérarchique est en quelque sorte forcée, c'est-à-dire que l'on applique une méthode de groupement qui, par nature même du processus produit une hiérarchie, sans obligatoirement savoir si, d'une part les objets que l'on cherche à déterminer sont hiérarchisés, et d'autre part si la méthode de groupement hiérarchique employée fournit bien des ensembles qui modélisent les objets recherchés.

Ici, le résultat de hiérarchisation (des cocons durs) a un autre caractère : la définition de certains sous-graphes possédant une propriété particulière (que l'on espère pertinente) conduit à des objets naturellement hiérarchisés.

Il est important de relever que, même dans le cadre d'un critère de pertinence très strict (cocons durs), les régions visuelles potentiellement pertinentes ne forment absolument pas une partition de l'image. Les modèles de cocons durs étayent au contraire la thèse d'une structuration hiérarchique des zones remarquables d'une image.

Insistons également sur le fait que le *f-linkage* est par nature un algorithme de groupement sur un critère *local*. Cependant, il conduit à des ensembles qui vérifient une propriété *globale* de contraste. Les objets globaux que l'on obtient par single linkage - l'arbre de poids minimal ou l'ultramétrique sous-dominante - étaient connus de longue date. Le formalisme des cocons caractérise les objets globaux que détermine le complete-linkage. En outre, et de manière importante, il montre que toutes les étapes d'un groupement hiérarchique ne sont pas pertinentes (ce que nous retrouverons également avec le modèle énergétique). La notion de cocon nous a également permis de mettre en évidence que les ensembles produits par les algorithmes de lien moyen ou médian vérifient tous une propriété globale de contraste, propriété qui est directement liée à la fonction de priorité employée pour le groupement.

Ces résultats fournissent donc une vision nouvelle sur ce que réalisent certains membres d'une famille classique d'algorithmes de groupement hiérarchique, qui - à notre connaissance - n'avaient jusqu'à présent qu'une justification intuitive (mis à part pour le single linkage).

Toutefois, comme nous l'avons vu, le modèle de cocon ne permet pas d'introduire de critère géométrique dans la segmentation. En outre, si l'on dépasse le cadre des cocons durs - qui le sont un peu trop - on retombe sur des algorithmes classiques. Le critère de contraste n'agit plus comme un critère de sélection de certaines étapes remarquables d'un groupement hiérarchique.

Le modèle de cocons nous a néanmoins conduit à envisager le partitionnement d'image en deux étapes clairement distinctes : construction d'une hiérarchie suivie d'une coupe dans cette structure. Cette idée soutend toute la méthodologie de segmentation multi-échelles s'appuyant sur un formalisme énergétique que nous proposons d'aborder maintenant.

[20] Après avoir écrit ces lignes, nous avons été solicités comme relecteur pour un article de Stella Yu et Jianbo Shi soumis en 2003 à IEEE trans. PAMI. Ces auteurs proposent dans cet article d'étendrent le critère de coupe normalisée au cas de segmentation en k classes avec k fixé. Ils proposent effectivement l'expression que nous donnons ci-dessus et font la même remarque que nous concernant l'incapacité du critère à décider du nombre de classes...

Analyse ensembles-échelle d'une image

Chapitre 6

Présentation de l'approche

Cette partie (chapitres 6 à 9) développe une deuxième approche multi-échelles de la segmentation d'images, mais qui s'appuie sur des principes totalement différents de ceux qui sont à l'origine des cocons.

À l'origine des cocons, il y avait l'idée de parvenir à caractériser certaines régions «remarquables» d'une image. Nous nous posions ainsi un problème de *décision* concernant les *parties* d'une image. Le développement d'un modèle adéquat dans le cadre de la théorie des graphes nous a alors conduit à des ensembles hiérarchisés : la structure multi-échelles était une *conséquence* du modèle.

A contrario, le modèle présenté ici part d'un problème d'*optimisation* formulé sur les *partitions* d'une image et *pose pour principe* de dégager une famille de solutions multi-échelles.

Nous aboutirons toutefois à la même structure de solution : une hiérarchie de régions indicée par un critère croissant que nous interpréterons comme un paramètre d'échelle (pour des raisons bien plus profondes que pour les cocons). Une telle structure sera appelée ici une *représentation ensembles-échelle (scale-sets)* d'une image car, comme nous le verrons, elle constitue le pendant orienté régions des représentations scale-space issues du domaine du filtrage et des théories de la diffusion. La méthode d'analyse qui est proposée sera donc appelée *méthode d'analyse ensembles-échelle* d'une image.

La méthode s'inscrit dans la lignée des approches par minimisation d'énergie de la segmentation d'images (approches variationnelles, bayesiennes, par codage minimal...). Ces différentes approches ont un point de départ commun : elles envisagent la segmentation comme une question de détermination d'un modèle de l'image, modèle défini par morceaux.

Commençons donc par examiner ce problème de modélisation et voir la réponse qu'y apportent chacune de ces théories.

6.1 Modélisation d'images par morceaux

L'idée d'envisager la segmentation (et d'autres problèmes apparentés, comme la restauration) comme un problème de modélisation est apparue au cours des années quatre-vingt.

On peut classer les différentes théories en fonction de la nature des images qu'elles considèrent (continues ou discrètes) et de la nature des modèles d'images qu'elles considèrent (modèles déterministes, probabilistes ou mixtes). Les formulations *variationnelles* envisagent des images continues et des modèles déterministes. Les formulations *bayesiennes* employées en image considèrent habituellement des images discrètes et des modèles probabilistes. Enfin, les formulations par *codage minimal* considèrent des images discrètes et peuvent intégrer des modèles mixtes, comportant des composantes déterministes et des composantes stochastiques.

Avant de détailler chacune de ces théories, examinons le problème général de modélisation.

6.1.1 Recherche d'un modèle «optimal»

Comparaison et unicité de solution

Considérons l'ensemble $\mathcal{I}$ des images de $\mathcal{D}$ dans $\mathcal{V}$. Notons alors $\mathcal{M}$ un ensemble de *modèles* possibles de ces images. Dans le cadre d'un problème de partitionnement, ces modèles sont définis par morceaux. Un modèle d'image segmentée est un couple (P, J) où $P = \{R_i\}_{i \in K}$ est une partition de l'image et $J = \{J_i\}_{i \in K}$ un ensemble de modèles de chaque sous-image R_i. Les J_i sont généralement des modèles paramétriques, déterministes (modèles fonctionnels des valeurs de l'image), stochastiques (modèle statistique de la distribution des valeurs) ou mixtes (modélisation probabiliste des écarts à un modèle fonctionnel). Notons Θ_i l'ensemble des paramètres déterminant J_i (si le modèle n'est pas paramétrique mais correspond par exemple à une fonction, Θ_i représente alors la donnée explicite des valeurs de cette fonction).

Face à une image observée $I \in \mathcal{I}$, l'objectif de la segmentation est alors de sélectionner dans $\mathcal{M}$ l'instance M qui «modélise le mieux» I. On parle également de problème «inverse» quand le modèle est conçu comme une «cause» des observations, cause à laquelle il s'agit de remonter ou qu'il s'agit de séparer d'autres facteurs (par exemple séparer la lumière émise par les surfaces du bruit introduit par le système de mesure).

Le problème d'inférence de modèle est alors conçu comme un problème d'optimisation, en vertu d'un principe explicité par Koepfler, Lopez, et Morel (1994) sous le nom de **principe de comparaison** : «Nous adoptons un principe sans lequel aucune discussion sur la segmentation ne peut avoir lieu, et que nous appelons le principe de comparaison. Selon ce principe, étant données deux segmentations différentes d'une même donnée, nous pouvons toujours décider laquelle des deux doit être considérée comme meilleure que l'autre (ou équivalente). Nous supposons donc l'existence d'un ordre total sur l'ensemble des segmentations possibles, ce qui se traduit par le fait qu'il existe une fonctionnelle réelle E telle que si $E(K_1) < E(K_2)$, alors la segmentation K_1 doit être condidérée comme "meilleure" que la segmentation K_2» [1].

Selon ce principe, la formulation générale d'un problème de segmentation revient à spécifier une fonction de coût[2], ou **énergie**, $\mathsf{E} : \mathcal{M} \times \mathcal{I} \to \mathbb{R}^+$ sur l'ensemble des segmentations et des images possibles et à définir la(les) solution(s) d'un problème de segmentation d'une image I comme la(les) segmentation(s) $M^*(I)$ qui minimise(nt) cette énergie, c'est-à-dire par

$$M^*(I) = \underset{M \in \mathcal{M}}{argmin} \quad \mathsf{E}(M, I).$$

En pratique, l'adoption du principe de comparaison est immédiatement suivie de l'adoption d'un deuxième principe, celui d'**unicité de la solution**, et c'est réellement la mariage de ces deux principes qui constitue l'axiomatique de base des méthodes de partitionnement par minimisation d'énergie[3].

En substance, ces deux principes (comparaison et unicité) expriment le fait qu'une méthode de sélection de solution ne doit relever d'aucun arbitraire : Si l'on pose le problème comme un problème de partitionnement et que l'objectif est de ne retenir qu'*une unique* partition du treillis alors la méthode qui permet de choisir cette partition ne doit pas être équivoque. Une manière d'exprimer cette condition est de formuler la segmentation comme un problème d'optimisation possédant un unique minimum. Ce n'est cependant pas la seule : un problème de décision peut également admettre une solution unique.

[1] «We shall adopt a principle without which no discussion about segmentation can even start, and which we call comparison principle. It states that given two different segmentations of a datum, we are always able to decide which of them is considered as better than (or equivalent to) the other. Thus we assume the existence of some total ordering over all possible segmentations, and this can be simply acheived only if this ordering is reflected by some real functional E such that if $E(K_1) < E(K_2)$, then the segmentation K_1 has to be considered "better" than the segmentation $K2$.»

[2] Koepfler et al. parlent de fonctionnelle car ils envisagent le problème pour des images «idéales», définies sur des domaines continus, auquel cas les segmentations sont des fonctions et les fonctions de coût des fonctionnelles.

[3] En continu, on suppose également l'existence d'une solution dans $\mathcal{M}$, ce qui est automatique sur un ensemble discret, qui est compact. Voir également ci-dessous la notion de problème variationnel «bien posé».

Pensons aux segmentations par ligne de partage des eaux ou aux modèles de cocons. On peut toujours formuler un problème de décision en termes d'optimisation, mais c'est souvent artificiel. Insistons sur le fait que les principes dont nous venons de parler portent sur la cohérence entre l'objectif poursuivi et la méthode qui permet de l'atteindre, indépendamment de la pertinence de l'objectif poursuivi : comme nous l'avons amplement discuté, il n'existe certainement pas d'unique «bonne» segmentation d'une image, ni de meilleur choix dans l'absolu. Nous interprétons finalement le principe de comparaison de Koepfler et al. comme une critique - fondée - à l'égard des méthodes classiques de segmentation - par prédicats - qui, bien que le formalisme ne détermine pas une solution unique, en sélectionnent tout de même une seule via une heuristique incorporée dans le processus de construction de la solution.

Une question de compromis

Tout problème de modélisation ou de description de données soulève un problème de *compromis* : il y a nécessairement un compromis à faire entre la **précision** du modèle et sa **simplicité**. Nous discuterons amplement ci-dessous de l'origine et de la signification de ce compromis, qui est au cœur de notre approche. Commençons par le mettre en évidence pour le problème de segmentation.

En partitionnement d'image, si l'espace $\mathcal{M}$ des modèles considérés est assez «riche», et que l'on n'exprime la qualité $\mathsf{E}(M, I)$ d'un modèle que sous la forme d'un écart entre le modèle et l'image, plusieurs problèmes surviennent.

Considérons pas exemple pour $\mathcal{M}$ l'ensemble $\mathcal{C}$ des images constantes par morceaux (fonctions étagées). Comme l'espace des images numériques est inclu dans $\mathcal{C}$, l'image elle-même est alors toujours solution du problème de minimisation : elle a une énergie nulle, quelle que soit la manière dont on exprime l'écart entre modèle et image. En outre, si l'image possède au moins deux valeurs égales, il y a au moins deux façons de la décrire par morceaux (éventuellement non connexes) : en regroupant les deux pixels égaux dans un même «morceau» ou en les séparant en deux «morceaux». Une telle formulation conduit donc à deux problèmes :

1. Il existe toujours une solution triviale (l'identité) qui ne nous intéresse évidemment pas.

2. Si le langage de description des éléments de $\mathcal{M}$ est tel que plusieurs descriptions définissent la même image alors la solution n'est pas nécessairement unique *dans l'espace des descriptions* (l'espace des paramètres de modèles).

Dans le cas qui nous intéresse, le deuxième point peut être considéré uniquement comme un problème technique : on peut toujours regrouper les descriptions qui conduisent à une même image dans une même classe d'équivalence et en choisir un représentant[4].

Le premier problème est plus profond. On a tout simplement oublié de spécifier un aspect essentiel du problème : nous recherchons un *modèle*, c'est-à-dire une description *simplifiée* de l'image.

Toutes les théories d'inférence de modèle proposent alors d'ajouter un terme à l'énergie E, portant uniquement sur le modèle et qui quantifie sa *complexité*[5]. L'inférence relève alors d'un compromis entre la complexité du modèle et sa capacité à décrire fidèlement les données. L'énergie à minimiser s'écrit finalement

$$\mathsf{E}(M, I) = \mathsf{C}(M) + \mathsf{D}(I, M) \tag{6.1}$$

où C est une mesure de la «complexité» de M et D est une mesure de l'«écart» entre M et I (nous verrons ci-dessous les différents noms que l'on donne à ces quantités en fonction du cadre théorique dont elles sont issues).

Notons qu'au moins pour les images constantes par morceaux le problème de non unicité de solution est également résolu par ce procédé : il suffit, pour deux modèles qui déterminent la même image, d'attribuer

[4]Cette remarque ne tient pas quand on se pose un problème d'interpolation ou d'extrapolation. C'est dans ce cas la technique de régularisation qui permet de sélectionner une solution unique.

[5]Le fait que ce terme porte uniquement sur le modèle n'est pas toujours vrai : nous verrons ci-dessous que les approches par codage minimal aboutissent à des termes de complexité qui dépendent également des données.

l'énergie C la plus faible à celui qui s'appuie sur la partition la plus grossière, en argumentant de sa plus grande «simplicité».

6.1.2 Formulations variationnelles

Dans les formalisations continues du problème de segmentation, la nécessité d'introduire un terme de complexité de modèle dans les fonctionnelles provient de la théorie de la régularisation. Si l'on n'utilise qu'un terme de distance entre M et I, le problème d'optimisation est généralement «mal posé»[6]. Dans ce cadre, le terme D est souvent appelé un terme *d'attache aux données* et C un terme de *régularisation*. Dans la théorie classique de la régularisation, dans laquelle les modèles sont des fonctions différentiables, les termes de régularisation portent sur l'amplitude des variations d'ordre p de la solution, favorisant donc, à l'ordre 1 les solutions les plus «lisses», à l'ordre 2 les moins «courbes», etc.

La fonctionnelle de Mumford et Shah (1989) est l'archetype des formulations variationnelles du problème de partitionnement. Mumford et Shah envisagent le problème de segmentation d'une image $I : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}$ comme un problème d'approximation par une fonction lisse par morceaux.

Comme en section 4.3 p. 53, définissons une partition de $\mathcal{D}$ comme une décomposition de $\mathcal{D}$ en $R_1 \cup \cdots \cup R_n \cup \Gamma$, où les régions R_i sont des ensembles ouverts disjoints tels que la fermeture de leur union est $\mathcal{D}$. Γ représente alors l'ensemble 1D des frontières des régions.

Une fonction f sur $\mathcal{D}$ est dite *lisse par morceaux* quand il existe une partition $P = (R_i, \Gamma)$ de $\mathcal{D}$ telle que f est différentiable sur $\mathcal{D}\backslash\Gamma$ mais pas sur l'ensemble Γ des frontières de la partition. Mumford et Shah ont appelé ce type de fonction un modèle *bande dessinée* (*cartoon model*) car il est constitué de plages régulières de couleur délimitées par des discontinuités franches.

Les auteurs proposent alors d'approximer une image I par la fonction lisse par morceaux qui minimise la fonctionnelle

$$\mathsf{E}(f,\Gamma) = \mu^2 \iint_{\mathcal{D}} (f-I)^2 dxdy + \iint_{\mathcal{D}\backslash\Gamma} ||\nabla f||^2 dxdy + \nu \int_{\Gamma} ds. \tag{6.2}$$

Le premier terme traduit la qualité de la régression de I par f, le terme de gradient exprime une préférence pour les fonctions les plus lisses à l'intérieur des régions et le troisième terme une préférence pour les partitions qui comportent la plus faible longueur totale de frontières. μ et ν sont des paramètres qui règlent l'importance relative de ces différents facteurs énergétiques. Si l'un des trois termes est supprimé, le problème dégénère en un problème mal posé.

Quand on impose à f d'être constante par morceaux, on aboutit ce que nous appellerons une *caricature par plages* de l'image. f est solution de

$$\mathsf{E}(f,\Gamma) = \iint_{\mathcal{D}} (f-I)^2 dxdy + \nu \int_{\Gamma} ds. \tag{6.3}$$

La valeur de f sur chaque région est alors donnée par la moyenne de I sur la région (si l'on fixe Γ, on aboutit à un problème trivial de minimisation quadratique sur chaque région).

Mumford et Shah ont montré d'importantes propriétés des minima des deux fonctionnelles qui précèdent (et d'un autre cas limite de 6.2 qui s'apparente à un problème de géodésique). Plusieurs variantes ont également été proposées autour de ces modèles. Les ouvrages de Blake et Zisserman (1987) et de Morel et Solimini (1995) sont consacrés à l'étude théorique de ce type de problèmes variationnels et à leur résolution numérique.

Dans le domaine variationnel, notons également les modèles de contours actifs et de régions actives (cf. 3.5.2 p. 40). Les énergies considérées comportent encore une fois des termes d'attache aux données (gradient, homogénéité...) et des termes de régularisation (courbure, torsion...). Dans ce cas, le rôle de la «régularisation» est essentiellement géométrique et non pas de «régulariser» un problème mal posé.

[6] Un problème variationnel est «mal posé» au sens d'Hadamard si l'une des conditions suivantes fait défaut : la solution existe, est unique et dépend continuement des données, voir (Poggio, Torre, et Koch 1985).

6.1.3 Modélisations probabilistes

Dans les approches probabilistes de la segmentation, l'image est vue comme une réalisation d'un processus aléatoire que l'on cherche à modéliser. Chaque modèle M est un élément d'une famille de modèles stochastiques paramétriques.

Inférence bayesienne

En inférence statistique de modèle, si l'*ordre* des différents modèles possibles est fixe, i.e. s'ils sont tous déterminés par le même nombre de paramètres, l'approche classique (due à Fisher) elle celle du **maximum de vraisemblance**. Elle consiste à sélectionner le modèle M qui maximise la probabilité $P_v(I|M)$ d'observer I sachant M. Si l'ordre des modèles est variable, comme c'est le cas en segmentation où la dimension de l'espace de paramètres dépend par exemple du nombre de régions, cette stratégie exhibe le modèle qui s'adapte le mieux aux données, indépendament de tout autre critère. Cette solution n'est généralement pas satisfaisante. Par exemple, si l'on modélise l'image comme le résultat d'un ensemble de tirages indépendants selon des lois gaussiennes de mêmes paramètres (moyenne, variance) sur chaque région, l'hypothèse la plus «vraisemblable» est alors de découper l'image en autant de régions que de pixels et de considérer que la valeur observée d'un pixel résulte d'un tirage selon une gaussienne centrée sur cette observation. On trouve donc une solution triviale.

Pour résoudre ce type de problème, la théorie de l'*inférence bayesienne* propose d'introduire une loi *a priori* P_p sur l'espace $\mathcal{M}$ des modèles. D'après la loi de Bayes, la probabilité *a posteriori* de M sachant I est alors donnée par

$$P_P(M|I) = \frac{P_p(M)P_v(I|M)}{P_i(I)}$$

où $P_i(I)$ désigne la probabilité d'observer l'image I parmi toutes les images possibles.

La stratégie la plus courante pour sélectionner M est alors la stratégie du **maximum a posteriori (MAP)** : sélectionner le modèle M auquel l'image I accorde la plus forte probabilité a posteriori $P_P(M|I)$. Comme $P_i(I)$ est constante sur $\mathcal{M}$, elle n'intervient pas dans l'optimisation. Maximiser $P_P(M|I)$ est alors équivalent à minimiser $-\log P_P(M|I)$, soit à minimiser

$$-\log P_p(M) - \log P_v(I|M). \tag{6.4}$$

On retrouve donc une expression à minimiser qui comporte une somme de deux termes. Notons que le cadre bayesien général n'indique pas comment choisir la loi a priori $P_p(M)$: elle est sensée traduire nos croyances, connaissances ou souhaits relatifs à la solution recherchée, ce qui dépend de chaque problème particulier.

Champs de Markov

Dans le cadre bayesien, une approche introduite par Geman et Geman (1984) et désormais classique (voir par exemple le livre de Li (2001)) consiste à modéliser une partition comme une réalisation d'un **champ de Markov** (*Markov Random Field*, **MRF**).

Un MRF est un ensemble $X = \{X_1 \dots X_n\}$ de variables aléatoires muni d'une relation de voisinage $U \subset X^2$ - i.e. (X, U) est un graphe symétrique - qui vérifie que la probabilité conditionnelle d'observer une réalisation x en un point X_i du champ sachant les réalisations pour les autres points du champ ne dépend que des réalisations effectives du champ pour les voisins de X_i. Le théorème de Hammersley-Clifford établit alors qu'un champ de v.a. est Markovien si et seulement si ses configurations obéissent à une loi de Gibbs, c'est-à-dire ssi leur probabilité peut s'écrire (à un facteur de normalisation près)

$$P(X = \mathsf{X}) \quad \propto \quad e^{-\sum_{c \in \mathcal{C}} \mathsf{E}(c = \mathsf{c})}$$

où $\mathcal{C}$ est l'ensemble des *cliques* induites par la relation de voisinage[7]. L'exposant, $\mathsf{E}(X) = \sum_{c \in \mathcal{C}} \mathsf{E}(c)$, s'appelle le potentiel de Gibbs du champ.

Pour la segmentation, on considère habituellement des champs à valeurs entières dont les réalisations sont interprétées comme des images d'étiquettes représentant des partitions du domaine de l'image. À chaque région i est alors attaché un modèle stochastique J_i de ses réalisations image, modèle qui comporte également certains paramètres propres, Θ_i, qui sont à estimer. Le modèle complet que l'on cherche à inférer est donc (X, Θ), où $\Theta = \{\Theta_i\}$ avec i situé dans l'espace de réalisation effectif de X.

Si l'on considère le critère du MAP (exprimé en opposé du «log-posterior» comme en 6.4), on cherche à trouver (X, Θ) qui minimise une expression de la forme

$$\mathsf{E}(X) - \log P_v(I|X, \Theta). \tag{6.5}$$

Cette minimisation peut se décomposer en deux : à X fixé, Θ est la solution du maximum de vraisemblance, pour laquelle on dispose souvent d'une expression analytique. Le problème «difficile» est alors de trouver la segmentation, c'est-à-dire la configuration du champ X qui conduit à l'énergie minimale.

Un potentiel de Gibbs classique est le potentiel de Potts qui considère un MRF s'appuyant sur des 4-voisinages et attribue un potentiel nul à deux voisins d'étiquettes semblables, et constant - de valeur α - sinon (le cas particuliers à deux classes s'appelle le modèle d'Ising). On aboutit alors à un analogue discret de l'énergie de longueur de frontières de Mumford-Shah :

$$\mathsf{E}(X) = \alpha \times \big|\{(i,j) \quad \text{4-voisins} \quad | \quad X(i) \neq X(j)\}\big|.$$

Associons ce potentiel a priori à un modèle d'image constante par région et bruitée par un bruit gaussien i.i.d. $\mathcal{N}(0, \sigma)$ de variance fixe. Le paramètre à estimer sur chaque région est donc la valeur constante de l'image non bruitée. À X fixé, la solution $\hat{\Theta}$ du maximum de vraisemblance est alors le vecteur des moyennes empiriques des valeurs de l'image sur chaque région. Si l'on note $\overline{I}_X$ l'image obtenue en reportant ces moyennes sur chaque pixel, on obtient alors

$$P_v(I|X, \hat{\Theta}) \quad \propto \quad \prod_{x \in \mathcal{D}} e^{-||I(x) - \overline{I}_X(x)||^2}$$

En développant l'expression 6.5 pour ces deux potentiels, on aboutit finalement à une version discrète du modèle constant par morceaux de Mumford-Shah. Le paramètre α règle le taux de régularisation géométrique de la segmentation.

D'autres formalisations Markoviennes s'affranchissent de la modélisation par étiquettage d'une partition en introduisant des «processus de bord», c'est-à-dire en modélisant explicitement la présence ou l'absence de frontière entre deux pixels voisins (Geman et Geman 1984).

L'approche par processus d'étiquettage Markovien et MAP peut être vue comme une extension des techniques de classification ponctuelle en k classes. L'introduction d'un modèle a priori (de type Potts) provoque deux types de régularisation. Elle permet de laisser libre k, i.e. de mettre en compétition des partitions d'ordre différent, et de trouver pour α assez grand une partition autre que la sur-partition absolue (le potentiel de Potts est maximal sur cette partition). Elle permet également d'introduire, entre deux k-partitions, une préférence pour la plus régulière géométriquement (au sens de la longueur totale de frontières).

Notons que la *représentation* d'une partition par une image d'étiquettes ne permet pas d'introduire de contrainte de connexité sur les régions de la solution. On sait pourtant par exemple que l'optimum du modèle constant par morceaux décrit ci-dessus est une partition à composantes connexes. Cette remarque a conduit Wang (1998) à proposer une méthode d'optimisation stochastique d'un critère de MAP qui s'appuie sur une représentation de la solution comme une partition d'un graphe en *composantes connexes*.

[7] Une clique est un sous-graphe complet d'un graphe. Par exemple pour un système de 4-voisinages sur une grille discrète, les cliques sont les points isolés et les couples de voisins horizontaux et verticaux.

Les «processus stochastiques objet» sont des extensions récentes des approches par MRF dans lesquelles les configurations recherchées ne sont pas des réalisations de champs d'étiquettes pixellaires mais des configurations d'ensembles d'objets en interaction : triangulations, formes géométriques élémentaires, e.g. triangles (Drot, Descombes, Le Men, et Zérubia 2002), ou plus complexes, e.g. modélisant des bâtiments à deux pans (Garcin, Descombes, Zérubia, et Le Men 2001). Les lois a priori sur les configurations d'objets sont généralement Markoviennes d'ordre faible : elles probabilisent les configurations géométriques via l'interaction entre paires d'objets (cliques d'ordre 2) : non recouvrement, alignement, etc.

6.1.4 Inférence par codage minimal

Comme la théorie bayesienne, les théories dites d'*inférence par codage minimal* (*Minimum Encoding Inference*, **MEI**) se positionnent dans le cadre de l'inférence statistique de modèle, dans le contexte non classique où des modèles d'ordres différents sont en compétition, mais proposent d'aborder le problème sous l'angle de la théorie de l'information.

Parmi ces théories, la théorie **MDL** (Minimum Description Length - Longueur Minimale de Description) de Jorma Rissanen est de loin la plus connue dans le domaine de l'analyse d'images. L'article de Rissanen de 1978 dans lequel il propose la première version du principe MDL (car il y en a plusieurs) est souvent cité comme l'article fondateur des théories d'inférence par codage minimal. Les idées principales avaient cependant été dégagées dix ans auparavant par Wallace et Boulton (1968). Wallace a par la suite développé ses idées et proposé avec Freeman (1987) la méthode dite **MML** (Minimum Message Length - Longueur Minimale de Message)[8].

Les théories MEI se présentent comme des formalisations du *principe de parcimonie* ou principe du *rasoir d'Occam*. Comme nous l'avons vu ci-dessus (sec. 3.2.2 p. 22), Ernst Mach pensait que l'élaboration de nos modèles mentaux relève fondamentalement d'un principe d'«économie de représentation». Dans cette lignée, les théories d'inférence par codage minimal érigent le principe de parcimonie en mécanisme fondamental d'inférence de modèle[9]. Yvan Leclerc, à qui l'on doit l'introduction de la théorie du codage minimal en segmentation d'images, propose ainsi de placer «la simplicité de description comme moteur de l'interprétation visuelle» (Leclerc 1989b), argumentant d'un principe de prägnantz Gestaltique qui descend directement de l'idée d'économie de représentation de Mach.

Techniquement, les théories MEI s'inspirent de la notion de complexité de Kolmogorov, pour laquelle «simple» est synonyme de «compressible», et proposent de considérer que le meilleur modèle parmi un ensemble de modèles possibles est celui qui permet de construire la plus *courte* description des données. Les descriptions doivent être *exactes* (ou complètes, sans perte) pour pouvoir être comparables, ce qui, comme nous le verrons, conduit à nouveau à des énergies à deux termes. Toutefois, à la différence des théories variationnelles et bayesiennes, les théories MDL et MML aboutissent à des énergies *sans paramètre*. Le principe de codage minimal se présente en effet comme un mécanisme systématique permettant de dériver des critères de sélection de modèle dans lesquels *in fine* aucune variable n'est laissée libre, ajustable par l'utilisateur.

Les théories MML et MDL partent d'un paradigme commun et aboutissent à des résultats proches. Il existe cependant de profonds désaccords entre les deux courants, comme on peut par exemple le constater à travers les discussions parues dans le numéro spécial du Computer Journal de 1999 consacré à

[8] Pour compléter l'historique, signalons que c'est Akaike (1973) qui a clairement mis en évidence les limites de la méthode du maximum de vraisemblance et le problème de fond que pose la comparaison de modèles d'ordres différents. Il a proposé ce que l'on appelle le critère AIC (pour «An Information Criterion» et non «Akaike Information Criterion») pour choisir entre des modèles d'ordres différents. Par la suite, Schwarz (1978, la même année que le premier article MDL) a proposé le critère BIC (Bayesian Information Criterion). AIC et BIC ne s'appuient cependant pas sur des principes de codage minimal. Nous verrons toutefois que BIC aboutit à des termes de complexité de modèle relativement similaires (asymptotiquement) à ceux obtenus par MML et MDL.

[9] Toutefois, comme le fait remarquer Cosma Shalizi dans sa critique internet (www.cscs.umich.edu/∼crshalizi/reviews) du livre de Rissanen (1989), Rissanen ne relève pas la filiation intellectuelle de ses idées avec les vieilles idées empirico-critiques d'Ernst Mach ou de Karl Pearson («he displays no awareness of this line of intellectual descent»).

la complexité de Kolmogorov. Outre un certain nombre de points relativement techniques, un point central de débat concerne la philosophie même de l'inférence et l'adhérence ou non à l'interprétation bayesienne des probabilités. Grosso-modo, les deux approches visent à quantifier la complexité d'un modèle afin de pouvoir choisir un modèle *unique* parmi un ensemble de modèles de complexités différentes. Mais alors que Wallace accepte que la «complexité» puisse dépendre de facteurs subjectifs et donc accepte d'intégrer une loi a priori dans le processus de qualification de la complexité des modèles, Rissanen, lui, le refuse totalement. Les deux approches conduisent à des critères sans paramètre - relativement proches d'ailleurs - mais qui véhiculent deux idées fondamentalement différentes : soit l'idée, profondément bayesienne, qu'une fois que l'on a convenablement modélisé l'ensemble des paramètres subjectifs d'un problème alors la solution doit être unique et donc relever d'un critère non paramétrique, ou bien l'idée qu'il existe une unique solution «objective» ou «universelle» à chaque problème de modélisation, que cette solution est systématiquement celle qui nous intéresse, et que l'on doit donc la rechercher par le biais d'un critère non paramétrique obtenu par un mécanisme qui ne relève d'aucun choix arbitraire.

Ces divergences de fond sont souvent mal connues et/ou mal comprises en analyse d'images. Plus généralement, comme le signale Lanterman (2000, p.2), on constate une «prépondérance de mauvaises interprétations et de mauvaises compréhensions [à propos des théories MML et MDL] dans la littérature de sciences appliquées» («preponderance of misinterpretations and misunderstandings in the applied litterature»). En analyse d'images, les méthodes d'inférence s'appuyant sur des critères de codage minimal sont généralement baptisées méthodes MDL et présentées comme des applications de la théorie de Rissanen. Nous avons relevé deux discours principaux à propos du principe «MDL», autour desquels tendent à se rassembler la plupart des auteurs.

Le premier discours suit globalement les revendications de Rissanen et présente le principe comme un outil d'inférence «objectif», principalement en raison du fait qu'il conduit à des critères non paramétriques. Nous argumenterons ci-dessous qu'une telle position est intenable au moins dans le domaine de l'*interprétation* d'images et plus particulièrement pour les questions de partitionnement pour lesquelles *il n'existe objectivement pas de solution «universelle»*.

Le deuxième discours consiste grosso-modo à dire que les critères «MDL» sont des équivalents informationnels de critères de MAP. C'est vrai sous certaines hypothèses, mais qui ne sont pas celles autour desquelles se développent les théories MML et MDL de Wallace et Rissanen. Dans leur cadre d'origine, MDL et MML s'écartent significativement des méthodes de MAP. En outre, tenir MDL et MAP pour équivalents suppose que l'on accepte l'interprétation bayesienne des probabilités, ce qui est contradictoire avec l'idée d'inférence «objective» par MDL.

Vu les difficultés conceptuelles que soulèvent les théories d'inférence par codage minimal, il nous semble judicieux de leur consacrer un peu plus d'attention.

Complexité de Kolmogorov

Les théories MEI s'inspirent de l'idée de *complexité algorithmique* ou *complexité de Kolmogorov* d'une séquence de chiffres binaires. Cette notion de complexité, et d'autres notion proches, ont été introduites au milieu des années 60, indépendamment par Andreï Kolmogorov, Ray Solomonoff et Gregory Chaitin, voir par exemple (Cover et Thomas 1991; Li et Vitanyi 1993; Solomonoff 1997a; Delahaye 1999). Kolmogorov et Chaitin s'intéressaient aux fondements de la théorie des probabilités et à la théorie de l'information. Solomonoff, lui, s'intéressait à l'IA, à la prédiction et à l'induction.

La complexité de Kolmogorov $\mathcal{K}_U(S)$ d'une suite finie de chiffres binaires est définie comme la *longueur du plus petit programme* qui, fourni à une Machine de Turing Universelle (MTU) U, produit la séquence S puis provoque l'arrêt de la machine. $\mathcal{K}_U$ possède d'importantes propriétés et est liée à de nombreux problèmes fondamentaux concernant la nature de l'aléatoire, les indécidables en mathématiques, etc. Ces problèmes fascinants dépassent cependant le cadre de notre propos et nous renvoyons le lecteur intéressé aux références citées, ainsi qu'au site internet très fourni de Gregory Chaitin[10] et à celui du Computer

[10] http ://www.cs.auckland.ac.nz/CDMTCS/chaitin/

Learning Research Centre de Londres[11] qui compte, entre autres, Ray Solomonoff, Chris Wallace et Jorma Rissanen (notons qu'Andreï Kolmogorov est décédé en 1987).

Un point essentiel est que $\mathcal{K}(S)$ est «universelle», au même sens technique que le sont certaines Machines de Turing : Si A et B sont deux MTU, pour toute séquence S

$$\mathcal{K}_A(S) \quad \leq \quad \mathcal{K}_B(S) + C_{A,B} \tag{6.6}$$

où $C_{A,B}$ est une constante *qui ne dépend pas de* S et qui s'interprète comme la longueur d'un programme permettant à A d'émuler le comportement de B. Il est alors usuel d'omettre la machine utilisée et d'écrire simplement $\mathcal{K}(S)$ (nous reviendrons ultérieurement sur l'importance du terme $C_{A,B}$).

Dans ce cadre, une séquence binaire S est «simple», ou peu aléatoire, si elle est fortement compressible, i.e. si $\mathcal{K}(S)$ est faible par rapport à la taille de S ($|S| - \mathcal{K}(S)$ est appelé la déficience aléatoire). Intuitivement, cela signifie que l'on peut décrire S de manière substantiellement plus courte que par simple énumération de ses chiffres. Par exemple l'instruction «imprimer 10^{10} fois 1» est très courte mais produit une très longue séquence, qui est donc très «simple». Une longue séquence de chiffres de π est également très «simple» au sens de Kolmogorov car un algorithme court permet de la produire. À l'inverse, une séquence que l'on ne peut pas - ou presque - compresser est très «complexe» ou aléatoire[12]. Il est alors remarquable que la règle est l'aléatoire : très peu de séquences sont compressibles, c'est-à-dire comportent des formes, apparentes - comme pour $10^{10} \times 1$ - ou cachées - comme pour les 10^{10} premiers chiffres binaire de π, qui sont *en apparence* totalement aléatoires. Notons avec Barbara Burke Hubbard (1995) que «cette définition du mot aléatoire diffère de celle, traditionnelle, donnée par la théorie des probabilités. Selon cette théorie, un procédé peut être aléatoire, mais il n'existe pas de signaux aléatoires ; si l'on jette un dé à dix faces dix fois, la séquence 3,8,5,9,10,4,2,7,6,8 est aussi peu probable que les séquences 2,2,2,2,2,2,2,2,2,2 ou 1,2,3,4,5,6,7,8,9,10. Toutefois nous sommes moins surpris lorsque nous obtenons la première séquence que lorsque nous obtenons les deux dernières». Pour expliquer cette surprise, Kolmogorov a en quelque sorte adopté un point de vue inverse du point de vue classique. Dans la position classique, on considère *un* système donné (modélisé de manière probabiliste) et l'on s'intéresse aux probabilités des différentes issues possibles du système. Kolmogorov a considéré *une* issue donnée et s'est s'intéressé aux différents systèmes possibles qui ont pu la produire. Cet angle d'attaque correspond bien à une optique inductive (il s'agit du même genre de renversement de point de vue que celui opéré quand on passe d'un test d'hypothèse par maximum de vraisemblance à une qualification de modèles par leur probabilité a posteriori).

$\mathcal{K}(S)$ est cependant incalculable, et ce dans un sens très fort[13].

Longueur de message ou de description

L'idée de Wallace et Rissanen est alors la suivante : intuitivement, la capacité à compresser un paquet d'information dépend de la capacité à en *exploiter* les régularités. Si l'on considère alors que l'objectif d'une modélisation est d'*exhiber* les régularités présentes dans des observations, alors un modéle réalise d'autant mieux cet objectif que si nous l'*exploitons au mieux* pour décrire les observations nous obtenons une description courte.

Or, si la complexité «universelle» $\mathcal{K}(S)$ est incalculable - car tous les «modèles» possibles et inimaginables sont susceptibles d'être incorporés dans le programme minimal - il n'en est pas de même quand on restreint l'ensemble des modèles envisagés pour décrire S (voir Wallace et Dowe (1999) et Vitanyi et Li (2000) pour des discussions détaillées sur les liens entre MEI et complexité de Kolmogorov).

[11] http ://www.clrc.rhul.ac.uk/

[12] Le terme historique de *complexité* pour désigner la quantité $\mathcal{K}(S)$ est relativement trompeur. $\mathcal{K}(S)$ correspond plutôt à une notion d'aléatoire qu'à une notion de «complexité».

[13] Il n'existe pas d'algorithme capable de trouver pour toute séquence S le programme minimal permettant de produire S, ni même de trouver $\mathcal{K}(S)$ ou une bonne approximation de $\mathcal{K}(S)$. Une démonstration élégante - plus élégante que la démonstration historique de Kolmogorov - est due à Ragnar Nohre, voir par exemple (Rissanen 1999, p.37).

Considérons donc des observations I et une famille $\mathcal{M}$ de modèles en compétition pour modéliser I. Supposons que l'on se restreigne, pour obtenir une description courte de I, à n'utiliser qu'un des modèles de $\mathcal{M}$. Comme en général un de ces modèles ne décrit pas parfaitement I, si l'on devait écrire un programme permettant à un ordinateur de reconstruire I bit pour bit, il faudrait alors décrire dans le programme :

1) Le modèle M de $\mathcal{M}$ que l'on utilise et
2) i. Si M est un modèle déterministe, les erreurs commises par M.
   ii. Si M est un modèle stochastique, les événements effectivement observés.

Dotons-nous donc d'un *langage* de description $\mathfrak{L}_1$ des modèles de $\mathcal{M}$, c'est-à-dire d'une procédure de codage de chaque $M \in \mathcal{M}$. Associons alors au langage $\mathfrak{L}_1$ la fonction $\mathcal{L}_1$ qui à tout $M \in \mathcal{M}$ fait correspondre le nombre de bits qu'il faut pour décrire M par $\mathfrak{L}_1$. $\mathcal{L}_1(M)$ est appelée la *longueur de description* de M (sous entendu : dans le langage $\mathfrak{L}_1$).

De la même manière, considérons un langage $\mathfrak{L}_2$ permettant de décrire, pour un modèle déterministe les corrections que l'on est susceptible de lui apporter, et pour un modèle stochastique, les événements possibles en supposant qu'ils obéissent à la distribution du modèle. À $\mathfrak{L}_2$ est également associé une fonction longueur-de-description $\mathcal{L}_2$ d'une image connaissant un modèle.

Étant donnée une image I et un modèle M de cette image, on peut alors faire une description *complète* (sans perte, exacte) de I en deux étapes : description de M par $\mathfrak{L}_1$, puis description des «différences» entre I et M par $\mathfrak{L}_2$. Cette description emploie alors

$$\mathcal{L}_{1,2}(I\langle M\rangle) = \mathcal{L}_1(M) + \mathcal{L}_2(I|M) \qquad \text{bits.} \tag{6.7}$$

$\mathcal{L}_{1,2}(I\langle M\rangle)$ est appelée la **longueur de message (MML)** ou **longueur de description (MDL)** de I selon une procédure de codage en deux étapes (two parts code[14]) utilisant le modèle M[15].

Le principe est alors de sélectionner le modèle $M \in \mathcal{M}$ qui minimise la longueur de message, i.e.

$$M^*(I\langle \mathcal{M}, \mathfrak{L}_1, \mathfrak{L}_2\rangle) \quad = \quad \underset{M\in\mathcal{M}}{argmin} \quad \mathcal{L}_{1,2}(I\langle M\rangle). \tag{6.8}$$

On voit qu'une fois une famille de modèles fixée, il y a deux paramètres : les deux langages de description $\mathfrak{L}_1$ et $\mathfrak{L}_2$.

$\mathfrak{L}_1$ et $\mathfrak{L}_2$ doivent correspondre à des schéma de codage *effectifs* : les éléments qu'ils décrivent doivent pouvoir être décodés sans ambiguïté, sans quoi l'approche n'a plus de sens. Donc, dans l'approche MEI, toutes les fonctions énergétiques sur les objets ne sont pas envisageables, leurs valeurs doivent correspondre à des longueurs de codes d'une procédure de description valide[16].

Pour une classe d'objets donnée, il y a bien sûr de nombreuses procédures valides de codage. La philosophie MEI est alors de sélectionner, parmi tous les langages possibles pour l'inférence, ceux qui sont *les plus économiques* (nous verrons ci-dessous dans quel sens), d'où l'appellation d'inférence par codage minimal. Il y a donc un *double principe d'économie* dans l'approche MEI : le meilleur modèle est celui qui donne lieu à la description la plus économique des données lorsqu'on utilise les langages de description les plus économiques. Et de ces deux principes, **le principe de fond est celui qui consiste à sélectionner les langages minimaux** et non celui qui consiste à rechercher un message de longueur

[14] La méthode d'inférence par un code à deux termes est la plus connue de la théorie MEI et, à notre connaissance, la seule à avoir été appliquée à l'image. Les théories MEI proposent cependant d'autres méthodes d'inférence qui s'appuient sur des procédures de description différentes.

[15] Remarque concernant la notation : Nous verrons dans un instant que la longueur de message $\mathcal{L}\left(I\langle M\rangle\right)$ s'apparente à la probabilité a posteriori $P_P(M|I)$ des approches bayesiennes. Cependant, dans l'approche par codage minimal, une notation du type $\mathcal{L}\left(M|I\right)$ est inapropriée car la description est celle de I et non de M. Par ailleurs, la notation sans référence au modèle $\mathcal{L}\left(I\right)$, qui est couramment utilisée, nous semble également inapropriée. La notation $\mathcal{L}(I\langle M\rangle)$ que nous utilisons doit être lue : «longueur de description de I par un code à deux termes employant le modèle M».

[16] Plus précisément, si l'on choisit d'utiliser des codes préfixés (ou auto-délimités), les longueurs $\mathcal{L}$ des mot-codes associés aux éléments de U doivent vérifier l'inégalité de Kraft : $\sum_{u\in U} 2^{-\mathcal{L}(u)} \leq 1$, cf. (Cover et Thomas 1991, p.82).

minimale dans un langage donné, cette dernière minimisation correspondant au principe de base de toutes les approches énergétiques (principe de comparaison).

Insistons sur le fait que si l'on ne comparait pas des messages écrits dans deux langages $\mathfrak{L}_1$ et $\mathfrak{L}_2$ *minimaux*, l'approche serait inconsistante. Un message à deux termes de type 6.7 met en effet sur un même pied d'équivalence la longueur de description du modèle et celle de ses erreurs. Il faut donc les évaluer en un point où elles sont comparables.

Recherche de langages optimaux

Quels sont donc les langages les plus économiques ? Dans quel sens ? Ces langages existent-ils ? Sont-ils uniques ? Bien sûr la question de l'unicité ne porte pas sur le code lui même mais sur la distribution d'énergie de codage qu'il induit (on peut arbitrairement permuter des mots de même longueur du dictionnaire).

Le problème d'exhiber un code optimal se pose différemment selon les quantités auxquelles on a affaire.

Code de Shannon-Fano. La théorie de l'information de Shannon (1948) s'applique à des événements d'un espace *discret*. Si l'on doit coder une séquence d'événements (e.g. d'images) ayant été produits de manière indépendante selon une loi P, alors le code préfixé de Shannon-Fano est optimal *en moyenne* : il minimise l'espérance de coût de codage d'une observation (il atteint asymptotiquement l'entropie de la distribution P). Ce code attribue à une issue x un mot-code de longueur $\lceil -\log P(x) \rceil$. Notons toutefois qu'un code de Huffmann est - entre autres - également optimal, et n'affecte pas aux événements les mêmes longueurs de mots-code que le code de Shannon-Fano[17]...

Modèles stochastiques paramétriques. Wallace et Rissanen se placent dans le cadre de l'inférence statistique de modèle, où les modèles de $\mathcal{M}$ sont probabilistes. Chaque $M \in \mathcal{M}$ est une distribution déterminée par un vecteur de paramètres Θ de taille *variable*.

Ils commencent par décomposer la famille $\mathcal{M}$ en différentes *classes* $\mathcal{M}_i$ de modèles (e.g. de Gauss, binomial...), chacune composée de modèles d'ordre fixe. Ils supposent que l'ordinateur (ou la personne) à qui l'on doit décrire les données connait les différentes classes envisagées et «sait» à quoi elles correspondent, i.e. qu'il sait ce qu'est un modèle «gaussien» ou «binomial». Nous n'avons pas à lui expliquer. Il suffit donc de lui spécifier quelque chose comme «nous utilisons une loi binomiale pour décrire les données» en entête du message, ce qui est supposé requérir une énergie constante - indépendante de la classe. Ce terme n'intervient donc pas dans l'inférence, ce qui revient à dire que la complexité «conceptuelle» d'une classe de modèle n'intervient pas.

La classe de modèle étant connue, la première partie du message 6.7 correspond alors à la description de l'instance particulière qui est considérée au sein de la classe, à travers le codage des valeurs de ses paramètres.

Le modèle étant connu, la deuxième partie du message 6.7 a pour principe de décrire I en supposant qu'elle obéit au modèle M et ce de la manière la plus économique possible. Pour tester une hypothèse M, il est donc naturel d'adopter un langage $\mathfrak{L}_2$ *qui serait optimal pour décrire I si I suivait effectivement la loi M*. Évidemment, les observations dont on dispose sont toujours quantifiées, les mesures sont faites à une précision limitée. On peut donc utiliser un code de Shannon-Fano pour décrire les observations[18]. La longueur de message s'écrit alors (en négligeant l'arrondi entier du code de Shannon-Fano et en supprimant les termes constants passés en entête) :

$$\mathcal{L}(\Theta) - \log P_v(I|M_\Theta). \tag{6.9}$$

[17] Il n'existe pas d'expression analytique pour ces longueurs : elles sont obtenues par un procédé algorithmique.

[18] Il s'agit d'une simplification car les modèles que l'on considère fournissent habituellement des valeurs réelles. Toutefois la quantification des *réalisations* des modèles à une précision donnée, celle des observations dont on dispose, introduit un terme constant dans le message, qui n'a pas d'influence sur l'optimum. Nous verrons que la quantification des *paramètres* des modèles pose, elle, beaucoup plus de problèmes.

Interprétation bayesienne de l'inférence par codage minimal. On voit alors que

i) ***si*** l'espace des paramètres du modèle considéré est également quantifié et

ii) ***si*** l'on définit une distribution de probabilité a priori sur cet espace, $P_p(\Theta)$, supposée connue du receveur, i.e. relevant d'un accord préalable entre les deux parties,

alors en employant à nouveau un code de Shannon-Fano pour décrire les paramètres du modèle, la longueur de message devient

$$-\log P_p(\Theta) - \log P_v(I|M_\Theta) \tag{6.10}$$

et le critère MEI coïncide avec un critère de MAP (voir l'équation 6.4), à ceci près qu'il s'applique uniquement à des modèles dont les paramètres sont soit discrets par nature, soit discrétisés par un procédé fixé au préalable.

Réciproquement, la spécification d'un dictionnaire valide pour coder les points d'un espace *discret* induit une loi de probabilité : l'inégalité de Kraft assure que la loi est normalisable. En vertu de cette équivalence pour les espaces discrets, Zhu et Yuille (1996) parlent de critère bayesien/MDL, identifiant les deux approches.

Notons avec Leclerc (1989a) que la théorie de l'information de Shannon établit un pont entre les domaines probabilistes et déterministes. Il peut parfois être malaisé d'envisager certains types d'objets comme le résultat d'un processus stochastique et difficile de définir une loi de probabilité a priori sur l'ensemble des réalisations possibles. Les envisager comme des quantités déterministes et en imaginer un schéma de description effectif peut alors être beaucoup plus naturel et définit implicitement une loi a priori sur ces objets (un point important est que l'on n'a pas à normaliser explicitement la loi). La force de l'équivalence entre longueurs de codes et probabilités est de pouvoir à convenance passer d'un registre à l'autre : spécifier des distributions de probabilité pour certaines quantités (e.g. bruit gaussien) et des schéma de description déterministes pour d'autres (e.g. géométrie de régions).

MDL et MML. Toutefois, la version MAP de l'«inférence par codage minimal» qui vient d'être décrite *ne correspond pas* aux théories développées par Rissanen et Wallace. MML et MDL s'intéressent en effet à des modèles dont les paramètres sont *continus* et s'écartent radicalement de l'inférence bayesienne dans leur traitement. Par ailleurs, comme nous l'avons évoqué ci-dessus, il y a une divergence de fond entre Wallace et Rissanen. Les défenseurs de l'approche MML interprètent effectivement la probabilité P_p ci-dessus comme «représentant des croyances, connaissances ou ignorances a priori sur les paramètres»[19] du modèle, alors que Rissanen soutient que «dans le principe MDL pour l'inférence statistique il n'y a pas besoin de s'embarasser des interprétations bayesiennes de la probabilité a priori sur les paramètres»[20]. L'objectif déclaré de Rissanen est d'aboutir à des méthodes «objectives» de sélection de modèle. Il refuse catégoriquement l'emploi de lois a priori «subjectives» et l'interprétation bayesienne des probabilités. Une grande part de ses travaux concerne la construction de lois a priori dites «non-informatives», «non-subjectives» ou «universelles» pour coder les paramètres de la manière la plus «neutre» possible. Jorma Rissanen est profondément anti-bayesien et l'association entre MDL et bayesianisme faite par Zhu et Yuille est une sorte d'oxymore...

Avant de discuter de questions d'a priori, voyons la manière dont MDL et MML traitent la quantification de paramètres réels.

Si les paramètres des modèles sont des réels - ce qui correspond à la plupart des problèmes standard de modélisation de données statistiques par des modèles probabilistes - alors plutôt que d'utiliser un schéma de quantification fixe, convenu au préalable comme en 6.10, MDL et MML proposent d'optimiser la quantification des paramètres. La démarche est la suivante.

[19] «representing prior beliefs, or knowledge, or ignorance about the parameters»(Baxter et Oliver 1994). Oliver et Baxter (1994) proposent une discussion détaillée sur les relations entre l'approche MML et le formalisme bayesien.

[20] «in the MDL principle for statistical inference there is no need for the awkward bayesian interpretations of the meaning of the prior probability on the parameters» (Rissanen 1993). N.d.T. : Les différentes traductions possibles d'*Awkward* sont 1) gauche, maladroit, disgracieux, balourd. 2) fâcheux, embarrassant, gênant. 3) incommode, peu maniable.

On ne peut pas coder un nombre réel avec une précision arbitraire. Le nombre de bits requis croit avec la précision. Par ailleurs, dans un message en deux parties de type 6.9, si les paramètres Θ du modèle sont transmis avec une certaine précision p, alors, pour que le code soit correctement décodable par le receveur, il faut transmettre les observations codées selon le modèle approximatif, de précision p. Les paramètres réels «idéaux», qui minimisent la deuxième partie du message, sont ceux qui maximisent la vraisemblance. Cependant, plus la précision de leur description diminue, plus la longueur de description des données en suivant le modèle quantifié s'écarte du maximum de vraisemblance et augmente. En conclusion, le premier terme de 6.9 croit en fonction de la précision de description des paramètres du modèle, alors que le second décroit (en tous cas autour du maximum de vraisemblance).

Dans la philosophie MEI, qui s'inspire de l'idée de complexité de Kolmogorov, on recherche le coût d'un programme minimal permettant de décrire *un* échantillon, celui *dont on dispose*. Dans cet esprit, il est alors naturel de rechercher la précision de description des paramètres qui minimise la longueur totale de message. On recherche donc un schéma de quantification Q des paramètres qui minimise

$$\mathcal{L}(\lceil\Theta\rceil_Q) - \log Pr(I|M_{\lceil\Theta\rceil_Q}) \tag{6.11}$$

où $\lceil\Theta\rceil_Q$ désigne le vecteur de paramètres Θ quantifiés selon une grille Q de quantification[21].

MML et MDL traitent différemment ce problème de quantification optimale. En outre, chaque courant a donné lieu à plusieurs variantes, à des versions pour le cas univarié et d'autres pour les distributions multivariées, etc. Il serait beaucoup trop long de détailler ces développements ici. Nous n'en donnerons-nous qu'un aperçu qualitatif et renvoyons le lecteur à l'excellente synthèse de Lanterman (2000).

Supposons que Θ est un vecteur de $\mathbb{R}^K$.

Il faut en premier lieu choisir une manière de quantifier l'espace. Rissanen (1983) s'appuie sur une quantification par des parallélépipèdes de taille q^K. Wallace et Freeman (1987) proposent d'utiliser une *grille de quantification optimale* (optimal quantizing lattice) dont la forme n'est connue que jusqu'à la dimension 3, mais pour laquelle ils n'ont besoin que du volume des cellules, dont on connait de bonnes approximations (Conway et Sloane 1993).

La première partie du message code alors la valeur tronquée des paramètres, c'est-à-dire la cellule de la grille de quantification. Les approches MML supposent que l'on dispose d'une distribution a priori sur l'espace des paramètres et utilisent naturellement le code de Shannon-Fano associé aux masses de probabilité contenues dans les cellules.

Rissanen, lui, propose d'indexer les cellules par des entiers, en spirale, et de coder l'indice d'une cellule par un code «universel» pour les entiers qui attribue à $n \in \mathbb{N}$ la longueur

$$\begin{aligned} \mathcal{L}^0(n) &= \log(2.865064) + \log^*(n) \qquad (6.12) \\ \text{où} \quad \log^*(n) &= \log(n) + \log\log(n) + \log\log\log(n) + \dots \quad \text{tant que les termes sont positifs.} \end{aligned}$$

Ces longueurs de code sont obtenues en résolvant un problème de minimax : si l'on doit coder par un code préfixé une séquence d'entiers ayant été produits selon une loi P inconnue de support *infini* et *croissante*, utiliser les longueurs $\mathcal{L}^0$ minimise la pire redondance moyenne sur $\mathbb{N}$ par rapport au code idéal de Shannon-Fano associé à P[22]. Si P est à support fini, Rissanen (1983) montre que la solution du même problème de minimax est une loi uniforme sur le support de P, ce qui correspond bien à notre intuition d'une loi non-informative (lire cependant Bernardo (1997) à ce sujet). Voir Cover et Thomas (1991) pour une description du schéma de codage effectif auquel correspondent les longueurs 6.12.

Avec ces schéma de quantification et de codage des paramètres, les auteurs cherchent alors à optimiser le pas de quantification q. Il emploient à nouveau différentes stratégies. Rissanen cherche à minimiser la

[21] Pour que le message soit effectivement décodable, il faudrait également décrire la grille Q utilisée. Celle-ci ne pourrait évidemment être décrite qu'à une précision finie, qu'il faudrait avoir codée au préalable, et ainsi de suite, ce qui conduit à un message récursif. Wallace et Rissanen s'affranchissent de ce problème d'une manière que nous ne détaillerons pas. Voir par exemple Wallace et Freeman (1987, p.245).

[22] Remarquons que le code 6.12 fait l'hypothèse que les entiers les plus faibles sont les plus probables. Toute permutation de $\mathbb{N}$ avant d'appliquer $\mathcal{L}^0$ fournit également un code «universel».

longueur de message 6.11 en considérant l'erreur maximale commise sur la cellule correspondant au maximum de vraisemblance. De leur côté, Wallace et Freeman (1987) font remarquer que la manière de paramétrer un modèle est arbitraire. Ils proposent alors de minimiser l'*espérance* de longueur de message sur toutes les reparamétrisations possibles du modèle (principe MML, voir Oliver et Hand (1994) pour une bonne introduction).

Les résolutions s'appuient alors sur des développements limités à l'ordre deux autour soit du maximum de vraisemblance, soit du MAP. Modulo différentes hypothèses, les auteurs dérivent des approximations des pas optimaux de quantification et des coûts de codage des paramètres[23].

Supposons que l'échantillon considéré comporte N observations et qu'on le modélise comme ayant été produit par N lois i.i.d. à K paramètres. Les approches MDL et MML aboutissent alors toutes deux à un terme de premier ordre pour l'expression de la complexité du modèle qui s'exprime :

$$\mathcal{L}\left(\lceil\Theta\rceil_{q^*}\right) \quad \simeq \quad \frac{K}{2}\log N. \tag{6.13}$$

Ce terme de complexité correspond également au terme principal obtenu par la théorie BIC de Schwarz (1978), et ce par une approche totalement différente, cf. (Lanterman 2000).

En segmentation, cette expression a par exemple été utilisée par Kanungo et al. (1995) pour représenter le coût de description de coefficients de polynômes de régression, et par Lee (2000) pour représenter celui de paramètres de distributions exponentielles. Pateux (1998) l'emploie également pour des questions de compression orientée régions. Nous l'avons récemment utilisée avec Franck Taillandier et Rachid Deriche (Taillandier, Guigues et Deriche 2003) pour représenter le coût de description de plans de régression dans le cadre d'une application de segmentation de MNE en facettes planes dont nous parlerons au chapitre 9.

Notons donc que les expressions obtenues par les théories MEI pour quantifier la complexité d'un modèle ne peuvent pas être interprétées comme des lois a priori, c'est-à-dire fixées *avant* d'avoir pris connaissance des données. Le codage du modèle est optimisé pour un échantillon donné, ce qui conduit en particulier à des expressions qui dépendent de la taille de l'échantillon à décrire (le critère AIC, par exemple, ne dépend que du nombre de paramètres).

Optimiser le codage pour chaque échantillon n'interdit cependant pas d'introduire une loi a priori «subjective» sur l'espace des paramètres. Comme nous l'avons déja signalé, les courants MML et MDL divergent toutefois radicalement sur ce point. Exposons notre point de vue à ce propos.

A priori et «objectivité»

Revenons tout d'abord sur la complexité de Kolmogorov.

On pourrait penser que si cette quantité était calculable, elle serait susceptible de conduire à un outil d'inférence «universel». C'est ce que soutiennent Li et Vitanyi (1993, 2000) qui argumentent grosso-modo que, quel que soit le problème d'induction, tout ce que nous avons à faire pour en trouver une bonne solution est d'exhiber un programme suffisament proche du programme minimal. Nous avons toutefois quelques doutes sur la pertinence de ce procédé d'inférence «universel».

Remarquons en premier lieu qu'un programme de longueur proche de $\mathcal{K}(S)$ n'est pas nécessairement «qualitativement proche» du programme qui réalise $\mathcal{K}(S)$. Parvenir à compresser ne serait-ce que d'un bit supplémentaire peut nécessiter de changer radicalement de principe de codage, soit de changer radicalement d'«explication» concernant l'origine des observations (i.e. concernant le processus algorithmique qui est supposé les avoir produites).

En outre, non seulement $\mathcal{K}(S)$ est beaucoup moins «universelle» qu'on ne le croît parfois, mais aussi et surtout, même si elle l'était, la solution qu'elle fournit n'est bien souvent pas celle que l'on recherche.

[23] En théorie, la quantification des paramètres augmente la valeur du deuxième terme, cependant l'augmentation est minime - nous l'avons estimée à moins de 1% pour des schema de codage gaussien par une méthode MML. Dans les utilisation pratiques des longueurs de message, on utilise finalement la valeur exacte du maximum de vraisemblance pour le deuxième terme. L'idée d'optimiser le compromis entre les deux termes ne sert finalement qu'à dériver une approximation de la complexité du modèle.

L'«universalité» bien relative de la complexité algorithmique a été clairement exprimée par Ray Solomonoff (dans le cadre des problèmes de prédiction) : «le choix de la machine [de Turing] de référence et celui du mode de représentation [des observations] peuvent être vus comme des modifications des probabilités a priori affectées aux séquences d'événements [...] Ceci rend certainement le résultat «subjectif». Si l'on quantifie l'objectivité, on peut mécaniquement réduire le choix de la machine et de la représentation à certaines valeurs universelles «par défaut» - mais il y a un compromis entre objectivité et précision. Pour obtenir la meilleure extrapolation, nous devons utiliser toute l'information disponible, et une bonne part de cette information peut être subjective»[24].

Quand il parle de l'influence du choix de la machine de référence, Solomonoff évoque la constante que nous avons notée $C_{A,B}$ dans la relation 6.6 ci-dessus, qui est donc loin d'être anodine[25]. Solomonoff insiste également sur l'influence du mode de représentation. L'idée est simple : le programme minimal doit permettre à une machine donnée de produire *bit pour bit* une séquence de codes qui représente certaines informations (mesures, observations). Or, c'est *nous* qui décidons de la manière de coder l'information sous forme d'une séquence binaire. Ce mode de représentation n'est évidemment pas unique et selon celui que nous choisissons, le programme permettant à une machine *donnée* de produire la séquence peut être plus ou moins long. En gros, le programme doit intégrer les instructions nécessaires à la *présentation* du résultat sous la forme que nous avons spécifiée. Les influences du choix de la machine de référence et du mode de représentation sont évidemment intimement liées. Pour donner un exemple concret, imaginons que l'on s'intéresse à la séquence binaire S résultant de la compression jpeg d'une image simple. Selon que la machine de référence intègre ou non une compression/décompression jpeg matérielle, les programmes minimaux qui permettent de produire S peuvent être radicalement différents.

Quand bien même $\mathcal{K}(S)$ était «universelle», il nous semble que dans bon nombre de problèmes d'inférence, la solution recherchée n'est pas une *explication* «universelle» mais bien au contraire une *interprétation* très «subjective». Illustrons-le pour le problème de vision avec un exemple proche de celui de la compression jpeg qui vient d'être donné.

La transformation du boulanger est souvent présentée comme l'archétype des transformations chaotiques, cf. e.g. (Arnoux et Chemla 1992). Notons la B. Quelques itérations de B appliquées à une image la rendent en apparence totalement aléatoire. Prenons une image I présentant un carré blanc sur fond noir. La décrire comme nous venons de le faire, comme présentant «un carré blanc sur fond noir» est certainement proche de la plus courte description possible de I. Considérons maintenant $J = B^{1000}(I)$: la transformation de I par 1000 itérations de B (en espérant que B^{1000} ne soit pas l'identité). La plus courte façon de décrire J est certainement de la décrire comme nous venons de le faire : décrire I, décrire B (la description de la transformation du boulanger tient en une ligne) et dire que ce que l'on observe est $B^{1000}(I)$. *Nous, humains*, ne percevons cependant pas J - d'apparence totalement aléatoire - comme une permutation d'une figure simple par un procédé mixing simple[26]...

Il nous semble que l'objectif de la vision artificielle est de reproduire l'inférence que *nous* faisons quand nous observons une image et non pas l'inférence «optimale» qu'un système extrêmement plus «perspicace» que nous serait capable de faire.

MDL et MML proposent précisément de résoudre le double problème posé par la complexité de Kolmogorov - trop forte «universalité» et incalculabilité - en restreignant l'espace des modèles possibles. Toutefois, comme nous l'avons vu, l'école MML juge légitime de considérer une distribution a priori «subjective» sur l'espace des modèles envisagés alors que l'école MDL s'y refuse catégoriquement.

[24] «Both choice of reference machine and mode of representation can be viewed as modifications of the apriori probabilities assigned to sequences of events [...] This certainly makes the result «subjective». If we value objectivity, we can routinely reduce the choice of machine and representation to certain universal «default» values - but there is a tradeoff between objectivity and accuracy. To obtain the best extrapolation, we must use wathever information is available, and much of this information may be subjective»(Solomonoff 1997b).

[25] Notons en particulier que $C_{A,B}$ peut être prépondérant face à $\mathcal{K}_A(S)$ et $\mathcal{K}_B(S)$ même pour une séquence S très longue : il suffit que cette séquence soit très «simple». $C_{A,B}$ n'est véritablement négligeable qu'à la limite.

[26] ou quasi-mixing, cf. (Cornu et Smorlarz 2001).

Notons alors $\mathcal{U}$ l'ensemble de tous les modèles possibles de l'«Univers» (grosso-modo ceux qu'est susceptible d'exploiter le programme minimal) et $\mathcal{M} \subset \mathcal{U}$ les seuls modèles que l'on s'autorise à inférer. Contraindre la solution à appartenir à $\mathcal{M}$ peut alors être vu comme l'attribution *a priori* d'une probabilité *nulle* à tous les modèles hors de $\mathcal{M}$. Dans ce cas, pourquoi se restreindre à une distribution uniforme[27] à l'intérieur de l'ensemble $\mathcal{M}$, c'est-à-dire finalement a une logique binaire sur l'ensemble $\mathcal{U}$: modèle possible ou impossible ? Nous pensons donc que, dès lors que l'on se restreint à un ensemble limité de modèles, l'idée d'inférence «non-subjective» comporte une sorte de contradiction interne.

C'est pourquoi, comme D'Agostini (1999), «contrairement à ceux qui essaient de trouver des priors objectifs qui donneraient à la théorie bayesienne un plus noble statut d'objectivité, nous préférons revendiquer explicitement la naturalité et la nécessité de priors subjectifs»[28].

Comme nous l'avons vu, c'est généralement ainsi que l'entendent les chercheurs en traitement d'image quand ils assimilent inférence par codage minimal et inférence bayesienne. Leclerc insiste ainsi sur le fait que l'inférence dépend du choix d'un langage et donc d'une loi a priori : «l'information a priori sur le monde et le capteur est incorporée dans le langage utilisé pour décrire le monde et le capteur, et le processus d'inférence consiste à trouver la description la plus simple (i.e. courte) dans ce langage qui reproduit exactement les images données»[29]. C'est encore plus clair chez Zhu et Yuille (1995) qui suggèrent d'apprendre les lois a priori sur des images réelles. Notons que tel que l'expriment ces auteurs, les lois a priori portent sur «le monde», «le capteur» ou l'ensemble des images «naturelles» d'une certaine classe, i.e. les images d'un certain type de scène faites par un certain type de système imageur. Mais outre les connaissances (ou croyances) a priori concernant l'*observé*, il nous semble qu'un point essentiel en segmentation est de modéliser les a priori de l'*observateur*, la manière dont il interprète les observations. En adoptant le point de vue positiviste radical d'Ernst Mach, on pourrait d'ailleurs dire que tout se ramène à cela...

Critères de codage pour la segmentation

Voyons finalement la structure typique des critères de codage considérés pour la segmentation d'images. Considérons pour cela une famille $\mathcal{S}$ de modèles d'images par morceaux, c'est-à-dire s'appuyant sur une partition du domaine de l'image. Appelons un élément de $\mathcal{S}$ une *segmentation*. Une segmentation sur un domaine $\mathcal{D}$ est un couple $S = (P, V)$, où $P = \{R_1, \ldots, R_N\}$ est une partition de $\mathcal{D}$, et $V = \{V_1, \ldots, V_N)$ est un ensemble de modèles dédiés à la description du contenu de l'image au sein de chaque région. Nous dirons que P est la composante *géometrique* du modèle et que V est sa composante *radiométrique*.

La description $I\langle S\rangle$ d'une image I utilisant un modèle S d'image segmentée peut alors consister en

1. Décrire la part géométrique de la segmentation, i.e. la partition P, éventuellement en utilisant un *modèle géométrique* G de partition provenant d'une classe $\mathcal{G}$.
2. Décrire, *connaissant* P, le contenu de l'image I dans chaque région $R \in P$ en tirant parti d'un *modèle radiométrique* V_R de ce contenu pris dans un ensemble $\mathcal{R}$.

Un tel message utilise deux modèles *emboîtés*. La modélisation de l'image comme une union de sous-images s'appuie sur une partition de l'espace, dont la description peut elle-même s'appuyer sur une modélisation géométrique, typiquement qui emploie des descriptions par morceaux de courbes paramétriques (segments, portions de côniques, splines...), c'est à dire une segmentation 1D.

[27] ou à une loi impropre «non-informative», ce qui d'après Bernardo (1997) n'existe pas.

[28] «In conclusion, contrary to those who try to find objective priors which would give the bayesian theory a nobler status of objectivity, I prefer to state explicitly the naturalness and necessity of subjective priors» (D'Agostini 1999, p.30).

[29] «prior information about the world and image sensor is incorporated in the language used to describe the world and sensor, and the inference process is to find the simplest (i.e. shortest) description in the language that exactly reproduces the given images.»(Leclerc 1989b)

Dans ce cadre, la longueur de message générale donnée par l'expression 6.7 devient

$$\begin{array}{rcccc} \mathcal{L}\left(I\langle S\rangle\right) & = & \mathcal{L}_{\text{géo}}(I\langle S\rangle) & + & \mathcal{L}_{\text{radio}|\text{géo}}(I\langle S\rangle) \\ & = & \mathcal{L}\left(P\langle G\rangle\right) & + & \mathcal{L}\left(I\langle P,V\rangle\right) \\ & = & \left(\mathcal{L}\left(G\right)+\mathcal{L}\left(P|G\right)\right) & + & \left(\mathcal{L}\left(V|P\right)+\mathcal{L}\left(I|P,V\right)\right). \end{array} \tag{6.14}$$

Dans la dernière expression, les deux termes $\mathcal{L}(G)$ et $\mathcal{L}(V|P)$ représentent les complexités des deux modèles utilisés - géométrique et radiométrique - tandis que les deux autres termes $\mathcal{L}(P|G)$ et $\mathcal{L}(I|V,P)$ sont respectivement des termes d'attache i) de la partition «observée» à son modèle géométrique, et ii) de la sous-image observée sur chaque région à son modèle radiométrique.

La plupart des auteurs (Leclerc 1989a; Kanungo et al. 1995; Lee 1998; Fuchs et Le Men 1999; Lee 2000) ne considèrent pas de modèle géométrique à proprement parler mais considèrent, en suivant l'idée initiale de Leclerc, une description exacte des frontières de la partition par des séquences de codes de Freeman. Le codage de Freeman est approximativement proportionnel au nombre d'éléments de frontière discrète de la partition, i.e. à la longueur totale de frontières exprimée en pixels. On retrouve donc un terme de régularisation à la Mumford-Shah ou Ising/Potts. Toutefois, la théorie MDL indique que la valeur énergétique à accorder à chaque pixel de frontière est de l'ordre de $\log k$ bits pour des frontières codées en k-connexité.

Nous verrons ci-dessous qu'employer des modes de description des formes s'appuyant sur des approximations polygonales permet de significativement améliorer la qualité des segmentations obtenues, indiquant bien que le choix du langage de description - ou de la classe de modèles envisagée, ce qui peut s'assimiler à un choix de langage - n'est pas anodin.

Pour ce qui est des modèles radiométriques, l'image est classiquement modélisée comme une surface polynomiale d'ordre faible - constante, plane, et au plus quadratique - noyée dans un bruit non corrélé spatialement, de type gaussien additif pour les images classiques, Poissonien pour les images numériques, ou multiplicatif - bruit de speckle - pour les images acquises par un système actif à lumière cohérente (e.g. images radar).

Leclerc a montré qu'un modèle *local* d'ordre deux - i.e. la modélisation de l'image au sein de *chaque pixel* par une surface d'ordre deux - était suffisant pour modéliser la plupart de images usuelles par des surfaces lisses.

Le modèle de «bruit» peut avoir des paramètres constants sur toute l'image ou variables entre les différentes régions (paramètres constants par région). Dans ce dernier cas, ce que l'on modélise est la variabilité de *texture* des différentes régions et non un «bruit» introduit par le capteur, qui ne saurait s'adapter aux formes de l'image. Nous verrons ci-dessous qu'un modèle gaussien i.i.d est suffisant pour discriminer bon nombre de textures, pour peu que l'on s'autorise à pousser l'algorithme bien au delà du domaine du bruit, explicitement dans le domaine textural.

Toujours au sujet du modèle de bruit, notons l'approche originale de Lee (1998) qui considère des bruits corrélés spatialement. Il est toutefois dans ce cas difficile de faire la part entre une variation spatiale des caractéristiques du bruit et une variation spatiale de l'image elle-même. Lee ne démontre la validité d'une telle approche que dans le cas d'images sous-jacentes constantes par morceaux. Les variations lentes sont alors imputables sans ambiguïté aux oscillations du bruit.

Nous n'avons pas encore parlé du coût de description du modèle radiométrique de chaque région. Comme nous l'avons vu, plusieurs auteurs emploient l'expression asymptotique 6.13. D'autres auteurs, comme Fuchs (2001), considèrent un coût de codage constant par région, ce qui correspond à un codage équiprobable des paramètres suivant une grille de quantification fixe. Voyant qu'attribuer des coûts de codage «naturels» par rapport à la quantification intrinsèque des données ne donnait pas de résultats satisfaisants *pour son application spécifique*, Frank Fuchs a proposé de faire varier ces coûts «fixes» pour collecter des segmentations à plusieurs degrés de détail. Nous verrons que sa proposition se présente comme un cas particulier de notre méthode de segmentation multi-échelles.

Pour notre part, dans nos premières expériences de segmentation par codage, nous avons cherché à raffiner l'expression 6.13 pour le cas particulier de modèles gaussiens multivariés. Nous avons pour cela

étendu au cas multivarié une approche MML proposée pour le cas 1D par Oliver et Hand (1994), aboutissant à une expression de l'espérance de longueur de message invariante par reparamétrisation de la loi normale nD et à une approximation à l'ordre 2 du coût de codage des paramètres minimisant cette espérance. L'approche a fourni certains résultats théoriques intéressants, comme une expression de la complexité du modèle décroissante avec l'isotropie de la distribution : plus la matrice de variance/covariance possède des valeurs propres proches plus on peut compresser la description. Toutefois, il s'est avéré que ces corrections n'influaient que très peu sur les résultats de segmentation. Nous avons par la suite abandonné cette «bataille» sur les bits de modèle qui s'est avérée peu fructeuse en pratique.

6.1.5 Exploration de l'espace des partitions

Voyons maintenant la manière dont on traite en pratique les problèmes d'optimisation qui viennent d'être décrits.

Optimisation combinatoire

Toutes les formulations par minimisation d'énergie qui possèdent un certain intérêt conduisent à des problèmes d'optimisation difficiles.

Ainsi, un des problèmes les plus simples de partitionnement de données, le partitionnement en k groupes G_i, avec k fixé, de manière à minimiser $\sum_i \sum_{(x,y)\in G_i} d(x,y)^2$, où d est une distance entre observations, est connu de longue date comme étant NP-difficile, cf. (Garey et Johnson 1979).

Dans le domaine du partitionnement d'images, Boykov et al. (1999) ont récemment montré que le problème de labellisation pixellaire, sous sa forme régularisée la plus simple utilisant un potentiel de Potts, conduit également à un problème NP-difficile.

Certains problèmes admettent des méthodes de résolution particulièrement adaptées. Nous pensons ici à l'approximation du critère de Coupe Normalisée de Shi et Malik (2000) par une technique de théorie spectrale des graphes (cf. 5.7.2 p. 105).

Cependant, quelle que soit l'origine de la formulation énergétique, on a habituellement recours à une méthode générique d'optimisation combinatoire (Toet 1995). Une telle méthode fonctionne globalement par exploration d'un sous-espace de l'espace S des solutions, en utilisant un ensemble de déplacements élémentaires dans cet espace (Charon et al. 1996). Ces déplacements sont le plus souvent *locaux*, ce qui conduit - selon l'espace de recherche considéré et/ou le type de représentation de la solution considéré - à des opérations de type :

- Fusion et division de régions quand le voisinage est considéré dans le treillis des partitions.
- Faible déformation de frontières de régions ou croissance de régions quand le voisinage est considéré dans le domaine spatial.
- Relabellisations ponctuelles quand le voisinage est envisagé dans l'espace des images de label.

On trouve alors encore une fois deux types d'algorithmes de recherche : déterministes et stochastiques. Par ailleurs, certains algorithmes trouvent un optimum local, d'autres - beaucoup plus coûteux - tentent d'exhiber le minimum global.

Recherche d'optimum local

On se trouve dans un minimum local quand tous les mouvements autorisés provoquent une croissance d'énergie. La méthode la plus générale pour trouver un minimum local est d'employer un algorithme de *descente de gradient* ou son analogue discret qui consiste à choisir à chaque pas le mouvement qui diminue le plus l'énergie.

En variationnel, le fait d'être dans un minimum local s'exprime par l'équation différentielle d'Euler-Lagrange, qui conduit après discrétisation à des schéma d'adaptation locale de la solution.

Introduite pour les approches par MRF, la méthode des Modes Conditionnels Itérés (Iterated Conditional Modes, ICM) permet d'atteindre un minimum local (Besag 1986). Elle consiste à fixer n variables

(généralement toutes sauf une, e.g. le label d'un pixel) et à rechercher la valeur des variables libres qui minimise l'énergie conditionnée par les valeurs des variables bloquées. Ce mode conditionnel peut souvent être calculé explicitement. Ce procédé est itéré jusqu'à convergence.

Quand on considère des mécanismes de fusion de régions, deux notions - plus ou moins locales - de minimalité sont définies à travers la notion de *normalité* (Koepfler et al. 1994) :

Définition 17 *Si I est une image définie sur un domaine $\mathcal{D}$ et E est une énergie sur $\mathbb{P}(\mathcal{D})$[30], une partition P est déclarée :*

- ***normale*** *quand*
$$\forall Q \in \mathbb{P}(\mathcal{D}) \quad P < Q \Rightarrow \mathsf{E}(P) < \mathsf{E}(Q).$$
- ***2-normale*** *quand*
$$\forall Q \in \mathbb{P}(\mathcal{D}) \quad P \prec Q \Rightarrow \mathsf{E}(P) < \mathsf{E}(Q).$$

Ces définitions possèdent bien sûr des analogues dans le treillis des partitions en composantes connexes. La normalité est le pendant énergétique de la notion de maximalité qui est définie dans le cadre des prédicats (cf. 5.2.2 p. 72). Notons bien que la normalité ne correspond pas à l'optimalité globale de la partition : l'ordre sur les partitions n'est que partiel, une partition normale est seulement d'énergie plus faible que ses sous-partitions ce qui ne préjuge en rien de sa valeur sur l'ensemble du treillis.

De nombreux auteurs, e.g. (Koepfler et al. 1994; Ballester et al. 1994; Kanungo et al. 1995; Fuchs et Le Men 1999; Lee 2000), argumentent en faveur de l'utilisation d'une méthode de descente associée à un mécanisme de fusion de régions. Les algorithmes de *Fusion de Régions par Descente (FRD)* sont initialisés à partir d'une segmentation S_1, supposée être une sur-segmentation de la solution optimale. S_1 peut être la sur-segmentation absolue de l'image en pixels individuels ou une sur-segmentation obtenue par pré-traitement, par exemple par croissance de régions à partir de germes (seeded region growing (Lee 2000)), ou par calcul d'une ligne de partage des eaux (Haris et al. 1998). Si $N = |S_1|$, une FRD produit alors une suite de N segmentations emboitées. S_{k+1} est définie comme la partition d'énergie minimale couvrant S_k. La chaîne de partitions $S_1 \prec \cdots \prec S_N$ est donc un chemin de plus grande pente depuis S_1 dans le treillis des partitions. Encore une fois, ces auteurs se placent dans l'espace des partitions en parties connexes et ne considèrent donc à chaque étape que la fusion de paires de régions voisines.

Il y a deux variantes pour sélectionner la solution finale. La plupart des auteurs arêtent la fusion quand plus aucune union ne permet de diminuer l'énergie, c'est-à-dire à la première segmentation 2-normale rencontrée. De son côté, Lee (1998, 2000) propose de continuer la fusion même si l'énergie croît à certaines étapes, jusqu'à ce que l'image entière soit fusionnée. Il sélectionne alors la meilleure segmentation parmi les N rencontrées au cours de l'exploration. La solution est également seulement 2-normale mais ce n'est pas obligatoirement la première de la chaîne $S_1 \prec \cdots \prec S_N$.

Certains algorithmes alternent différents types de mouvements, comme par exemple l'algorithme de *compétition de régions* de Zhu et Yuille (1996) qui considèrent une croissance à partir de germes, des déplacements de la frontière entre deux régions voisines en «compétition», et la fusion de régions voisines, le tout guidé par la minimisation d'une énergie globale dite bayesienne/MDL (comme nous l'avons vu, il serait plus juste de dire bayesienne/MML).

Boykov et al. (1999) considèrent des mouvements de grande amplitude dans l'espace des images de label (relabellisation d'un nombre arbitraire de pixels de label identique, échange de labels entre deux groupes de pixels de labels identiques de taille arbitraire) et proposent des algorithmes s'appuyant sur la théorie des graphes (flot maximal) pour trouver le mouvement qui diminue le plus l'énergie à chaque étape. Une stratégie de descente est alors appliquée. Vu l'amplitude des mouvements considérés, cet algorithme peut être qualifié de semi-local. Les auteurs prouvent en outre que l'énergie de la solution calculée est au pire double de celle de l'optimum global (ce n'est pas en soi une borne très fine, mais elle assure que l'algorithme ne peut pas faire «n'importe quoi»). La méthode est cependant relativement gourmande en temps de calcul.

[30] La dépendance à l'image est implicite.

Recherche d'optimum global

À coté des méthodes approchées, on trouve quelques méthodes d'optimisation globale. Les méthodes les plus utilisées en segmentation s'organisent en deux grandes familles : la famille des méthodes de recuit (simulé, déterministe, par champ moyen) et la famille des méthodes dites de non-convexité graduelle (Blake et Zisserman 1987) ou de continuation (Leclerc 1989a). La méthode d'optimisation multi-échelles proposée ci-dessous peut s'envisager comme une méthode de continuation. Décrivons donc plus en détail ce type de méthode et renvoyons le lecteur à la littérature concernant les autres méthodes.

Si $\mathsf{E}(x)$ est une énergie non convexe à minimiser, une méthode de continuation plonge E dans une famille d'énergies $\mathsf{E}(x,s)$ telle que

i) $\mathsf{E}(x,0) = \mathsf{E}(x)$,

ii) Pour s suffisament grand $\mathsf{E}(x,s)$ soit convexe, i.e. admette un unique minimum local et global,

iii) $\mathsf{E}(x,s)$ soit continue selon s.

L'idée est de simuler un lissage progressif de l'énergie au cours duquel les minima locaux disparaissent progressivement en se fusionnant à d'autres, et ce jusqu'à ce qu'il n'en subsiste qu'un. Au cours de ce processus, l'optimum global de l'énergie d'origine est lissé et se déplace selon une trajectoire continue. Aussi, si l'on inverse le processus et que l'on suit la ligne la plus profonde, on doit pouvoir remonter la trajectoire qu'il a suivi.

Le principe est alors de partir de s_0 suffisament grand pour que $\mathsf{E}(x,s_0)$ soit convexe. On minimise cette énergie par une méthode de descente de gradient, qui est rapide et qui, comme l'énergie est convexe fournit son minimum global. On diminue alors légèrement s, l'énergie est légèrement remodelée, et on applique une nouvelle descente à partir de la solution qui vient d'être trouvée. On diminue à nouveau légèrement s et on réitère. Sous certaines hypothèses sur E et son plongement, on prouve que si s diminue de manière *continue*, on obtient une convergence presque sûre vers l'optimum global de E quand s tend vers 0 (la convergence en probabilité provient des problèmes de bifurcations). En pratique, on procède bien entendu par paliers sur s, typiquement selon une suite géométrique $s_0 r^n$, $0 < r < 1$. Si les pas sont suffisamment rapprochés (r proche de 1), on a toutes les chances de ne pas perdre le minimum global.

Comme Leclerc l'a remarqué, cette technique peut être considérée comme une poursuite de l'optimum global dans une représentation scale-space de l'énergie, de l'échelle la plus grossière à la plus fine. Nous utiliserons ci-dessous une stratégie analogue, à la différence près - qui est énorme - que l'objectif sera de calculer *toutes les solutions*, à toutes les échelles, d'une énergie dépendant d'un paramètre se comportant comme un paramètre d'échelle. Nous verrons en outre que notre «continuation» sera véritablement continue.

Illustrons ce qui vient d'être dit en voyant l'énergie utilisée par Leclerc et le plongement graduellement convexe qu'il considère.

En règle générale, les termes qui posent problème dans une énergie de segmentation sont ceux qui portent sur les discontinuités du modèle - i.e. précisément sur la recherche du bon découpage (typiquement, nous avons vu que si l'on fixe la partition, l'optimisation devient souvent triviale, e.g. correspond à un maximum de vraisemblance d'un problème simple).

Leclerc propose de minimiser une énergie de partition issue du formalisme MDL qui comporte un terme C de description géométrique de la partition par un codage de Freeman des chaînes de pixels, ce qui revient à un terme de longueur de frontière discrète. Leclerc écrit alors cette longueur de frontière

$$\mathsf{C}(L) = \frac{1}{2} \sum_{x \leftrightarrow y} 1 - \delta(L(x) - L(y)) \tag{6.15}$$

où L est l'image d'étiquettes de la partition et δ est le delta de Kronecker. Il remarque ensuite que $\forall x \in \mathbb{R}$

$$\lim_{s \to 0} e^{-x^2/s} = \delta(x)$$

et remplace alors l'énergie $\mathsf{C}(L)$ d'origine par la famille d'énergies en fonction de s

$$\mathsf{C}(L,s) = \frac{1}{2}\sum_{x\leftrightarrow y} 1 - e^{-(L(x)-L(y))^2/s}.$$

Les autres termes énergétiques considérés par Leclerc sont des termes d'attache aux données pour lesquels l'unique minimum global - d'énergie nulle - est l'image elle-même. Quand s n'est pas infini mais suffisament grand, le terme de régularisation $\mathsf{C}(L,s)$ est encore assez peu influent et le minimum global est toujours l'image d'origine. Leclerc initialise la continuation à un tel point.

6.2 Modélisation multi-échelles d'images par morceaux

Ce tour d'horizon des principales approches énergétiques du problème de *partitionnement d'image* étant fait, voyons comment nous proposons d'aborder l'analyse *multi-échelles* d'une image en régions.

L'approche proposée comporte deux aspects : un aspect *énergétique* et un aspect *structurel*. Voyons tout d'abord l'aspect énergétique, qui fait directement suite à ce qui vient d'être discuté.

6.2.1 Le cadre énergétique

Énergies affines

Nous venons de voir que les énergies considérées pour le partitionnement d'image prennent globalement la forme d'une opposition entre deux termes antagonistes : un terme d'attache aux données et un terme de régularisation. Nous considérons les énergies qui s'écrivent de manière relativement générale comme une combinaison de deux termes énergétiques :

Définition 18 *Nous appelons* ***énergie affine*** *une énergie qui s'écrit*

$$\mathsf{E}_\lambda(P) \quad = \quad \lambda \mathsf{C}(P) + \mathsf{D}(P) \tag{6.16}$$

où P est une partition du domaine des images (la dépendance à l'image est implicite) et $\lambda \in \mathbb{R}^+$ est un paramètre qui règle l'importance relative des deux termes C et D.

À P fixée, nous envisagerons ci-dessous $\mathsf{E}_\lambda(P)$ comme une fonction de λ, d'où la terminologie d'énergie affine.

Remarque : En raison du sens physique qu'ils revêtent, C, D et λ sont a priori supposés prendre des valeurs réelles *positives*. Nous verrons ci-dessous que certaines de ces contraintes peuvent être relâchées - au moins mathématiquement.

Le tableau 6.1 résume les différentes interprétations possibles d'une énergie affine en fonction du formalisme dont elle est issue. Examinons de plus près le rôle du paramètre λ.

Dans un cadre variationnel, λ est simplement un paramètre de pondération du terme de régularisation de la fonctionnelle. Pour la fonctionnelle constante par morceaux de Mumford-Shah, Koepfler, Lopez, et Morel (1994) l'appellent également **paramètre d'échelle** en remarquant qu'il joue globalement sur la finesse des partitions qui minimisent l'énergie. Comme nous l'avons vu, si $\lambda = 0$, la solution est la sur-partition absolue (l'inf du treillis). À l'autre extrême, pour λ assez grand, la solution est la partition sans frontière, la sous-partition absolue (le sup du treillis). Nous montrerons ci-dessous que dans un cadre hiérarchique, le paramètre λ de cette fonctionnelle se comporte rigoureusement comme un paramètre d'échelle.

Dans un cadre bayesien, λ joue sur la forme de la distribution a priori. La probabilité a priori d'une partition P s'exprime en effet

$$P_p(P) = \frac{e^{-\lambda \mathsf{C}(P)}}{\sum\limits_{Q\in\mathbb{P}(X)} e^{-\lambda \mathsf{C}(Q)}}$$

	Formalisme		
Terme	Variationnel	Bayesien (MAP)	Codage Minimal
D	Attache aux données	Potentiel de vraisemblance	Longueur de description des erreurs de modèle (déterministe) / des observations sachant le modèle (stochastique)
C	Régularisation	Potentiel a priori	Longueur de description du modèle
λ	Poids de la régularisation / paramètre d'échelle	Paramètre	Facteur de redondance dans le codage d'une des deux parties du message
$\lambda\mathsf{C}+\mathsf{D}$	Énergie	Potentiel a posteriori	Longueur de message

TAB. 6.1 – Différentes interprétations d'une énergie affine.

où le dénominateur normalise à 1 la mesure sur $\mathbb{P}(\mathcal{D})$[31]. Quand $\lambda = 0$, toutes les partitions sont a priori équiprobables, de probabilité $1/|\mathbb{P}(\mathcal{D})|$, et la solution du MAP s'identifie à celle du maximum de vraisemblance (*argmin* de D). Augmenter λ a pour effet de progressivement accentuer les maxima de la distribution a priori. À la limite, quand λ tend vers $+\infty$, seuls les maxima globaux de C possèdent une probabilité non nulle. Si l'on note $\mathcal{C}^*$ l'ensemble des maxima globaux de C, $P_p(P)$ vaut $1/|\mathcal{C}^*|$ si $P \in \mathcal{C}^*$ et est nulle sinon. La solution du MAP correspond alors à sélectionner la partition de $\mathcal{C}^*$ qui maximise la vraisemblance. Les deux cas extrêmes $\lambda = 0$ et $\lambda = +\infty$ correspondent donc à deux solutions de maximum de vraisemblance mais sur deux espaces différents.

L'interprétation du paramètre λ dans le cadre du codage minimal mérite de plus amples commentaires. Les théories d'inférence par codage minimal (MML/MDL) sont en effet sensées être sans paramètre.

Si $\mathsf{C}(P)$ représente le coût de codage d'un modèle de l'image alors $\lambda\mathsf{C}(P)$ peut s'interpréter comme la longueur d'un message dans lequel :

- Si $\lambda < 1$: le modèle est transmis *avec perte*, soit incomplètement.
- Si $\lambda = 1$: le modèle est transmis complètement et à moindre coût.
- Si $\lambda > 1$: le modèle est transmis *avec redondance*.

Dans l'esprit MEI d'un message à deux termes, transmettre le modèle incomplètement est absurde : l'image n'est plus reconstructible. Remarquons toutefois que les minima de l'énergie sont inchangés par multiplication de l'énergie par une constante positive. Minimiser $\lambda\mathsf{C}+\mathsf{D}$ avec $0 < \lambda < 1$ est alors équivalent à minimiser $\mathsf{C}+\frac{1}{\lambda}\mathsf{D}$, soit $\mathsf{C}+\mu\mathsf{D}$ avec $\mu > 1$. On peut donc toujours considérer que le message ne comporte aucune perte mais que l'un des deux termes, C ou D, est transmis avec un certain facteur de redondance. λ est donc intrinsèquement un terme de redondance *relative* de C par rapport à D.

Ce facteur de redondance peut être interprété dans le cadre d'une *transmission par un canal bruité*. La théorie de l'inférence de modèle par codage minimal suppose en effet que le message est reçu par le décodeur exactement tel qu'il a été construit. Imaginons au contraire qu'il doive passer par un canal bruité avant décodage. Si l'on souhaite que le message puisse être correctement décodé par le receveur, il faut alors introduire des *codes correcteurs d'erreurs* dans le message. Ces codes allongent le message mais augmentent la probabilité que le message soit correctement reconstruit. On peut donc interpréter λ comme correspondant à l'introduction *inégale* de tels codes entre les deux parties du message - modèle et erreurs. λ mesure ainsi l'importance relative que le constructeur du message accorde à la bonne réception de chacune d'elles.

[31] on l'appelle également «facteur de partition» mais ça n'a rien a voir avec le fait qu'ici la probabilisation porte sur l'espace des partitions du domaine $\mathcal{D}$.

Un problème d'optimisation sous contrainte

Voyons maintenant une deuxième interprétation de λ, comme le paramètre de Lagrange associé à un problème d'optimisation sous contrainte. Reprenons pour cela le problème de modélisation à sa base et proposons une formalisation du principe de parcimonie différente de celle proposée par les approches MEI.

Notre traduction du principe de parcimonie. Notons à nouveau $\mathcal{M}$ l'ensemble des modèles d'images envisagés et dotons-nous de deux fonctions réelles, D et C qui mesurent respectivement la «distance» entre un modèle et une image, et la «complexité» d'un modèle.

Que nous dit alors le principe de parcimonie ? «le modèle le plus simple est le meilleur» ? Il ne faut évidemment pas prendre cette expression au pied de la lettre, sans quoi le meilleur modèle serait toujours l'image vide (dans notre cas la sous-partition absolue, n'ayant qu'une seule classe), et ce quelles que soient les observations. En fait, la règle de sélection par simplicité intervient *après* celle par fidélité. L'expression exacte du principe de parcimonie - telle que l'entendent par exemple les physiciens - est en effet : «de deux modèles qui décrivent *également bien* les observations, il faut retenir le plus simple».

Donc, *pour une précision de description ε donnée*, ce principe nous enjoint à sélectionner le modèle le plus simple. Formellement, on aboutit à un problème d'optimisation sous contrainte : il s'agit de trouver le modèle solution de

$$\min_{M\in\mathcal{M}} \mathsf{C}(M) \quad \text{sous la contrainte} \quad \mathsf{D}(I,M)=\varepsilon. \tag{6.17}$$

où ε contrôle l'erreur d'approximation.

Notons bien que deux modèles de précisions différentes sont a priori *incomparables*, ils sont *qualitativement* différents. Par exemple, le modèle gravitationnel d'Einstein n'est pas meilleur *dans l'absolu* que celui de Newton. Si l'on juge que la gravitation Newtonienne est plus «simple» (techniquement) que la gravitation relativiste alors tant que les mesures à décrire ne sont pas assez précises pour que l'on ressente les effets relativistes, d'après le principe de parcimonie, la gravitation de Newton *doit* être utilisée : elle est aussi exacte à cette précision et à la fois plus simple que son homologue relativiste.

Évidemment, si un modèle M tel que $\mathsf{D}(I,M)<\varepsilon$ est plus simple que tous les modèles tels que $\mathsf{D}(I,M')\leq\varepsilon$ on doit également le retenir. On peut donc transformer l'égalité en inégalité dans le terme de contrainte qui intervient dans 6.17. On aboutit alors à un problème que nous noterons S_ε :

$$\mathsf{S}_\varepsilon(I) \quad : \qquad \min_{M\in\mathcal{M}} \mathsf{C}(M) \quad \text{sous la contrainte} \quad \mathsf{D}(I,M)\leq\varepsilon. \tag{6.18}$$

Transformer l'égalité en inégalité permet aussi de gérer le cas où l'espace des modèles est discret : si $\mathcal{M}$ est un espace discret, I une image arbitraire et ε une borne arbitraire de distance, il est en effet très improbable qu'il existe dans $\mathcal{M}$ un modèle qui soit *strictement* à distance ε de I.

À S_ε est associé un problème **dual**, que nous notons $\mathsf{S}^\star_\gamma$, pour lequel les rôles de C et D sont échangés :

$$\mathsf{S}^\star_\gamma(I) \quad : \qquad \min_{M\in\mathcal{M}} \mathsf{D}(M,I) \quad \text{sous la contrainte} \quad \mathsf{C}(M)\leq\gamma. \tag{6.19}$$

$\mathsf{S}^\star_\gamma$ traduit un principe de parcimonie dual : «de deux modèles également complexes, il faut retenir celui qui rend le mieux compte des observations». γ borne ici la complexité du modèle recherché.

Transposition lagrangienne. La technique classique pour résoudre un problème d'optimisation sous contrainte s'appuie sur la minimisation de son Lagrangien. Le Lagrangien associé à S_ε est

$$\mathsf{E}_\lambda(M,I) \quad \triangleq \quad \lambda\mathsf{D}(M,I)+\mathsf{C}(M) \tag{6.20}$$

où λ est appelé le paramètre de Lagrange de l'énergie E_λ.

Pour une image I donnée, notons M_λ le modèle qui minimise le Lagrangien E_λ de paramètre λ.

Sous certaines hypothèses qui seront vérifiées pour nos modèles, on sait alors que M_λ correspond également à la solution de S_ε quand λ est réglé de telle manière que $\mathsf{D}(\mathsf{M}_\lambda, I) = \varepsilon$ (ou approximativement dans le cas discret, ce point sera précisé ci-dessous).

Le Lagrangien associé au problème dual échange naturellement les rôles de D et C :

$$\mathsf{E}^\star_\mu(M, I) \quad \triangleq \quad \mu\mathsf{C}(M) + \mathsf{D}(M, I) \tag{6.21}$$

et son minimum $\mathsf{M}^\star_\mu$ correspond à la solution du problème $\mathsf{S}^\star_\gamma$ quand $\mathsf{C}(\mathsf{M}^\star_\mu) = \gamma$.

Les énergies affines peuvent donc se concevoir comme les expressions lagrangiennes associées à des problèmes d'optimisation sous contrainte. Chaque valeur du poids relatif entre les deux termes de l'énergie est alors associée à une valeur de contrainte soit sur la complexité du modèle, soit sur sa précision.

Nous nous intéresserons principalement par la suite à l'expression affine 6.16, qui est analogue à 6.21 et correspond donc au principe de parcimonie dual 6.19, dans lequel la contrainte porte sur la complexité du modèle. Sous certaines hypothèses sur les énergies, nous aboutirons cependant à un algorithme qui produit des solutions identiques pour les deux problèmes duaux.

Compression avec pertes - théorie débit/distorsion

On distingue classiquement - depuis Claude Shannon (1948, 1959) - deux types de problèmes de compression d'information numérique : la compression *sans* perte et la compression *avec* pertes. L'articulation entre les deux domaines a un nom célèbre : l'entropie. Si l'on souhaite compresser un signal en dessous de son entropie, on encourt nécessairement des pertes d'information. La question est alors de compresser au plus en occasionnant le minimum de pertes[32].

On traduit cette question par deux problèmes d'optimisation sous contrainte qui sont formellement identiques aux problèmes 6.18 et 6.19. Si la fonction C représente le volume de l'image compressée - appelé *débit* quand on se place dans le cadre d'une communication par un canal numérique - et si la fonction D mesure l'importance des pertes - la *distorsion* de l'image d'origine qu'entraîne la compression - alors

i) Le problème 6.18 traduit le soucis d'obtenir la meilleure compression (minimisation de débit) dans une limite de dégradation donnée (distorsion bornée). Régler cette limite de dégradation correspond par exemple à régler le facteur de «qualité» dans un algorithme jpeg.

ii) Le problème 6.19 traduit le soucis de minimiser la dégradation quand on spécifie un taux de compression (débit borné). Ce cas correspond par exemple à la volonté de transmettre un aperçu le plus fidèle possible par un canal de capacité limitée (transmission par un réseau, e.g. par internet), ce qui revient à spécifier un temps maximal de transmission si l'on connait la capacité du canal.

La théorie *débit/distorsion* (*rate/distortion*) s'intéresse alors à la relation entre débit et distorsion dans les problèmes de compression avec perte. La théorie a été introduite par Shannon (1959), qui a aussitôt dégagé les résultats principaux. Le résultat central de Shannon est relativement intuitif. Il stipule qu'en général, débit et distorsion sont reliés par une relation *monotone*. Autrement dit, augmenter le taux de compression ne peut qu'augmenter les pertes, voir par exemple Clarke (1995, pp.12-14) ou Guillois (1996, pp.22-25) pour un présentation simple. Plus précisément, Shannon s'est intéressé, pour une perte maximale donnée, au meilleur taux de compression moyen qu'il est possible obtenir pour les représentants d'une classe donnée de signaux aléatoires. Autrement dit, Shannon s'est intéressé aux limites intrinsèques de compressibilité de certains types de signaux (voir le chapitre 13 de (Cover et Thomas 1991) qui développe en détail la théorie débit/distorsion)[33]. Contrairement aux problématiques de débit/distorsion

[32] Une question centrale est également de sélectionner l'«information» que l'on s'autorise à perdre et celle que l'on souhaite conserver. Cette question n'est pas abordée par la théorie classique de l'information, à laquelle on a souvent reproché de traiter de l'«information» abstraite, c'est-à-dire des symboles, des codes, et non de ce qu'ils représentent, voir par exemple (Atlan 1982).

[33] Mentionnons à ce propos que Donoho (2000) a récemment mis en évidence d'importants parallèles entre la théorie débit/distorsion de Shannon et la théorie de l'approximation de Kolmogorov. Kolmogorov s'est intéressé à l'approximation

«théoriques» de Shannon, le problème de débit/distorsion 6.18 (tout comme son problème dual) est formulé pour *un* élément à comprimer, et pose la question de minimiser le débit, à distorsion bornée, par *un* procédé de compression donné. On dit qu'il s'agit d'un problème de débit/distorsion *opérationnel*, car il correspond à une instance concrète d'un problème général de compression.

Nous aboutirons ci-dessous à un résultat de monotonie entre débit et distorsion pour une large gamme de problèmes opérationnels de compression orientée régions, i.e. s'appuyant sur un codage d'une partition de l'image puis d'un modèle de chacune des sous-images correspondant à chaque région.

Débit/distorsion et codage minimal - Régularisation par complexité

Le point de vue débit/distorsion nous semble particulièrement intéressant vis-à-vis de la théorie de l'inférence par codage minimal (MDL/MML). Cette théorie pose en effet le problème d'inférence comme un problème de compression optimale, mais *sans perte*, contrairement aux problèmes de débit/distorsion qui autorisent une perte contrôlée et recherchent le modèle le plus simple satisfaisant cette contrainte - ce qui nous paraît être la traduction juste du principe de parcimonie (dual). Les approches MDL et MML, théoriquement sans paramètre, proposent en définitive qu'il n'existe qu'une seule inférence pertinente : celle qui réalise le «meilleur» compromis global, c'est-à-dire qui met sur un même pied d'équivalence simplicité et fidélité du modèle. Nous avons pourtant parfois besoin d'un modèle particulièrement simple et nous sommes prêts à payer cette simplicité par une perte de fidélité, ou inversement. Le point de vue débit/distorsion permet d'introduire ce facteur. Nous y voyons finalement un enrichissement de la théorie stricte du codage minimal, trop centrée sur l'idée - à notre avis flouée - qu'il existe une solution «universelle» aux problèmes d'inférence de modèle, qui répond à tous les besoins.

De manière fort intéressante, le paramètre de Lagrange associé au problème sous contrainte peut s'interpréter comme un facteur de redondance dans le codage d'une partie du message. Appliquons alors le principe d'équivalence lagrangienne, mais en sens inverse de ci-dessus, en partant de l'énergie affine :

Minimiser $\mathsf{E}_\lambda = \lambda\mathsf{C}+\mathsf{D}$ pour λ *fixé* (e.g. à 1 pour MDL) est équivalent à minimiser D sous la contrainte $\mathsf{C} \leq \gamma$, où γ est la valeur, dépendante de λ, qui assure que la solution s'identifie au minimum de E_λ. En termes probabilistes, cela revient à maximiser la vraisemblance ($\propto e^{-\mathsf{D}}$) sur l'ensemble

$$\mathcal{P}_\gamma = \{M \in \mathcal{M} \mid e^{-\mathsf{C}(M)} \geq e^{-\gamma}\}$$

des modèles dont la probabilité a priori $P_p(M) \propto e^{-\mathsf{C}(M)}$ est supérieure à une valeur $e^{-\gamma}$ donnée.

De ce point de vue sous contrainte, le critère de MAP apparaît donc comme un critère de maximum de vraisemblance, mais posé sur un sous-ensemble restreint de modèles. La loi a priori agit donc comme un filtre sur les solutions possibles (remarquons que l'ensemble $\mathcal{P}_\gamma$ peut être vu comme une section supérieure de la fonction $P_p(M)$). Sous cet angle, le seuil γ, qui définit l'ensemble des modèles autorisés, semble bien arbitraire et les solutions de MAP apparaîssent comme constituant un *continuum* de solutions depuis le maximum de vraisemblance jusqu'au maximum a priori (nous avons déja parlé page 132 des solutions extrêmes pour $\gamma = 0$ et $\gamma = +\infty$; revoir également la discussion de la page 125 où nous interprétons la délimitation de l'espace $\mathcal{M}$ des modèles comme correspondant à attribuer une probabilité nulle aux modèles hors de $\mathcal{M}$).

---

d'un espace infini $\mathcal{F}$ de fonctions par un ensemble *fini* $\mathcal{D}$ de fonctions (un dictionnaire), qu'il a appelé un *réseau* (net). Si $\mathcal{F}$ est compact pour la norme $||\cdot||$, $\mathcal{D}$ est un ε-réseau si $\sup_{f\in\mathcal{F}} \min_{g\in\mathcal{D}} ||f-g|| \leq \varepsilon$. Kolmogorov définit alors l'ε-entropie $H_\varepsilon(\mathcal{F})$ de $\mathcal{F}$ comme le logarithme du cardinal du plus petit ε-réseau de $\mathcal{F}$. L'approche est similaire à la théorie débit/distorsion de Shannon à la différence que Shannon, en considérant l'espérance de perte, s'intéresse à la distance *moyenne* entre les objets à coder (ou représenter, approximer) et les éléments du dictionnaire et non à la distance *maximale* comme le fait Kolmogorov (le sup dans la définition d'un ε-réseau ci-dessus). Donoho (2000) montre que si l'on identifie ε^2 à la distorsion D, alors quand D tend vers 0, l'ε-entropie de $\mathcal{F}$ se rapproche du pire débit $R(D,X)$ possible pour un signal aléatoire X qui se réalise en probabilité dans $\mathcal{F}$. Formellement, $H_\varepsilon(\mathcal{F}) = \sup\{R(D,X) : P(X \in \mathcal{F}) = 1\}(1+o(1)), \quad D = \varepsilon^2 \to 0$. Autrement dit, les théories de Shannon et de Kolmogorov se rejoignent quand on s'approche de l'entropie, i.e. du domaine de la compression sans pertes. Donoho établit également des liens profonds entre ces théories et la théorie du risque de Wald (minimax) et la théorie Bayésienne.

Nous rejoignons finalement - tout au moins sur la forme - la théorie de la *régularisation par complexité* de Moulin et Liu (2000). Ces auteurs proposent de formaliser toute une gamme de problèmes inverses d'imagerie (essentiellement de restauration) comme des questions de compression optimale avec perte. L'originalité est que le «moteur» d'inférence est un algorithme effectif de compression d'image (codeur jpeg, codeur prédictif...). La distorsion est exprimée via un modèle probabiliste de formation de l'image qui donne lieu à une vraisemblance $P_v(I|M)$. Réglé pour une certaine perte, l'algorithme de compression associe alors à une image I un modèle de complexité $\mathcal{L}(M)$ (volume de l'image compressée). L'inférence consiste alors grosso-modo à trouver les réglages de l'algorithme de compression qui correspondent à la solution du problème de débit/distorsion opérationnel de Lagrangien

$$\mathsf{E}_\lambda(I, M) = \lambda\mathcal{L}(M) - \log P_v(I|M). \tag{6.22}$$

Moulin et Liu prouvent différents résultats théoriques concernant la qualité d'un tel procédé d'inférence (en terme de distance de Kullback-Leibler entre l'estimateur et la solution du problème inverse). Sous diverses conditions complexes, ils montrent en particulier qu'une «bonne» solution théorique ne peut être obtenue qu'en réglant $\lambda > 1$, c'est-à-dire en réglant λ strictement au dessus de la valeur préconisée par l'approche MDL (les hypothèses sur lesquelles reposent ce résultat doivent cependant être examiné avec soin avant d'en tirer quelque conclusion hâtive).

Le principe d'indifférence d'échelle

Même si Moulin et Liu abordent l'inférence par codage sous un angle débit/distorsion, ce qui les conduit à introduire un paramètre de pondération entre les deux termes énergétiques, ils n'envisagent pas de faire varier ce paramètre pour obtenir plusieurs solutions au problème de modélisation. Or, si *chaque* problème d'optimisation sous contrainte - pour une borne de fidélité ou de complexité donnée - peut être dérivé d'un principe de parcimonie, en revanche, ce principe ne permet pas de comparer des modèles de précisions ou de complexités différentes. Nous proposons donc de nous intéresser à la famille de *toutes les solutions* du problème de modélisation, pour toutes les valeurs des contraintes, considérant qu'elles sont toutes potentiellement utiles.

Nous rejoignons ici l'argumentaire du chapitre 3 en faveur d'une approche multi-échelles de la segmentation de bas niveau. Étayons cet argumentaire dans le cadre présent :

Une image $I : \mathcal{D} \to \mathcal{V}$ possède deux «dimensions» : l'espace ($\mathcal{D}$) et la mesure ($\mathcal{V}$). Si l'énergie D mesure l'écart entre les valeurs observées dans une image et les valeurs données par un modèle, alors la précision de modélisation ε qui intervient dans le problème sous contrainte 6.18 peut s'interpréter comme une résolution d'analyse selon l'axe des valeurs de l'image. Par ailleurs, si l'énergie C de «complexité» d'un modèle s'appuyant sur une partition P de $\mathcal{D}$ attribue une complexité d'autant plus forte à un modèle que la partition sur laquelle il s'appuie est fine alors la borne γ sur la complexité de modèle qui apparaît dans le problème 6.19 peut s'interpréter comme une résolution d'analyse spatiale. Or, les deux dimensions $\mathcal{D}$ et $\mathcal{V}$ sont nécessairement quantifiées lors de l'acquisition d'une image et ce *a priori*, c'est-à-dire avant l'opération de mesure et donc indépendamment de son résultat. On peut alors concevoir la segmentation comme un procédé de re-quantification a posteriori, dont le rôle est de produire une description discrète *adaptée* aux données. Dans ce cas, une opération de segmentation ne doit pas à nouveau fixer une résolution d'analyse *a priori* mais bien *rechercher quelles résolutions d'analyse sont pertinentes*, ce qui impose de s'intéresser à l'ensemble des descriptions obtenues pour les différents choix possibles.

Nous proposons donc de rechercher une méthode A qui à une image I associe une famille de partitions $\mathsf{A}(I) = \{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$ indexée par un réel positif λ qui soit «la plus proche possible» - dans un sens qui reste à définir - des minima d'une énergie affine donnée.

Les raisons précises pour lesquelles nous posons le problème en termes de «proximité» apparaîtront clairement ci-dessous. Une première raison, d'ordre pratique, est qu'en général les problèmes d'optimisa-

tion qui interviennent en segmentation sont hors de la classe P. Un algorithme utilisable de segmentation multi-échelles ne peut donc qu'approximer la famille des minima d'une énergie affine.

Notre objectif est donc de parvenir à une méthode d'analyse orientée régions qui produise des descriptions multi-échelles *continues* en échelle, comme le sont les analyses scale-space d'une image. L'approche proposée repose alors sur la *structure* souhaitable pour une analyse multi-échelles ensembliste et la manière dont cette structure peut être *représentée*. Décrivons maintenant ce deuxième aspect de l'approche.

6.2.2 Le cadre structurel

Faisons temporairement abstraction du cadre énergétique qui vient d'être décrit et intéressons nous au résultat recherché, à savoir une famille de partitions $\mathsf{A}(I) = \{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$.

Causalité structurelle et non délocalisation de frontières

Pour que la famille $\mathsf{A}(I)$ puisse être regardée comme une description multi-échelles de I, de paramètre d'échelle λ, nous requérons qu'elle vérifie un *principe de causalité* analogue au principe qui intervient dans les théories scale-space (cf. 3.4.2 p. 32).

Comme nous l'avons vu, le principe de causalité d'échelle consiste à dire qu'une diminution de précision d'analyse ne peut pas créer d'information, i.e. que les «structures» «vues» à grande «échelle» (fort λ) doivent correspondre à certaines «structures» «vues» aux plus petites «échelles».

Intéressons-nous alors à des images définies sur un domaine $\mathcal{D}$ muni d'une topologie et considérons, dans l'esprit de la section 3.4.1 p. 31, que la «structure» d'une partition de $\mathcal{D}$ est donnée par ses **frontières**.

Une famille $\{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$ de partitions possède alors une structure causale quand pour tout couple $\lambda_2 > \lambda_1$ d'échelles :

i) Tout élément de frontière présent dans la partition $P(\lambda_2)$ possède un correspondant (une «cause») dans $P(\lambda_1)$

ii) La correspondance inverse est univoque, i.e. tout élément de frontière de $P(\lambda_1)$ correspond au plus à *un* élément de frontière de $P(\lambda_2)$.

Autrement dit, pour tout $\lambda_2 > \lambda_1$, les frontières de $P(\lambda_2)$ sont en bijection avec les frontières d'une *sous-partition* de $P(\lambda_1)$.

En outre, pour que la définition soit pertinente topologiquement, nous demandons que la bijection en question soit spatialement continue, ce qui conduit à la définition

Définition 19 *Une famille $\{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$ de partitions d'un domaine topologique $\mathcal{D}$ est dite* ***causale*** *quand $\forall(\lambda_1,\lambda_2) \in {\mathbb{R}^+}^2$, si $\lambda_2 \geq \lambda_1$ alors il existe un homéomorphisme h de $\mathcal{D}$ dans lui-même et une sous-partition Q de $P(\lambda_1)$ tels que $P(\lambda_2) = h(Q)$.*

Il semble alors naturel de requerir que les déformations de frontières qui interviennent dans cette définition soient également continues en échelle, i.e. que si λ_2 se rapproche de λ_1 alors la déformation h associée à $P(\lambda_2)$ et $P(\lambda_1)$ se rapproche de l'identité. Nous imposerons pour notre part un principe plus fort, à savoir que les analyses multi-échelles que nous recherchons **ne doivent pas délocaliser les frontières**. Autrement dit, nous imposerons que l'homéomorphisme qui intervient dans l'expression générale de la causalité structurelle soit systématiquement l'identité, ce qui conduit à la définition

Définition 20 *Une famille $\{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$ de partitions d'un domaine $\mathcal{D}$ telle que*

$$\forall(\lambda_1,\lambda_2) \in {\mathbb{R}^+}^2 \qquad \lambda_2 \geq \lambda_1 \quad \Rightarrow \quad P(\lambda_2) \geq P(\lambda_1) \tag{6.23}$$

sera appelée une ***analyse ensembliste multi-échelles*** *de $\mathcal{D}$.*

Une analyse ensembliste multi-échelles est donc simplement une famille de partitions $\{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$ *monotone* en fonction de λ.

Koepfler et al. (1994) ont appelé la relation 6.23 le *principe de causalité* ou principe de *structure pyramidale*. Nous venons de voir que la propriété de structure pyramidale à proprement parler peut s'envisager comme résultant de la combinaison de deux contraintes appliquées aux *frontières* des partitions, une contrainte de correspondance *topologique* selon la dimension d'échelle, dictée par le principe de causalité d'échelle lui-même, et une contrainte de correspondance *géométrique*, dictée par le souhait que les analyses à échelle grossière (fort λ) ne délocalisent pas les limites des régions. Pour faire un parallèle qualitatif avec les théories scale-space, une analyse ensembliste multi-échelles s'apparente donc aux analyses scale-space *non linéaires*. Le scale-space linéaire (gaussien) entraine en effet une délocalisation progressive des structures analysées (des courbes de niveau) : l'analyse devient de plus en plus grossière à la fois structurellement (disparition de courbes) *et* géométriquement (caricature des courbes).

Un algorithme A qui à toute image I sur $\mathcal{D}$ associe une analyse ensembliste multi-échelles $\{P(\lambda)\}_{\lambda\in\mathbb{R}^+}$ de $\mathcal{D}$ sera appelé un algorithme de segmentation multi-échelles. Le paramètre λ sera alors appelé le paramètre d'échelle de l'analyse.

Représentation ensembles-échelles d'une analyse multi-échelles

L'idée d'obtenir une analyse multi-échelles qui dépend d'un paramètre d'échelle *continu* est relativement séduisante. Toutefois, d'un point de vue purement pratique, l'axe des échelles étant continu, une telle description ne peut a priori pas être représentée ou calculée exactement. Les analyses scale-space sont par exemples contraintes à *échantillonner* l'axe des échelles pour pouvoir explorer la dimension d'échelle.

Il est néanmoins possible de représenter *exactement* et *exhaustivement* une analyse multi-échelles ensembliste. Il faut pour cela changer de point de vue sur cet objet et envisager une famille de partitions monotones de façon *implicite*, comme la *famille des coupes naturelles d'une hiérarchie indicée* (cf. 4.5.4 p.66). Détaillons ce changement de représentation, qui est une clef de voûte de notre approche.

Considérons un algorithme multi-échelles A et une image I définie sur un domaine $\mathcal{D}$ discret borné[34]. Notons $n = |\mathcal{D}|$. Considérons A comme une fonction de λ qui, pour $\lambda \in \mathbb{R}^+$ donné, renvoit la partition $\mathsf{A}(\lambda)$. Notons alors $S = \{\mathsf{A}(\lambda)\}_{\lambda\in\mathbb{R}^+}$ l'analyse multi-échelles complète obtenue sur I.

Comme $\mathcal{D}$ possède n points, les éléments de S forment une séquence d'au plus n partitions distinctes. Nous pouvons donc réécrire S sous la forme

$$S = \{P_1 < P_2 < \cdots < P_k\}, \quad \text{avec} \quad k \leq n.$$

Pour chaque partition P_i, considérons alors l'ensemble Λ_i des valeurs de λ pour lesquelles A produit la partition P_i, c'est-à-dire considérons l'image *inverse* de P_i par A :

$$\Lambda_i \triangleq \mathsf{A}^{-1}(P_i). \tag{6.24}$$

Si A est défini sur tout $\mathbb{R}^+$ et est univoque, alors les Λ_i constituent une partition de $\mathbb{R}^+$.

Si A est un algorithme multi-échelles, qui vérifie 6.23, alors les Λ_i ne sont pas entrelacés, i.e.

$$\forall (i,j) \in \langle 1,k\rangle^2 \quad \forall (x,y) \in \Lambda_i \times \Lambda_j \ : \quad i < j \ \Rightarrow \ x < y.$$

Les Λ_i constituent donc une suite d'*intervalles* contigus, de la forme :

$$\forall i \in \langle 1,k\rangle \quad \Lambda_i = [\lambda_i, \lambda_{i+1}[$$

[34] La présentation est faite dans le cas d'un domaine discret et borné à la fois pour simplifier et parce qu'il s'agit du cas concret qui nous intéresse. La construction est cependant directement extensible au cas d'un domaine continu. Les hiérarchies sont alors potentiellement infinies «vers le bas» car l'espace est infiniment divisible. Si le domaine est non borné, les hiérarchies peuvent également être infinies «vers le haut»...

avec $\lambda_1 = 0$ et $\lambda_{k+1} = +\infty$.

L'analyse multi-échelles S peut donc être représentée par une séquence d'au plus n partitions P_i, chacune attribuée par une échelle caractéristique λ_i qui représente *la plus petite échelle* pour laquelle l'analyse fournit la partition P_i. A propose la partition P_i jusqu'à ce que λ atteigne la valeur λ_{i+1}, valeur à laquelle P_i est remplacée par P_{i+1}.

La transition entre P_i et P_{i+1} correspond alors à la réunion de certaines régions de P_i en une seule région de P_{i+1}, tandis que d'autres régions de P_i se retrouvent intactes dans P_{i+1}. Si l'on ne considère que ces événements de réunion, chaque transition se traduit donc par l'*apparition* de nouvelles régions, à une échelle spécifique, en remplacement de certaines régions anciennes.

Finalement, l'analyse multi-échelles S peut complètement être décrite par un ensemble de **régions** $H = \{R_i\}_{i=1...m}$ caractérisées par leurs **échelles d'apparition** $\lambda^+ = \{\lambda_i\}_{i=1...m}$ dans un résultat de l'algorithme A.

H est défini par

$$H \triangleq \bigcup_{\lambda \in \mathbb{R}^+} \mathsf{A}(\lambda)$$

et si l'on envisage λ^+ comme une fonction $\lambda^+ : H \to \mathbb{R}^+$ alors $\forall x \in H$,

$$\lambda^+(x) \triangleq \min\{\lambda \in \mathbb{R}^+ | x \in \mathsf{A}(\lambda)\}.$$

Le couple (H, λ^+) possède la structure suivante :

- Les R_i sont hiérarchisées et si H contient chaque singleton de $\mathcal{D}$ et $\mathcal{D}$ lui-même - c'est-à-dire si
  - quand $\lambda \in [0, \lambda_2[$, $\mathsf{A}(\lambda)$ est la sur-partition absolue de $\mathcal{D}$
  - quand $\lambda \in [\lambda_k, +\infty[$, $\mathsf{A}(\lambda)$ est la sous-partition absolue de $\mathcal{D}$

  alors H est une *hiérarchie* sur $\mathcal{D}$.
- λ^+ vérifie :

$$\forall (x, y) \in H^2 \quad x \subset y \Rightarrow \lambda^+(x) \leq \lambda^+(y).$$

  c'est-à-dire est un *critère croissant* pour l'inclusion.

(H, λ^+) est donc une **hiérarchie indicée**.

Nous dirons que (H, λ^+) est la **représentation ensembles-échelle (scale-sets)** de l'analyse multi-échelles S. Les partitions renvoyées par A correspondent alors à la famille des **coupes naturelles** de cette hiérarchie indicée. Dans la représentation graphique naturelle de (H, λ^+) - obtenue en attribuant au nœud x une ordonnée qui correspond à son échelle d'apparition $\lambda^+(x)$ - ces coupes correspondent à des sections selon des *droites horizontales* (voir la figure 6.1).

Il est donc possible d'obtenir une représentation exacte d'une analyse multi-échelles ensembliste[35]. Il faut pour cela concevoir les partitions solutions du problème d'origine de façon **implicite**, comme des coupes dans une structure qui modélise véritablement un objet ayant une dimension de plus que le domaine de l'image. Le point clef est que la structure complète d'une analyse multi-échelles ensembliste est totalement déterminée par un ensemble d'événements d'apparition de régions (ou de manière équivalente de disparition de frontières), qui sont intrinsèquement des phénomènes *discrets* (et dans le cas de l'analyse d'images numériques sont en nombre *fini*). Ces événements peuvent alors être explicités, assortis de leurs **intervalles de vie en échelle**. L'atome conceptuel d'une représentation ensembles-échelle est donc la **région** et non la partition. Techniquement, s'intéresser aux intervalles de vie des objets revient à considérer l'application inverse A^{-1} (cf. equ. 6.24). Conceptuellement, cela signifie que plutôt que de se

[35] En pratique, les valeurs de λ^+ sont évidemment calculées à précision finie, e.g. en double précision. Le point important est que pour obtenir cette précision en échelle - qui est largement suffisante pour toutes les applications - il n'est pas nécessaire d'échantillonner l'axe des échelles à cette précision.

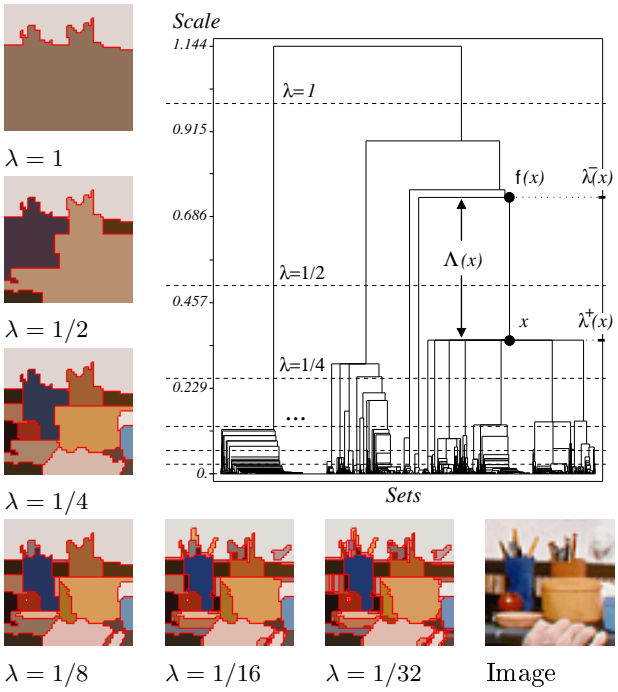


FIG. 6.1 – Une description ensembles-échelle et certaines de ses coupes. Cette analyse ensembles-échelle a été obtenue par l'algorithme d'escalade associé à la fonctionnelle constante par morceaux de Mumford et Shah (la méthode sera décrite au chapitre 8).

poser la question : «que voit-on à l'échelle λ?», il faut s'interroger : «à quelles échelles voit-on l'objet x?». L'ensemble de la méthodologie proposée repose sur ce renversement de point de vue. En outre, nous verrons qu'expliciter les intervalles de vie en échelle des ensembles donne toute leur richesse aux descriptions produites.

6.2.3 Positionnement par rapport aux approches existantes

Avant de poursuivre, positionnons l'approche proposée par rapport aux méthodes de segmentation existantes.

Processus hiérarchiques ou multi-résolution

Il est important de faire la distinction entre les méthodes de segmentation *utilisant* une stratégie hiérarchique ou multi-résolution pour construire une partition, et les méthodes *produisant* un résultat hiérarchique, multi-résolution ou multi-échelles (nous discuterons ci-dessous des différences entre ces trois types de résultats).

L'utilisation de processus hiérarchiques ou de stratégies multi-résolution (e.g. par rafinement progressif, «coarse to fine») pour résoudre un problème de partitionnement est très ancienne. Tout processus de partitionnement à base de fusion de régions balaye par exemple implicitement une hiérarchie. Comme nous l'avons vu, le fait d'expliciter une structure hiérarchique remonte au travaux de Pavlidis et de ses collaborateurs dans des années 70 (Tanimoto et Pavlidis 1975; Horowitz et Pavlidis 1976). Les hiérarchies considérées sont cependant des hiérarchies régulières (des quad-trees), de structure *prédéterminée*. Les approches correspondantes s'apparentent donc à des approches multi-résolution. Mais surtout, qu'il soit implicite ou explicite, régulier ou irrégulier, le processus d'exploration hiérarchique est systématiquement associé à un critère d'arrêt, s'appuyant soit sur un prédicat soit sur la minimisation d'une énergie, et dont

le but est de sélectionner une *partition unique* durant l'exploration de l'image.

Un certain nombre de processus multi-résolution ont ainsi été proposés dans le domaine markovien (Bouman et Shapiro 1994; Kato, Berthod, et Zerubia 1996; Laferté, Pérez, et Heitz 2000). Ces approches s'appuient sur une structure de quad-tree et définissent une séquence de champs de Markov, chaque champ étant défini sur un niveau du quad-tree. Outre ces dépendances intra-échelle, des dépendances markoviennes anti-causales sont alors également considérées (la réalisation d'une cellule d'un niveau donné dépend uniquement des réalisations de quelques cellules du niveau immédiatement supérieur). Le but est d'obtenir une labellisation de l'image d'origine, i.e. de la base du quad-tree, par raffinement progressif de labellisations plus grossières. Dans ces modèles, l'emploi d'une stratégie multi-résolution est essentiellement motivé par les importants temps de calcul qui sont nécessaires pour obtenir des simulations intéressantes de champs de Markov. L'approche markovienne multi-résolution permet de réduire ces coûts.

Analyses hiérarchiques, pyramidales ou multi-échelles

Même si elles utilisent un procédé ayant un certain caractère multi-échelles (ou plutôt multi-résolution), les méthodes que nous venons de décrire ne fournissent qu'un résultat mono-échelle. Nous proposons pour notre part de *produire une description multi-échelles* des données. Les propositions allant dans ce sens sont alors relativement peu nombreuses.

Analyses hiérarchiques – D'après Shi et Malik (2000), «l'action de partitionnement est intrinsèquement hiérarchique. Il est donc plus judicieux de penser à renvoyer la structure arborescente qui correspond à un partitionnement hiérarchique plutôt qu'une seule «plate» partition»[36]. Cette remarque de Shi et Malik est toutefois intimement liée au fait que la méthode de segmentation par coupe normalisée produit des partitions en deux classes. Elle doit donc être itérée pour obtenir des partitions de cardinal arbitraire, (voir 5.7.2 ci-dessus pour plus de détails). À notre connaissance, les auteurs n'ont pas développé cette idée d'obtenir une hiérarchie de régions par coupes normalisées itérées.

Poggi et Ragozini (1999) ont par contre développé l'idée de division en deux récursive dans le cadre Markovien. Encore une fois, la motivation première de ces auteurs est de réduire les importants temps de calcul qu'il faut déployer pour obtenir une réalisation d'un MRF qui soit proche du MAP. Plutôt que de rechercher une labellisation directe en K classes par un seul champ «plat», Poggi et Ragozini proposent alors d'itérer des labellisations en 2 classes, i.e. de découper l'image en deux grâce à un premier champ binaire, puis de redécouper chaque région obtenue par deux nouveaux champs binaires indépendants, etc. Les auteurs montrent que cette approche par MRF à structure arborescente (Tree-Structured MRF, TS-MRF) donne des résultats comparables à ceux obtenus par un seul MRF, tout en réduisant considérablement les temps de calcul. Ils notent également que conserver l'arbre entier de divisions peut s'avérer intéressant.

Dans les processus de fusion de régions, le processus d'exploration et le critère d'arrêt sont généralement profondément imbriqués, à la fois conceptuellement et pratiquement. Salembier et Garrido (2000) ont récemment proposé d'introduire une «séparation claire entre l'ordre de fusion et le critère d'arrêt»[37], ce qui les conduit à clairement séparer une étape de construction d'une hiérarchie d'une étape ultérieure d'exploitation de cette hiérarchie. Ils mettent alors en avant le fait que de nombreuses tâches d'analyse d'image peuvent être résolues efficacement à partir d'une description hiérarchique représentée sous forme arborescente. Nous reviendrons ci-dessous sur certaines applications proposées par ces auteurs.

Approches pyramidales – Différents auteurs proposent des algorithmes qui produisent des *pyramides de segmentations*, c'est-à-dire des séquences de partitions monotones.

[36] «the partitioning is inherently hierarchical. Therefore, it is more appropriate to think of returning a tree structure corresponding to a hierarchical partition instead of a single «flat» partition».

[37] «a clear separation between merging order and merging criterion».

Les algorithmes de pyramides stochastiques (Meer 1989), irrégulières (Montanvert, Meer, et Rosenfeld 1991) ou adaptatives (Jolion et Montanvert 1992) s'appuient sur un procédé de décimation stochastique des nœuds d'un graphe d'adjacence de régions utilisant la notion d'ensemble stable (Meer 1989). Itéré, ce processus assure une réduction exponentielle de la taille du graphe en fonction de l'itération. Partant d'un graphe de voisinage de pixels et si n est le nombre de pixels de l'image, on obtient donc une pyramide de graphes ayant statistiquement $\log n$ niveaux, comme dans les pyramides régulières par quad trees et dans les pyramides multi-résolutions. L'idée est toutefois de considérer des ensembles de géométrie irrégulière afin de pallier aux importants défauts des pyramides à géométrie rigide, qui ne respectent pas certaines invariances élémentaires (Bister, Cornelis, et Rosenfeld 1990).

Le procédé initial de décimation stochastique de Peter Meer ne s'appuie pas sur le contenu de l'image mais vise à proposer des grilles d'analyse multi-résolution qui possèdent les mêmes propriétés statistiques qu'une multi-résolution à cellules régulières. Des travaux théoriques ultérieurs ont porté sur le procédé de décimation lui-même et sa formalisation dans le cadre de la théorie des graphes, aboutissant en particulier à la notion de noyaux de contraction équivalents (Kropatsch 1995) ou à des modélisation du processus en termes de cartes combinatoires, i.e. en termes de graphe dual.

Pour appliquer ce principe de construction de structures irrégulières à la segmentation, l'idée consiste alors à biaiser le processus de décimation pour qu'il favorise le regroupement des pixels les plus similaires, produisant donc des pyramides adaptatives au contenu de l'image (Jolion et Montanvert 1992). Bertolino (1995) a proposé des améliorations des algorithmes initiaux, reposant en particulier à la fois sur des critères «régions» et «contours». Notons également l'idée de Cho et Meer (1997) qui proposent d'exploiter l'aspect stochastique des pyramides pour créer des «pyramides de consensus» à partir de plusieurs résultats du même algorithme sur une image. Cette idée est assez proche de l'idée de Leclerc (1989a) de solution «stable» et de l'idée de persistance en échelle que nous développerons ci-dessous.

Remarquons pour conclure sur ces approches que dans une pyramide irrégulière ou adaptative, le paramètre d'«échelle» est simplement le niveau dans la pyramide - l'itération du procédé de décimation - et ne revêt aucun sens fort (en particulier parce que la décimation est forcée à chaque étape).

Quelques auteurs ont proposé des algorithmes de segmentation pyramidale qui s'appuient sur des formulations par minimisation d'énergie. Koepfler et al. (1994) proposent de faire varier le poids du terme de régularisation dans la fonctionnelle constante par morceaux de Mumford-Shah pour obtenir des pyramides. Ballester et al. (1994) procèdent de façon identique mais utilisent une variante de la fonctionnelle de Mumford-Shah. De son côté, Fuchs (2001) s'appuie sur un critère «MDL» pour décomposer un MNE en facettes planes et propose d'augmenter le coût de codage d'un paramètre du modèle pour trouver des solutions de différents finesses.

Dans ces trois approches, les pyramides de segmentations sont construites en itérant un algorithme de fusion de régions sur un critère de descente énergétique (FRD, cf. 6.1.5 p.129). Une suite $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ de valeurs croissantes du paramètre d'«échelle» et une sur-partition initiale P_0 sont considérées. Une séquence de k partitions emboîtées est alors obtenue en calculant P_i par descente depuis P_{i-1} en réglant l'énergie à l'échelle λ_i. Le critère d'arrêt est la 2-normalité.

Seulement quelques partitions sont ainsi calculées, correspondant à une suite de valeurs d'échelle fixées *a priori* (dans toutes ces approches $\lambda_i \propto 2^i$). L'axe des échelles est donc échantillonné. Mais plus fondamentalement, les valeurs de λ attachées à chaque partition ne revêtent pas de sens profond car elles ne correspondent pas à des échelles caractéristiques des objets.

Une dernière catégorie importante d'approches pyramidales est issue de la théorie morphologique des opérateurs connexes (Serra et Salembier 1993; Salembier et Pardàs 1994; Salembier et Serra 1995; Salembier, Oliveras, et Garrido 1998). Les pyramides sont obtenues par l'application de séquences de filtres connexes, qui élargissent systématiquement les zones plates d'une image et produisent donc des partitions en zones plates de finesse décroissante. Dans la lignée classique de la morphologie mathématique,

un filtre connexe sur les images en noir et blanc peut être construit à partir d'un opérateur connexe sur les ensembles, via la décomposition en sections des images. Une catégorie importante de filtres connexes sont les filtres par reconstruction.

Pour conclure sur les approches pyramidales, même si le modèle qui permet de générer l'analyse dépend explicitement d'un paramètre d'«échelle» continu, seulement quelques coupes à échelles fixes sont produites, l'axe des échelles est échantillonné. Qualitativement, le résultat est une multi-résolution.

Analyse multi-échelles – À notre connaissance, seule l'approche par arbre de poids minimal de Fernand Meyer (1999a, 1999b) propose un résultat multi-échelles au sens où nous l'entendons, à savoir une analyse à finesse variable dépendant d'un paramètre d'«échelle» *continu*.

L'approche de Meyer découle de la remarque que l'algorithme classique de calcul d'une ligne de partage des eaux, par simulation d'immersion (cf. annexe A), produit naturellement une analyse multi-échelles. En effet, au fur et à mesure que le niveau d'eau monte, des bassins se mélangent (des régions fusionnent) et ce à certains niveaux spécifiques, qui correspondent à l'altitude du plus bas point séparant deux bassins, habituellement situé sur un col. Meyer remarque alors que cette analyse multi-échelles s'obtient en i) construisant le graphe d'adjacence des bassins versants du relief ii) attribuant ses arêtes par l'altitude minimale sur la frontière correspondante et iii) calculant l'arbre de poids minimal (MST) recouvrant ce graphe, qui représente implicitement la hiérarchie indicée obtenue par single-linkage sur le graphe ou encore l'espace ultramétrique correspondant.

La coupe naturelle de niveau λ de cette analyse s'obtient alors en supprimant toutes les arêtes de l'arbre qui ont un poids inférieur à λ et en calculant les composantes connexes de la forêt résultante. Ces composantes correspondent aux régions de la coupe.

Cette analyse est habituellement calculée sur un gradient de l'image d'origine. Meyer remarque cependant que l'analyse qui résulte de l'algorithme classique d'immersion privilégie le contraste des objets, indépendamment de leur taille. L'analyse exhibe donc à grande échelle les détails les plus marqués plutôt que des unités visuelles globales. Meyer propose alors plusieurs variantes du mécanisme d'immersion, appelées immersions *asynchrones* car le niveau d'eau monte à une vitesse différente dans chaque bassin. Une première variante correspond à inonder chaque bassin par une source virtuelle ayant un débit proportionnel à la *surface* du bassin. Ce procédé produit des analyses dont les coupes comportent des objets de tailles homogènes, indépendamment de leur contraste. Le critère finalement retenu par Meyer correspond à régler le débit proportionnellement au *volume* des bassins, un critère qui mélange le contraste et la taille des régions. Ces variantes reviennent à modifier les dissemblances initiales portées par les arêtes du RAG et la règle de mise à jour des priorités après fusion.

D'un point de vue applicatif, Meyer (1999b), Zanoguera, Marcotegui, et Meyer (1999) et Zanoguera (2001) ont montré qu'une analyse multi-échelles pouvait servir à la réalisation de systèmes de segmentation interactive très efficaces.

La représentation par MST d'une analyse multi-échelles proposée par Meyer et la représentation ensembles-échelle que nous proposons peuvent s'envisager comme deux représentations **duales**, au sens de la dualité régions/contours.

Il s'agit de deux représentations sous forme d'arbre attribué. Toutefois, l'arbre considéré par Meyer est un arbre spatial dont les *arêtes* - qui correspondent aux frontières de la partition de base - sont attribuées par l'échelle à la laquelle elles *disparaissent*. L'arbre que nous considérons est un arbre d'inclusion de régions dans lequel chaque ensemble est attribué par l'échelle à laquelle il *apparaît*. Dans l'approche par MST les ensembles et surtout leurs intervalles de vie sont seulement *implicites*. On peut évidemment à loisir passer d'une représentation à l'autre, toutefois privilégier l'une ou l'autre conduit à des idées différentes. Notons également que les hiérarchies obtenues par MST sont systématiquement des hiérarchies binaires.

Enfin, de manière plus générale, les formalisations de la segmentation via un prédicat d'homogénéité ou via une mesure de dissemblance reposent sur *un seul* critère de qualité sur les régions, critère lié à l'aspect *radiométrique* des objets. Ce type d'approche ne permet pas naturellement d'intégrer de facteur de qualité *géométrique* sur les objets.

6.3 Description de la démarche

Notre approche peut globalement s'envisager comme un mariage entre l'aspect «énergétique» des approches de Koepfler et al. (1994), Ballester et al. (1994) ou Fuchs (2001) et l'aspect «structurel» de l'approche de Meyer (1999a).

Du point de vue énergétique, nous avons toutefois mis en évidence que l'existence d'un paramètre d'«échelle» était inhérente au problème de modélisation. L'«échelle» est le compromis entre la «simplicité» et la «fidélité» du modèle. Un unique principe, le principe de parcimonie, nous a conduit à deux expressions duales relativement générales du problème de modélisation/description par parties d'une image, qui correspondent à deux problèmes de débit/distorsion opérationnel. Les énergies que nous considérons - les énergies affines - ont alors été dérivées par équivalence lagrangienne.

Du point de vue structurel, ensembliste, l'idée d'échelle est associée à l'idée d'analyse à finesse variable. En termes de partitions, une analyse multi-échelles est donc une séquence de partitions monotones. Adoptant une approche similaire à l'approche scale-space, nous avons montré qu'une telle structure pouvait se déduire de la combinaisons de deux principes élémentaires : le principe de causalité et un principe de non délocalisation des limites des objets. Nous posant le problème pratique de la représentation exhaustive d'une analyse multi-échelles, nous avons conclu qu'il fallait abandonner le point de vue des partitions, et proposé de représenter une analyse multi-échelles sous forme de hiérarchie indicée. Cette représentation a été baptisée représentation ensembles-échelle pour mettre en évidence le parallèle avec une représentation scale-space.

Il s'agit maintenant de marier ces deux aspects. La question qui vient naturellement à ce point est alors : les partitions minimisant une énergie affine constituent-t-elles effectivement une analyse multi-échelles ? La réponse est négative en général.

6.3.1 Non monotonie a priori des minima d'une énergie affine

Intéressons-nous par exemple à l'énergie proposée par Mumford et Shah pour les modèles d'images constants par morceaux. Rappelons que dans ce cas, C représente la longueur totale des frontières de la partition et D est la somme des écarts quadratiques à la moyenne de chaque région. En continu, il est évident que la famille des minima de cette fonctionnelle ne constitue absolument pas une analyse ensembliste multi-échelles : les frontières se déforment continuement (se tendent) quand λ augmente et ne font pas que disparaître. Et même si l'on force les frontières à ne pas se déplacer quand l'échelle augmente, on peut obtenir des solutions non monotones.

La figure 6.2(a) l'illustre sur un exemple. L'image considérée présente trois régions A, B et C totalement homogènes ($\mathsf{D}(A) = \mathsf{D}(B) = \mathsf{D}(C) = 0$) avec A voisine de B et B voisine de C mais A et C non adjacentes.

De manière relativement évidente, la solution à échelle nulle comporte les trois régions séparées en trois pièces distinctes (rigoureusement, cette solution est obtenue en tant que solution unique pour $\lambda = \varepsilon > 0$ assez faible). Imposons-nous alors de rechercher les solutions à plus grande échelle parmi les sous-partitions de $\{A, B, C\}$. On sait que les minima de la fonctionnelle de Mumford-Shah sont des partitions en ensembles connexes. Les sous-partitions en ensembles connexes de $\{A, B, C\}$ sont $\{\{A \cup B, C\}, \{A, B \cup C\}, \{A \cup B \cup C\}\}$.

Supposons que $|A| = |B| = |C| = 1$. Supposons que les valeurs de l'image sur A et B soient très proches mais que la valeur sur C en soit très différente. Notons α la variance de l'image sur $A \cup B$, β

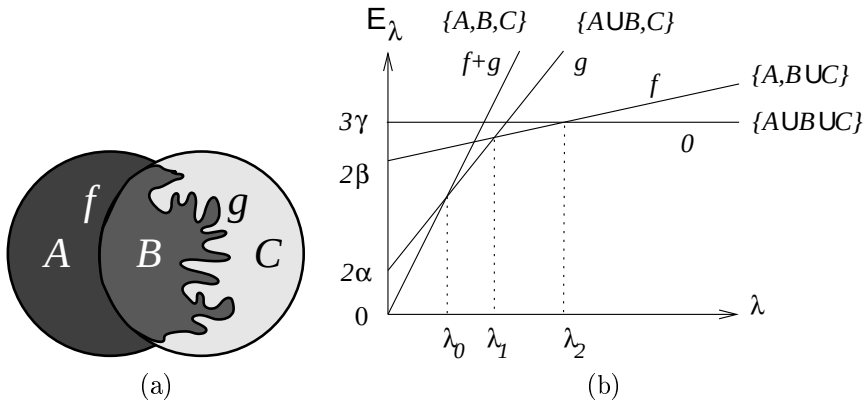


FIG. 6.2 – Exemple de non monotonie de partitions minimales en fonction de l'échelle. Voir texte.

celle sur $B \cup C$ et γ celle sur $A \cup B \cup C$. Soit f la longueur de la frontière séparant A et B et g celle de la frontière entre B et C. Supposons que $g \gg f$. Le tableau qui suit résume les composantes des énergies affines associées à chaque partition et la figure 6.2(b) représente leur évolution en fonction de λ. Le terme D correspond à l'ordonnée à l'origine de la droite qui représente l'évolution de l'énergie en fonction de λ et le terme C représente sa pente.

P	C	D
$\{A, B, C\}$	$f + g$	0
$\{A \cup B, C\}$	g	2α
$\{A, B \cup C\}$	f	2β
$\{A \cup B \cup C\}$	0	3γ

On voit alors que l'on peut choisir f,g,α,β et γ de telle manière qu'il existe $\lambda_0 < \lambda_1 < \lambda_2$ tels que :

- quand $\lambda_0 \leq \lambda < \lambda_1$ la partition optimale est $\{A \cup B, C\}$
- quand $\lambda_1 \leq \lambda < \lambda_2$ la partition optimale est $\{A, B \cup C\}$
- quand $\lambda \geq \lambda_2$ la partition optimale est $\{A \cup B \cup C\}$

Dans cet exemple, le non emboîtement des partitions optimales provient d'un fort désaccord entre le terme de régularisation et celui d'attache aux données quant à la meilleure fusion : $B \cup C$ ou $A \cup B$. Pour l'énergie C, fusionner B et C est préférable par rapport à fusionner A et B car cela diminue nettement plus la longueur totale de frontières. C'est le contraire pour l'énergie D car les valeurs de l'image sur A et B sont beaucoup plus voisines que celles sur B et C. En fonction du poids relatif accordé à C et D l'une ou l'autre de ces solutions antagonistes est alors optimale.

Remarquons que ces deux solutions sont également conflictuelles d'un point de vue visuel : nous avons tendance à grouper A et B sur un critère de proximité de teinte et B et C sur un critère de simplicité de forme. Desolneux, Moisan, et Morel (2001b) parleraient de deux gestalts partielles conflictuelles.

Notons que Kervrann, Hoebeke et Trubuil (1999, 2002) ont montré que pour les images mono-canal et pour certains choix particuliers des énergies C et D, les solutions obtenues correspondent à sélectionner des *courbes de niveau* de l'image, dont nous avons vues en 5.5.2 p. 87 qu'elles constituent une structure multi-échelles (toutefois dans un sens différent que celui auquel nous nous intéressons ici). Les énergies C utilisées ne sont cependant pas des termes de régularisation géométrique mais dépendent uniquement du nombre de parties dans la partition ou de leurs cardinaux. Nous verrons ci-dessous que ce type de modèle est très proche des modèles par prédicats d'homogénéité. Ces derniers peuvent se formuler en termes de minimisation d'énergie et les termes de régularisation obtenus sont du type de ceux utilisés par Kervrann et al.

Remarquons finalement que même s'il est possible - comme nous venons de le faire - de construire

des exemples de solutions non monotones, ces exemples correspondent à des cas relativement extrêmes (notons la parenté de l'exemple donné ci-dessus avec l'exemple de la figure 3.2 p. 38 de non hiérarchisation de groupements perceptuels dûe à une transparence où à une ombre portée).

6.3.2 Présentation de la démarche

Comment alors concilier l'aspect énergétique et l'aspect structurel ?

De manière encore informelle, nous proposons de rechercher un algorithme d'analyse multi-échelles dont le résultat est une description ensembles-échelle dont les coupes naturelles soient «aussi proches que possible» des minima d'une énergie affine.

La démarche pour obtenir cet algorithme repose alors sur la remarque qu'une analyse ensembles-échelle possède deux composantes : une composante structurelle - la hiérarchie - et un axe d'échelle - l'indice. Ces deux composantes sont nécessaires pour déterminer complètement les solutions du problème envisagé. La hiérarchie définit les régions qui sont produites par l'algorithme. Elles ne détermine cependant qu'un ordre *partiel* sur ces régions. L'indice définit quant à lui l'ordre *total* dans lequel ces régions sont produites et - implicitement - les partitions que l'on trouve à échelle donnée. Dans la ligne d'idée de Salembier et Garrido (2000), ces deux composantes suggèrent une décomposition du problème de segmentation multi-échelles en deux étapes :

i) Construction d'une hiérarchie de régions H.
ii) Définition d'un indice sur H, correspondant à une famille de coupes monotones de H.

La méthodologie proposée s'appuie alors sur une étude de ces problèmes *en ordre inverse* :

1. Dans un premier temps, le chapitre 7 commence par supposer que la hiérarchie H a déja été construite - par un moyen quelconque - et étudie le problème de coupe dans cette hiérarchie, selon un principe de minimisation d'énergie. Nous montrons qu'une classe relativement générale d'énergies affines conduit effectivement à des familles de coupes minimales monotones et nous dégageons un algorithme exact pour calculer ces familles.

2. Dans un second temps, le chapitre 8 effectue un retour sur la construction de la hiérarchie elle-même. Par une sorte de procédé de bootstrap, les résultats du chapitre 7 nous conduisent en effet à un principe naturel, baptisé principe d'**escalade**, pour construire une analyse multi-échelles complète en s'appuyant sur la minimisation d'une énergie affine. Le chapitre aboutit finalement à une formalisation précise du problème de segmentation multi-échelles, qui repose sur la définition d'un critère d'optimalité pour les *hiérarchies* plutôt que pour les partitions. Nous montrons alors que l'escalade peut s'envisager comme une stratégie d'optimisation gloutonne pour résoudre ce problème, qui produit des solutions vérifiant plusieurs propriétés remarquables (invariances, solutions identiques pour les deux problèmes duaux de débit/distorsion).

Le chapitre 9 met finalement en application la méthodologie proposée. Il présente des résultats expérimentaux obtenus pour différentes énergies et étudie en détails le comportement de l'algorithme.

Chapitre 7

Minimisation d'énergies multi-échelles dans une hiérarchie

Ce chapitre développe la première étape de la démarche que nous venons de présenter.

Il est divisé en deux parties, 7.1 et 7.2.

La première partie s'intéresse à la recherche d'*une* coupe minimale dans une hiérarchie pour une énergie E donnée. Nous voyons en 7.1.1 que si E est une énergie *séparable*, cette coupe se calcule efficacement par un algorithme de programmation dynamique. La section 7.1.2 détaille alors la classe des énergies séparables et le type de modèle auquel elles correspondent. La section 7.1.3 s'intéresse ensuite à des questions de combinatoire : combien y a-t-il de coupes dans une hiérarchie ? Comment cette combinatoire se situe-t-elle par rapport à celle d'autres espaces de partitions ? La section 7.1.4 compare finalement la stratégie d'optimisation proposée aux algorithmes classiques de fusion par descente de gradient.

La deuxième partie du chapitre s'intéresse à la famille des coupes minimales que l'on obtient quand on fait varier le paramètre λ d'une énergie affine séparable. La section 7.2.1 montre que sous une hypothèse relativement naturelle sur l'énergie de régularisation, ces coupes sont monotones, i.e. constituent une famille de solutions multi-échelles. La section 7.2.2 propose alors une extension de la méthode de programmation dynamique vue en 7.1, qui permet d'obtenir efficacement la représentation ensembles-échelle exacte de cette famille de solutions. La section 7.2.3 discute les résultats obtenus et propose quelques extensions. La section 7.2.4 revient finalement sur les deux problèmes duaux d'optimisation sous contrainte dont les énergies considérées sont les Lagrangiens et montre que la méthode permet d'obtenir toutes leurs solutions, pour toutes les valeurs de contrainte, quand on s'impose de rechercher ces solutions parmi les coupes d'une hiérarchie.

7.1 Coupe minimale d'une hiérarchie

Comme nous l'avons vu, trouver la partition de $\mathbb{P}(X)$ qui minimise une énergie E est généralement un problème difficile. Cependant, si l'on dispose d'une hiérarchie H sur X et que l'on ne s'intéresse qu'aux partitions que l'on peut former avec des éléments de H, c'est-à-dire aux coupes de H, alors le problème se résout facilement par un algorithme de programmation dynamique.

7.1.1 Algorithme de programmation dynamique

Il faut pour celà que l'énergie considérée soit **séparable**, c'est-à-dire s'exprime pour une partition P comme une somme d'énergies portant sur les régions de P :

$$\mathsf{E}(P) = \sum_{R \in P} \mathsf{E}(R). \tag{7.1}$$

Nous utiliserons la même lettre pour désigner une fonction d'énergie définie sur les parties d'un ensemble et l'énergie séparable définie sur les partitions qui lui est associée.

Soit H une hiérarchie sur X. Pour tout $x \in H$, le sous-ensemble $H(x)$ de H défini par

$$H(x) \quad \triangleq \quad \{y \in H \mid y \subseteq x\}$$

est de toute évidence également une hiérarchie. $H(x)$ a pour sommet x et sa base est

$$\mathcal{B}(x) \quad \triangleq \quad \{y \in \mathcal{B}(H) \mid y \subseteq x\}.$$

Définition 21 *Nous appelons $H(x)$ la **hiérarchie partielle** issue de x et $\mathcal{B}(x)$ la **base partielle** issue de x.*

La figure 7.1 illustre ces définitions. Notons qu'évidemment, $\forall H \in \mathbb{H}(X) \ : \ H(\vee H) = H$.

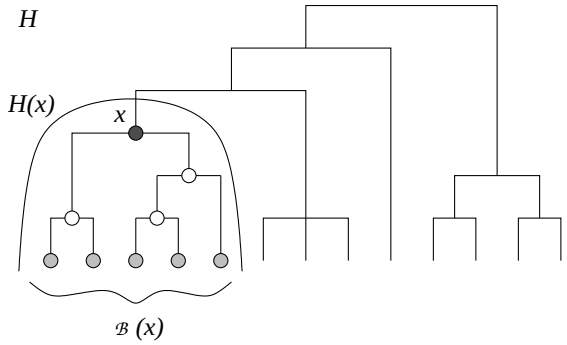


FIG. 7.1 – Hiérarchie partielle. $H(x)$ est la hiérarchie partielle issue de x et $\mathcal{B}(x)$ est sa base.

Considérons alors une énergie $\mathsf{E} : \mathbb{P}(X) \to \mathbb{R}^+$ séparable.

Pour tout $x \in H$ notons $C^*(x)$ la coupe de la hiérarchie partielle $H(x)$ d'énergie E minimale, i.e.

$$C^*(x) \triangleq \underset{C \in \mathbb{C}ut(H(x))}{\operatorname{argmin}} \mathsf{E}(C) = \underset{C \in \mathbb{C}ut(H(x))}{\operatorname{argmin}} \sum_{y \in C} \mathsf{E}(y)$$

et notons pour simplifier $\mathsf{E}^*(x) \triangleq \mathsf{E}(C^*(x))$ l'énergie de cette coupe. Avec ces notations,

Proposition 5 *Les énergies E^* des coupes minimales des hiérarchies partielles d'une hiérarchie H sont reliées par la relation de programmation dynamique*

$$\forall x \in H \qquad \mathsf{E}^*(x) = \min \Big\{ \mathsf{E}(x), \sum_{y \in \mathcal{F}(x)} \mathsf{E}^*(y) \Big\}. \tag{7.2}$$

où $\mathcal{F}(x)$ désigne l'ensemble des fils de x dans H.

Preuve : Pour tout $x \in H$ définissons

$$U(x) \triangleq \bigcup_{y \in \mathcal{F}(x)} C^*(y)$$

qui représente l'union des coupes optimales des hiérarchies partielles issues des fils de x.

Soit B une branche de $H(x)$. B passe par un et un seul fils y de x. $B \backslash x$ est donc une branche de $H(y)$. Comme $C^*(y)$ est une coupe de $H(y)$, la branche $B \backslash x$ l'intersecte une et une seule fois. En outre, $B \backslash x$ n'intersecte aucune autre coupe $C^*(z), z \in \mathcal{F}(x), z \neq y$, car les hiérarchies partielles issues des fils de x sont disjointes. U est donc une coupe de $H(x)$.

Réciproquement, on voit immédiatement que toute coupe de $H(x)$ qui n'est pas la partition triviale $\{x\}$ se décompose en une union de coupes des hiérarchies partielles issues des fils de x. Soit V une telle coupe. Elle se décompose en

$$V = \bigcup_{y \in \mathcal{F}(x)} V(y).$$

L'énergie E étant séparable, on a alors

$$\mathsf{E}(V) = \sum_{y \in \mathcal{F}(x)} \mathsf{E}(V(y)).$$

Cette décomposition vaut en particulier pour la coupe $U(x)$, aussi comme les $C^*(y)$ sont les coupes optimales des $H(y)$, on obtient

$$\forall V \in \mathbb{C}ut(H(x)) \backslash \{x\} \quad \mathsf{E}(U(x)) \leq \mathsf{E}(V).$$

En conclusion :

$$\begin{array}{llll} si & \mathsf{E}(x) \leq \mathsf{E}(U(x)) & C^*(x) = \{x\} & et \quad \mathsf{E}^*(x) = \mathsf{E}(x). \\ sinon & & C^*(x) = \bigcup_{y \in \mathcal{F}(x)} C^*(y) & et \quad \mathsf{E}^*(x) = \sum_{y \in \mathcal{F}(x)} \mathsf{E}^*(y). \end{array}$$

■

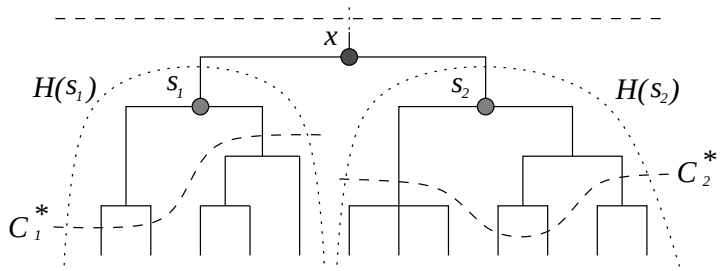


FIG. 7.2 – Un pas de programmation dynamique pour le calcul d'une coupe minimale. La coupe de $H(x)$ minimisant E est soit $\{x\}$, soit l'union des coupes optimales des hiérarchies partielles issues des fils de x.

La relation de programmation dynamique 7.2 permet de trouver la coupe optimale de H - et son énergie - en une passe ascendante dans la hiérarchie. La figure 7.2 illustre une étape de l'optimisation.

Les nœuds doivent être visités en respectant l'ordre partiel des inclusions. Si l'on dispose de la hiérarchie H sous forme d'arbre, cet ordre peut être trouvé en une passe descendante de H, par exemple par une procédure récursive similaire à celle proposée en section 4.5.3 p. 65 pour le dessin planaire d'une hiérarchie. Notons toutefois que la hiérarchie a généralement été calculée par une méthode de fusion ou de division, qui produit les éléments de H selon un ordre compatible avec l'ordre des inclusions. Si c'est le cas, on a tout intérêt à mémoriser cet ordre au cours de la construction de H.

Supposons donc que l'on dispose d'une liste ordonnée $L = (r_1 \dots r_N)$ des $N = |H|$ nœuds de H qui vérifie que $i < j \Rightarrow r_i \subset r_j$ ou $r_i \cap r_j = \emptyset$. Par définition, $\forall x \in H, x \subseteq \vee H$ et donc $r_N = \vee H$ (par contre les $|\mathcal{B}(H)|$ premiers nœuds de L ne sont pas nécessairement les nœuds de base).

L'algorithme d'optimisation s'écrit alors

Algorithme 2 **Coupe minimale(H,L,E)**

Pour i variant de 1 à N

1. Calculer ou lire $\mathsf{E}(r_i)$.

2. Faire $\mathsf{E}^*(r_i) \leftarrow \mathsf{E}(r_i)$; $C^*(r_i) \leftarrow \{r_i\}$; $S \leftarrow +\infty$.
3. Calculer $S \leftarrow \sum_{y \in \mathcal{F}(r_i)} \mathsf{E}^*(y)$.
4. Si $S < \mathsf{E}^*(r_i)$ faire $\mathsf{E}^*(r_i) \leftarrow S$ et $C^*(r_i) \leftarrow \bigcup_{y \in \mathcal{F}(r_i)} C^*(y)$.

fin Pour.

Fin de l'algorithme

Le dernier nœud visité est le sommet de H, $C^*(r_N)$ donne la coupe minimale de H et $\mathsf{E}^*(r_N)$ l'énergie de cette coupe.

Complexité de l'algorithme :

Une hiérarchie sur X contient au plus $N = 2|X| - 1$ éléments (quand il s'agit d'un arbre binaire) et donc au plus $M = 2|X| - 2$ arêtes, d'après le théorème d'Euler ou la relation cyclomatique (relation 4.2 p. 51) appliquée à un arbre. On voit que les sommes de l'étape 3 de l'algorithme 7.1.1, mises bout à bout, correspondent à balayer une et une seule fois chaque arête de H. L'application de 7.2 à tous les nœuds de H conduit donc à une complexité en $M + N$ et requiert exactement N évaluations de la fonction d'énergie E sur les parties de H. L'algorithme de calcul de l'*énergie* de la coupe minimale est donc linéaire en fonction de $|X|$ (si l'on ne prend pas en compte le coût de calcul de E, i.e. si l'on suppose que l'énergie de chaque région est une entrée de l'algorithme d'optimisation ou si l'on suppose que le calcul de $\mathsf{E}(x)$ s'effectue en temps constant sur une partie x).

Notons toutefois que les unions de coupes partielles de l'étape 4, si elles sont exécutées sans précaution, peuvent conduire à une complexité quadratique, à la fois en temps et en mémoire. En effet, si l'on calcule U par *recopie* des éléments des coupes $C^*(y), y \in \mathcal{F}(r_i)$, alors la mémoire totale nécessaire comme le nombre d'opérations élémentaires requises est $T = \sum_{i=1}^{N} |C^*(r_i)|$. T peut être en N^2 si la hiérarchie a N niveaux et que la coupe optimale est la sur-partition absolue de X (la coupe optimale est dans ce cas mémorisée à chaque niveau). On pourrait simplement vider les valeurs de C^* qui ne sont plus utiles, i.e. dès que la recopie a été faite, mais cela ne résoudrait pas le problème de temps de duplication. La solution la plus efficace pour résoudre les deux problèmes - mémoire et temps - est d'utiliser des listes chaînées pour représenter les coupes optimales $C^*(r_i)$. La création de U à l'étape 3 ne se fait alors pas par recopie des C^* des fils du nœud courant r_i mais par concaténation de ces listes en une seule, ce qui nécessite $\mathcal{O}(|\mathcal{F}(r_i)|)$ opérations de dépointage de successeur dans une liste chaînée.

Une autre solution pour calculer la coupe en temps et espace linéaire consiste à effectuer deux traversées de la hiérarchie, une passe ascendante suivie d'une passe descendante. Au cours de la passe ascendante, on ne calcule que l'énergie de la coupe optimale tout en marquant chaque nœud n_i qui vérifie que la meilleure coupe dans la hiérarchie partielle $H(r_i)$ dont il est le sommet est lui même, i.e. est $\{r_i\}$. Ceci se produit quand $\mathsf{E}(r_i) < \sum_{y \in \mathcal{F}(r_i)} \mathsf{E}^*(y)$ dans l'équation de programmation dynamique 7.2. Après la remontée, les plus hauts nœuds marqués sont ceux qui optimisent la coupe globale. Une redescente de la hiérarchie (en profondeur ou largeur d'abord) permet de les collecter. Cette décomposition en deux de l'optimalité globale d'une région sera à la base de la méthode de coupe multi-échelles de la deuxième partie de ce chapitre (section 7.2).

Discussion

Le principe de programmation dynamique, inventé par Richard Bellmann (1957), est généralement connu pour son application aux problèmes de recherche de plus court chemin (pcc) dans un graphe. Nous l'avons par exemple exploité avec Jean-Marc Viglino (Viglino and Guigues 2000) pour rechercher un réseau d'hypothèses de routes traversant une image aérienne[1]. La programmation dynamique a également trouvé de multiples applications dans les problèmes de mise en correspondance linéaire, qui se formalisent

[1] La méthode s'appuie sur une variation autour du principe standard de recherche de plus court chemin, qui permet - si le graphe est un graphe orienté sans cycle possédant deux extrêmités a et b particulières (e.g. le graphe représentatif d'un treillis) - de calculer pour tout nœud c du graphe, le plus court chemin reliant a à b en passant par c.

souvent comme des problèmes de pcc (mise en correspondance de textes, alignement moléculaire (Danchin 2001), stéréo-vision le long de lignes épipolaires (Baillard 1997)...). La recherche d'une approximation polygonale d'une courbe (Pierrot Deseilligny 1990) ou le recalage d'une ligne polygonale (Jibrini 2002) peuvent également se concevoir comme des recherches de pcc. Notons que pour ces problèmes de plus court chemin, la longueur d'un chemin est formulée comme la somme des longueurs des arcs qui le composent, ce qui est l'analogue de la condition de séparabilité de l'énergie dans l'algorithme de coupe minimale que nous venons de décrire.

Le principe de programmation dynamique possède cependant une portée bien plus générale. Comme l'exprime Danchin (2001) : «Cette méthode [...] est plus une «philosophie» qu'un algorithme particulier». Globalement, la «philosophie» consiste à plonger le problème d'origine dans une famille de problèmes d'optimisation dont les solutions sont reliées par une relation de récurrence (Gondran et Minoux 1995, Annexe 4). Dans l'algorithme d'optimisation hiérarchique qui vient d'être présenté, la famille de problèmes considérée correspond à la famille des hiérarchies partielles de la hiérarchie d'origine et chaque problème partiel est de même nature que le problème global. Quand il s'agit de segmentation d'image, les résolutions partielles correspondent à des résolutions du problème de segmentation sur des *sous-images*. La relation de récurrence est alors *ascendante* : les solutions sur les petites régions se combinent itérativement pour donner celles sur les grandes régions d'image. Nous y verrons ci-dessous une relation profonde avec le principe de causalité des théories multi-échelles.

Breiman, Friedman, Olshen, et Stone (1984) sont - à notre connaissance - les premiers à avoir proposé l'algorithme 7.1.1 pour trouver une coupe minimale dans une hiérarchie. Ils ne le présentent cependant pas comme un algorithme de programmation dynamique. Ils l'utilisent pour des problèmes de classification (algorithme CART).

L'algorithme *best-basis* (meilleure base) de Coifman et Wickerhauser (1992) emploie également cette méthode. Face au très vaste choix d'ondelettes possible pour l'analyse d'un signal, ces auteurs se posent le problème de construire la meilleure base pour un problème de compression donné (voir une présentation simple dans (Burke Hubbard 1995)). Les hiérarchies sont dans ce cas des hiérarchies régulières (quad-trees) supportant les coefficients d'ondelettes. Donoho (1997) fait le lien entre la méthode CART de Breiman et al. et l'algorithme *best-basis* de Coifman et Wickerhauser (c'est par cet article de Donoho que nous avons eu connaissance de l'utilisation de l'algorithme de coupe minimale dans CART et best-basis). Donoho (1999) réemploie le même genre de technique pour la compression en «wedgelets», adaptée au codage des arêtes. Ndili, Nowak, et Figueiredo (2001) utilisent les wedgelets dans une méthode de partitionnement s'appuyant sur une exploration par pyramide régulière (quad-tree) pour minimiser un citère MDL.

Salembier et Garrido (2000) ont également proposé d'utiliser l'algorithme 7.1.1 pour résoudre un problème de caricature d'image orienté région. La hiérarchie est une hiérarchie de régions obtenue par un algorithme classique de fusion de régions et la caricature est posée comme un problème de débit/distorsion opérationnel. Nous détaillerons ci-dessous leur approche à laquelle nous devons l'interprétation actuelle de notre méthode multi-échelle en termes de problème d'optimisation sous contrainte.

Muchnik et Mottl (1998) proposent une méthodologie pour résoudre différents problèmes de vision par ordinateur (segmentation, stéréo-vision,...) qui s'appuie également sur l'application du principe de programmation dynamique à des structures arborescentes. Cependant, les arbres qu'ils considèrent sont extraits du *graphe d'adjacence spatiale* des pixels de l'image, alors que ceux que nous considérons sont extraits du *treillis des parties* de l'image. L'idée de Muchnik et Mottl correspond globallement à approcher la solution d'un problème d'optimisation surfacique (2D) par la combinaison des solutions optimales de problèmes d'optimisation sur des arbres (1D) recouvrant la surface («spanning trees» et plus précisément dans ce cas «space filling trees»).

Nous venons donc de voir que pour le problème de partitionnement par minimisation d'énergie, restreindre l'espace de recherche au sous-espace des partitions correspondant à des coupes dans une hiérarchie permet de minimiser *exactement* des énergies séparables et ce *en temps linéaire* par rapport à la taille de

l'image.

La méthode de coupe minimale par programmation dynamique est soumise à deux conditions : séparabilité de la fonction de coût et organisation hiérarchique des parties élémentaires pouvant composer les solutions. Deux questions se posent donc naturellement :

1. Quelles fonctions de coût sur les partitions sont séparables, et que permettent-elles de modéliser ?
2. Quelle est l'ampleur de la restriction de l'espace complet des partitions à l'espace des coupes d'une hiérarchie ?

Les sections 7.1.2 et 7.1.3 qui suivent examinent successivement ces deux questions.

7.1.2 Modèles énergétiques séparables

Commençons par voir que la plupart des énergies proposées pour la segmentation sont séparables, puis expliquons à quelle classe de modèles ce type d'énergie correspond.

Fonctionnelle de Mumford-Shah

Considérons à titre d'exemple la fonctionnelle de Mumford et Shah (1989). Rappelons que cette fonctionnelle s'exprime

$$E(f,\Gamma) = \mu^2 \iint_{\mathcal{D}} (f-I)^2 dxdy + \iint_{\mathcal{D}\setminus\Gamma} ||\nabla f||^2 dxdy + \nu \int_{\Gamma} ds \tag{7.3}$$

où le domaine $\mathcal{D}$ de l'image I est décomposé en $R_1 \cup \cdots \cup R_n \cup \Gamma$. Γ est l'ensemble des frontières de la partition et $\int_\Gamma ds$ représente sa longueur totale.

Si l'on considère une image I sur un domaine $\mathcal{D}$ *ouvert*, Γ correspond alors à l'ensemble des frontières communes à deux régions, i.e.

$$\Gamma = \bigcup_{(i,j)} \delta R_i \cap \delta R_j.$$

On exclue ainsi la frontière $\delta\mathcal{D}$ entre l'image et son «extérieur». Chaque élément linéaire de frontière (à l'exception des «jonctions» ponctuelles, mais qui sont de longueur totale nulle) est alors partagé par exactement deux régions. Comme la mesure de longueur d'une courbe est additive sur les unions de courbes disjointes, on peut réécrire 7.3

$$\begin{aligned} E(f,\Gamma) &= \sum_{i=1}^{n} \Big(\mu^2 \iint_{R_i} (f-I)^2 dxdy \quad + \quad \frac{\mu^2}{2} \int_{\delta R_i \cap \mathcal{D}} (f-I)^2 dxdy \\ &+ \iint_{R_i} ||\nabla f||^2 dxdy \quad + \quad \frac{\nu}{2} \int_{(\delta R_i)\cap\mathcal{D}} ds \Big). \end{aligned} \tag{7.4}$$

Cette séparation de la fonctionnelle de Mumford et Shah impose que Γ ne comprenne pas les éléments de la frontière du domaine, qui ne sont adjacents qu'à une seule région. En continu, il suffit de poser le problème sur un ouvert. Remarquons que, de toutes façons, l'exclusion des frontières du domaine est insensible car leur contribution énergétique est identique pour toutes les partitions.

Modèle séparable général

Plus généralement, le modèle séparé suivant couvre une classe importante d'énergies couramment utilisées pour la segmentation :

$$\mathsf{E}(P) = \sum_{R\in P} \mathsf{E}_I(R) + \mathsf{E}_F(\delta R) \tag{7.5}$$

avec P une «partition» en régions (en continu : «pseudo-partition» à la Mumford-Shah).

E_I est un terme d'énergie interne des régions, et E_F est terme d'énergie liée aux frontières des régions.

Dans les formulations continues, E_I est une intégrale bidimensionnelle sur la surface des régions et E_F est une intégrale monodimensionnelle, suivant les frontières des régions. Dans la fonctionnelle de Mumford et Shah, E_I comporte une composante d'attache aux données (deux premiers termes de 7.4) et un terme de régularité de la solution (troisième terme de 7.4) De son côté, E_F est uniquement un terme de régularisation, indépendant de l'image, lié à la géométrie de la partition (quatrième terme). E_F peut également comporter un terme d'attache aux données, par exemple dépendant du gradient de l'image le long des frontières. Li (1995) discute différents termes de régularité des frontières d'une partition dits «adaptatifs aux discontinuités de l'image».

Les énergies employées dans les formulations variationnelles des méthodes d'ensembles de niveau ou de snakes classiques (qui portent sur des k-partitions ou seulement sur des courbes) peuvent se mettre sous la forme séparée 7.5.

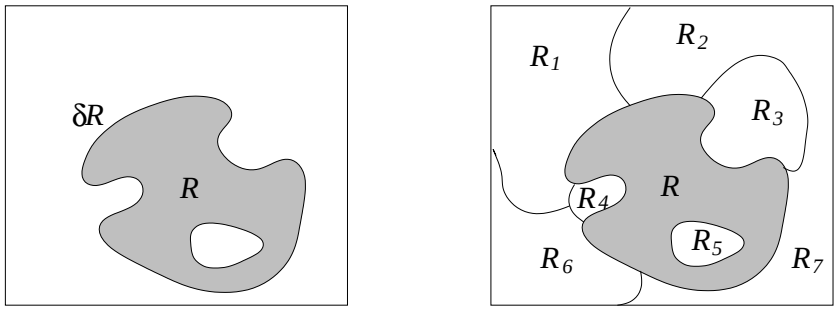
Dans le domaine markovien pixellaire, les potentiels associées à la labellisation Bayésienne de champs de Markov du premier ordre sont généralement séparables (Boykov et al. 1999). C'est le cas par exemple du modèle de Potts (Geman et Geman 1984). Les énergies considérées par Wang (1998) dans son modèle de relaxation stochastique sur les partitions à composantes connexes sont également séparables.

Enfin, toutes les approches par codage minimal (MDL) que nous connaissons aboutissent à des longueurs de description séparables, citons les modèles de Leclerc (1989a), Kanungo et al. (1995)[2], Lee (2000) et Fuchs (2001).

La fonction de coût 7.5 peut également être vue comme une formulation générale d'un modèle mixte régions/frontières, parfois appelé modèle «hybride» (Haris et al. 1998).

Modèles non contextuels

Fondamentalement, toutes ces formulations aboutissent à des énergies séparables parce qu'elles considèrent - plus ou moins explicitement - un *modèle non-contextuel*, c'est-à-dire un modèle pour lequel la qualification des différentes régions d'une image peut se faire indépendament les unes des autres (voir figure 7.3). En termes probabilistes, l'image est modélisée comme une union de processus régionaux indépendants, la probabilité de l'union est alors le produit des probabilités des sous-images (nous voyons donc que les modèles Markoviens pixellaires de Potts peuvent s'interpréter comme modélisant des régularités géométriques *propres* aux objets plutôt que des relations *entre* objets).



(a) (b)

FIG. 7.3 – Modèles énergétiques contextuels et non-contextuels. Dans un modèle non contextuel (a) l'énergie d'une région est indépendante des autres régions de la partition. L'énergie d'une région R peut prendre en compte des mesures surfaciques, concernant l'intérieur, l'extérieur ou le voisinage de R, ainsi que des mesures géométriques ou radiométriques sur sa frontière, et des propriétés topologiques. Les modèles contextuels (b) nécessitent a priori la connaissance de l'ensemble de la partition pour qualifier chaque région. Souvent, seule la structure locale de la partition est considérée (i.e. des voisinages d'ordre faible).

[2]L'énergie de codage de frontières de Kanungo et al. n'est a priori pas séparable mais pour une raison assez subtile. Ils la traitent cependant comme si elle l'était et ne semblent pas avoir vu le problème. Nous en parlerons ci-dessous.

Le caractère global de la solution intervient alors dans la contrainte de déterminer une *partition* de l'image en régions. C'est par la coopération - ou compétition - des différentes parties élémentaires qu'une certaine forme de contexte est introduite. Ce type de relation contextuelle ne permet cependant pas d'incorporer de préférence pour certains agencements globaux réguliers, c'est-à-dire pour la formation de Gestalts à un niveau supra-régional.

Remarque 1 : Il est important de ne pas confondre interprétation non contextuelle et interprétation «aveugle» à l'image à l'extérieur de la région analysée. On peut très bien incorporer dans l'énergie d'une région R des mesures liées à son complémentaire $\overline{R}$. En général, une région doit être vue comme une bi-partition de l'image et non comme une sous-image.

Remarque 2 : Autoriser les régions non connexes permet d'introduire des relations, par exemple géométriques, entres des unités distantes. Toutefois, il s'agit de relations entres composantes d'un seul objet, considéré comme un tout dans l'analyse. Nous verrons ci-dessous qu'une des perspectives du modèle multi-échelles proposé est de réintroduire certaines relations contextuelles par couplage des solutions entre les différentes échelles.

A contrario, les modèles contextuels ne sont pas séparables. Ils sont cependant moins nombreux.

On trouve toutes les formulations markoviennes qui mettent en jeu des probabilités conditionnelles dépendant des contenus des régions, ou les champs de Markov impliquant des voisinages d'ordre supérieur à un. Ces modèles sont cependant rarement considérés. En effet, dans un scénario non supervisé, il est peu vraisemblable qu'il existe certaines adjacences privilégiées entre couleurs ou textures, qui induiraient des conditionnement non uniformes. Par ailleurs, les champs de markov d'ordre supérieur à un sont rarement utilisés pour des raisons pratiques de difficulté de simulation en temps raisonnable.

Dans le domaine variationnel, le seul exemple à notre connaissance de modèle non séparable est la version affine-invariante de la fonctionnelle de Mumford-Shah de Ballester et al. (1994). Dans ce modèle, l'énergie associée à la géométrie de la frontière d'une région dépend de la géométrie de l'ensemble des courbes de la partition. Nous détaillerons cette énergie en 9.2.6 p. 236.

Notons que l'utilisation d'une énergie séparable est souvent tout simplement liée au problème technique de sa minimisation, même approximative. Un schéma d'optimisation combinatoire, pour être rapide (ou tout simplement faisable), nécessite que de faibles modifications d'une hypothèse de solution n'entrainent que de faibles calculs pour actualiser l'énergie, par exemple qu'un mouvement local n'entraine que des variations locales d'énergie. La séparabilité conduit à des schéma d'actualisation locaux : seule l'énergie des régions qui ont changé nécessite d'être recalculée[3].

Deux degrés de séparabilité d'une énergie de frontières

Énergies de frontières fortement séparables. Nous appelons énergies **fortement séparables** les énergies de frontière E_F que l'on peut décomposer en termes énergétiques portés par les portions de frontières séparant des régions voisines, i.e. qui s'écrivent

$$\mathsf{E}_F(P) = \frac{1}{2} \sum_{R \in P} \sum_{S \in P} \mathsf{F}(\delta R \cap \delta S). \tag{7.6}$$

Le facteur 1/2 rappelle que chaque arête de G sépare deux régions et que si l'on souhaite ne totaliser qu'une seule fois la contribution énergétique de chaque portion de frontière il faut diviser par 2 l'énergie totale de chaque région.

Pour que l'énergie totale soit effectivement séparable, il faut alors que F soit une fonction **additive** sur les portions de courbes disjointes, i.e. vérifie

$$\forall f_1, f_2 \mid f_1 \cap f_2 = \emptyset \qquad \mathsf{F}(f_1 \cup f_2) = \mathsf{F}(f_1) + \mathsf{F}(f_2). \tag{7.7}$$

[3] Attention, en aucun cas nous n'insinuons la réciproque : des énergies non séparables peuvent admettre des schéma d'actualisation rapide, c'est par exemple ce qui rend utilisable la fonctionnelle de Ballester et al. (1994).

Cette condition est illustrée par la figure 7.4.

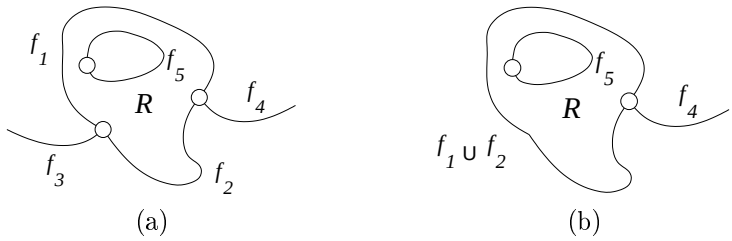


FIG. 7.4 – Les énergies locales non additives ne conduisent pas à des énergies de partitions séparables. Voir texte.

Sur cette figure, pour la partition (a), l'énergie de R vaut $\mathsf{F}(f_1) + \mathsf{F}(f_2) + \mathsf{F}(f_5)$ et pour la partition (b) elle vaut $\mathsf{F}(f_1 \cup f_2) + \mathsf{F}(f_5)$. Si F n'est pas additive, l'énergie de R n'est pas la même selon que la frontière f_3 entre deux régions *autres que* R existe ou non. Le modèle est donc *contextuel* (d'ordre 1) : l'énergie d'une région dépend de la configuration de ses voisines.

On voit alors que la seule façon de satisfaire 7.7 pour tous les morcellements possibles d'une courbe f est que F s'exprime sous la forme

$$\mathsf{F}(f) = \sum_{x \in f} \mathsf{L}(f, x). \tag{7.8}$$

où $\mathsf{L}(f, x)$ représente la contribution énergétique de la portion élémentaire x de la courbe f. En discret $\mathsf{L}(f, x)$ représente la contribution d'un pas élémentaire de frontière pixellaire ou interpixellaire. En continu, c'est celle d'un élément infinitésimal de f. Les énergies géométriques fortement séparables reposent donc nécessairement sur l'aggrégation de caractéristiques *locales* aux courbes.

L'archétype des énergies fortement séparables est l'énergie de longueur employée dans la fonctionnelle de Mumford-Shah, le potentiel de Potts ou le codage de Freeman de frontières discrètes.

Réciproquement, ce n'est pas parce qu'une énergie de courbe F s'exprime comme une somme de caractéristiques locales le long de la courbe (eqn. 7.8) qu'elle conduit à une énergie de partition (eqn. 7.6) séparable. La condition essentielle est l'additivité (eqn. 7.7). La courbure totale (somme des courbures en chaque point) est par exemple une mesure locale mais n'est pas additive : lorsque que l'on concatène deux portions de frontières, il faut ajouter la courbure au point de jonction. On peut finalement se demander s'il existe d'autres mesures - pertinentes - que la mesure de longueur qui soient distribuables sur les frontières entre régions et qui conduisent à des énergies totales séparables...

La condition 7.7 d'additivité de F pour que l'énergie totale soit séparable est relativement insidieuse. En effet, on peut toujours définir une énergie F non additive sur les frontières de G et l'énergie totale d'une partition par 7.6. Mais il faut alors bien se rendre compte que l'énergie totale n'est pas séparable et la gérer comme telle. Kanungo et al. (1995) proposent ainsi un schéma de codage d'un graphe de frontières dans lequel le coût de codage d'une frontière du graphe est une fonction qui a un comportement sous-additif par concaténation de deux portions de frontières (à cause d'un terme de codage de la longueur de la frontière qui emploie un code «universel» de Rissanen pour les entiers, qui est approximativement en $\log |f|$, cf. equ. 6.12 p. 123). Très bien, mais les auteurs indiquent ensuite - en toute bonne foi - que la diminution du coût de codage de la partition lors d'une fusion de deux régions voisines correspond au coût de codage de la frontière séparant les deux régions. La figure 7.4 montre bien que ce n'est pas le cas : supprimer f_3 ne diminue pas seulement l'énergie totale de $\mathsf{F}(f_3)$ mais également de $\mathsf{F}(f_1) + \mathsf{F}(f_2) - \mathsf{F}(f_1 \cup f_2)$.

Énergies faiblement séparable. A contrario, une énergie de frontière qui *ne peut pas* s'exprimer sous la forme 7.6 sera dite **faiblement séparable**.

Plaçons-nous à nouveau dans une optique de codage, pour laquelle l'énergie géométrique correspond à la longueur d'une description de la partition P. Nous venons de voir que les énergies fortement séparables proviennent d'une description «*orientée frontières*» de P, qui consiste à décrire indépendamment chaque

portion connexe de frontière f séparant deux régions. Mais on peut également imaginer de faire une description «*orientée régions*» de P qui consiste à décrire indépendamment la géométrie complète de chaque région. Une telle description conduit a priori à une énergie qui ne peut pas se distribuer sur les frontières de P, i.e. qui s'écrit en dernier lieu

$$\mathsf{E}_F(P) = \sum_{R \in P} \mathsf{F}(\delta R). \tag{7.9}$$

Une description orientée régions est redondante : chaque frontière est décrite deux fois. Elle n'est donc pas économique du point de vue du coût de codage de la géométrie de P.

Toutefois, une telle description est *qualitativement* plus riche qu'une description orientée frontières. Elle permet de prendre en compte certaines régularités globales des formes analysées et d'introduire des modèles géométriques. Suivant cette idée, nous proposerons ci-dessous de mesurer la complexité géométrique d'une région à la complexité d'une approximation polygonale de sa frontière. Ce type de description privilégie les frontières rectilignes en leur attribuant une énergie plus faible qu'aux frontières sinueuses de même longueur.

Notons bien que l'on pourrait imaginer de faire une description *orientée frontières* de P qui utilise des modèles géométriques, e.g. qui décrive succintement les portions de frontières approximativement rectilignes. Toutefois, d'une part, comme nous l'avons vu l'énergie obtenue ne serait pas séparable, d'autre part les alignements entre deux portions de frontière séparées par une jonction ne seraient pas pris en compte. Typiquement, la «bonne continuation» que l'on peut observer en figure 7.4 entre les deux portions f_1 et f_2 de la frontière de R ne serait pas prise en compte dans la partition (a). Dans la partition (b), c'est la bonne continuation de $f_1 \cup f_2$ au niveau de sa fermeture, au point de jonction avec f_4, qui ne serait pas relevée - l'énergie serait la même que si la frontière se fermait en faisant un angle saillant. Une description orientée regions permet donc de quantifier la simplicité globale de la forme d'une région, en tenant compte de la façon dont les différents éléments de sa frontière s'agencent. Dans le cadre d'un modèle non contextuel, une telle description est bien plus naturelle qu'une description orientée frontières : elle est centrée sur les unités indépendantes qui nous intéressent, les régions.

Revenons pour conclure ce paragraphe sur la redondance globale d'une description orientée régions : supposons que l'on dispose d'un schéma de description orienté régions E_F qui attribue à P l'énergie 7.9. Un tel schéma décrit globalement deux fois chaque élément de frontière. Il est alors légitime de supposer que l'on pourrait trouver un schéma E'_F de description *non redondante* de P - pas nécessairement séparable - qui tire parti des mêmes régularités géométriques que le schéma E_F et tel que le coût d'une description de P par E'_F n'excède pas la moitié du coût d'une description par E_F. Autrement dit, on peut estimer le coût d'une description orientée régions non redondante de P à $\frac{1}{2}$ du coût donné par l'équation 7.9. Notons que contrairement au coefficient $\frac{1}{2}$ qui apparaît dans l'équation 7.6, le facteur $\frac{1}{2}$ a ici un sens statistique.

Finalement, nous préférons concevoir toutes les énergies séparables relatives à la géométrie d'une partition comme des énergies correspondant à des descriptions orientées régions de la partition, et les flanquer d'un facteur $\frac{1}{2}$ qui traduit la redondance globale de la description. Il se trouve alors que si l'énergie d'une région est proportionnelle à la longueur de sa frontière alors il existe effectivement un schéma de description orienté frontières qui induit la même énergie, une fois redistribuée sur les régions.

Invariance par focalisation

Pour finir sur les énergies séparables et les modèles non contextuels, voyons qu'ils sont intimement liés à une propriété d'invariance dont nous n'avons trouvé aucune trace dans la littérature sur la segmentation : l'invariance par changement de domaine d'étude, que nous appelons **l'invariance par focalisation**. Relevons toutefois que ce type d'invariance est crucial en tatouage d'image (digital watermarking). Un algorithme de tatouage doit être robuste à l'extraction de sous-parties (crop resistance), i.e. le filigrane numérique doit pouvoir être détecté même sur un petit extrait de l'image tatouée.

Soit $\mathcal{I}$ l'ensemble des images de $\mathcal{D}$ dans $\mathcal{V}$. Si $I \in \mathcal{I}$ est une image et $A \subset \mathcal{D}$ est une partie du domaine de I, nous notons I_A l'extrait de I de domaine A.

Considérons un algorithme P de partitionnement d'image. À une image I définie sur $A \subset \mathcal{D}$ qui lui est fournie, P fait correspondre une partition $P(I) \in \mathbb{P}(A)$.

Notons - abusivement - $P \cap B \triangleq \{R \cap B\}_{R \in P}$ la *restriction* d'une partition P d'un domaine A à un domaine B.

Définissons également la *restriction sans occultation* d'une partition P d'un ensemble A à un ensemble B. Cette restriction, que nous notons $P \overset{o}{\cap} B$, consiste à ne conserver de P que les ensembles qui sont entièrement inclus dans B (voir la figure 7.5). Formellement $P \overset{o}{\cap} B \triangleq \{R \in P | R \subset B\}$[4].

Notons finalement $X \ominus \mathcal{B}_r$ l'érosion morphologique d'un ensemble X par une boule $\mathcal{B}_r$ de rayon r.

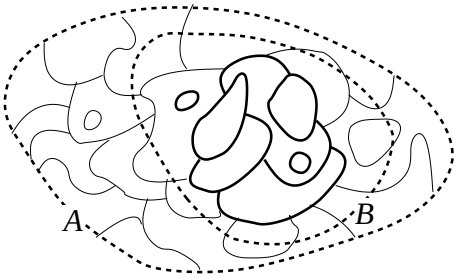


FIG. 7.5 – Restriction sans occultation d'une partition à un domaine. Les ensembles de la partition de A qui appartiennent à sa restriction sans occultation au domaine B sont représentés en gras.

Nous distinguons alors deux degrés d'invariance par focalisation :

• **Invariance forte par focalisation.**

Nous disons qu'un algorithme de segmentation est fortement invariant par focalisation quand, pour toute image I et tout couple de domaines (A, B), $P(I_A)$ et $P(I_B)$ coïncident sur leur domaine d'étude commun, $A \cap B$, sauf éventuellement sur une bordure de $A \cap B$ de taille fixe (indépendante des domaines et de l'image) et raisonnable (quelques pixels).

Formellement, $\exists \varepsilon \geq 0$ tel que $\forall I \in \mathcal{I} \quad \forall (A, B) \in \mathcal{P}^2(\mathcal{D})$:

$$P(I_A) \cap [(A \cap B) \ominus \mathcal{B}_\varepsilon] \quad = \quad P(I_B) \cap [(A \cap B) \ominus \mathcal{B}_\varepsilon] \,.$$

Cette propriété est vérifiée pour tous les algorithmes *locaux*, par exemple les algorithmes de segmentation qui s'appuient sur le résultat de filtres, tel qu'un algorithme de ligne de partage des eaux calculé sur un gradient de l'image. Elle est peu réaliste pour les algorithmes non locaux, par exemple qui utilisent des statistiques sur le contenu image des régions, au moins à cause des effets d'échantillon partiel.

• **Invariance faible par focalisation.**

Nous disons qu'un algorithme de segmentation est faiblement invariant par focalisation quand $\exists \varepsilon \geq 0$ tel que $\forall I \in \mathcal{I} \quad \forall (A, B) \in \mathcal{P}^2(\mathcal{D})$:

$$P(I_A) \overset{o}{\cap} [(A \cap B) \ominus \mathcal{B}_\varepsilon] \quad = \quad P(I_B) \overset{o}{\cap} [(A \cap B) \ominus \mathcal{B}_\varepsilon] \,.$$

Un tel algorithme donne des segmentations identiques pour tous les objets *intérieurs* au domaine d'étude commun, c'est-à-dire pour tous les objets dont l'une ou l'autre analyse n'a pas pu être biaisée par une visibilité partielle.

On sent bien que cette invariance faible est intimement liée à l'utilisation d'un modèle non contextuel, et donc dans un cadre énergétique, à l'utilisation d'une énergie séparable. En dépit de cette intuition,

[4] Cette opération peut se concevoir comme un filtrage morphologique par reconstruction : sont supprimés de P tous ses ensembles qui sont reconstructibles à partir du complémentaire de B.

nous n'avons pas pu prouver de résultat général concernant les modèles séparables, à part la propriété descendante triviale : si $P = \{R_j\}_{j\in J}$ est la partition qui minimise une énergie séparable E sur un domaine $\mathcal{D}$, alors pour tout $K \subset J$ le minimum de E sur le sous-domaine $\bigcup_{k\in K} R_k$ est la partition $\{R_k\}_{k\in K}$.

7.1.3 L'espace des coupes d'une hierarchie

Intéressons-nous maintenant à la dimension des espaces de partitions définis par les coupes d'une hiérarchie, c'est-à-dire au cardinal de $\mathbb{C}ut(H)$, où H est une hiérarchie.

$|\mathbb{C}ut(H)|$ dépend bien entendu de la structure de H. Par un argument de programmation dynamique analogue à celui de la proposition 5, on établit aisément que les tailles des ensembles de coupes des hiérarchies *partielles* de H obéissent à la récurrence suivante

$$|\mathbb{C}ut(H(x))| = \prod_{y\in\mathcal{F}(x)} |\mathbb{C}ut(H(y))| + 1. \tag{7.10}$$

Un examen rapide de cette récurrence permet de se convaincre que le nombre de coupes d'une hiérarchie :

- Décroit avec l'*arité* des relations père-fils. A l'extrême, la hiérarchie creuse à 2 niveaux, uniquement constituée d'un sommet et d'une base de taille N, reliés par une relation d'arité N, n'admet que 2 coupes : une à chaque niveau.
- Croît avec l'*équilibre* de la hiérarchie[5].

Hiérarchies binaires parfaitement équilibrées

D'après ce que nous venons de dire, les hiérarchies binaires sont celles qui sous-tendent le plus de coupes et parmi celles-ci, les mieux équilibrées sont les plus riches. Examinons-donc leur cas.

Soit X un ensemble dont le cardinal $N = 2^n$ est une puissance de 2. Considérons les hiérarchies sur X qui ont une structure d'*arbre binaire parfaitement équilibré* (la condition $N = 2^n$ est nécessaire pour que tous les nœuds puissent être équilibrés). Notons BE_N ces hiérarchies. Elles correspondent à une dichotomie itérée de X en ensembles de tailles égales (voir l'exemple de BE_8 ci-contre).

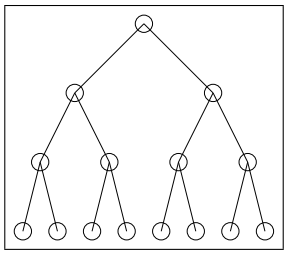


Fig. 7.6: Arbre binaire parfaitement équilibré BE_8

On a alors

Proposition 6

i) Les cardinaux $c_N = |\mathbb{C}ut(BE_N)|$, définis pour tout $N = 2^n$, $n \in \mathbb{N}$, obéissent à la récurrence :

$$\begin{cases} c_1 &= 1 \\ c_{2N} &= c_N^2 + 1 \end{cases}$$

ii) La suite $\sqrt[N]{c_N}$ converge et si β est sa limite alors $\forall N = 2^n > 0$

$$\beta^{N - \frac{4}{7\ln\beta}\frac{1}{N^2}} < c_N < \beta^N$$

Une évaluation numérique à 10^{-9} près donne $\beta = 1,502836801$.

Preuve :

i) Par construction, BE_{2N} est constituée de deux arbres partiels BE_N, dont les sommets sont les fils du sommet de BE_{2N}. Donc d'après la relation 7.10

$$|\mathbb{C}ut(BE_{2N})| = |\mathbb{C}ut(BE_N)|^2 + 1.$$

[5]Nous disons qu'une hiérarchie est équilibrée au nœud x quand les cardinaux des hiérarchies partielles portées par les fils de x sont tous égaux.

En outre, BE_1 admet une seule coupe.

ii) Posons $d_n = c_{2^n}$. d obéit à la récurrence

$$\left\{ \begin{array}{rcl} d_0 & = & 1 \\ d_{n+1} & = & d_n^2 + 1 \end{array} \right.$$

Définissons alors la suite u par

$$u_n = \frac{\ln d_n}{2^n}.$$

Comme $d_n^2 + 1 = d_n^2(1 + \frac{1}{d_n^2})$ et donc $\ln d_{n+1} = 2 \ln d_n + \ln(1 + \frac{1}{d_n^2})$, on a la récurrrence sur u :

$$\left\{ \begin{array}{rcl} u_0 & = & 0 \\ u_{n+1} & = & u_n + \frac{1}{2^{n+1}} \ln(1 + \frac{1}{d_n^2}) \end{array} \right.$$

qui donne

$$u_{n+1} = \sum_{k=0}^{n} \frac{1}{2^{k+1}} \ln(1 + \frac{1}{d_k^2}) = \frac{1}{2} \sum_{k=0}^{n} \frac{1}{2^k} \ln(1 + \frac{1}{d_k^2}).$$

Or $\forall k \geq 0$: $d_{k+1} \geq d_k^2$ et donc $d_k \geq 2^k$. On obtient alors la majoration de u_{n+1} :

$$u_{n+1} \leq \frac{1}{2} \sum_{k=0}^{n} \frac{1}{2^k} \ln(1 + \frac{1}{2^{2k}}) < \frac{1}{2} \sum_{k=0}^{n} \frac{1}{2^k} \times \frac{1}{2^{2k}} = \frac{1}{2} \sum_{k=0}^{n} \frac{1}{2^{3k}}.$$

u et croissante et majorée par une somme convergente. Elle converge donc.
Notons $\ln \beta$ sa limite. On a alors

$$\begin{array}{rcl} \ln \beta - u_n & = & \frac{1}{2} \sum_{k=n}^{\infty} \frac{1}{2^k} \ln(1 + \frac{1}{d_k^2}) \\ & < & \frac{1}{2} \sum_{k=n}^{\infty} \frac{1}{2^{3k}} = \frac{1}{2} \times \frac{1}{2^{3n}} \sum_{k=0}^{\infty} \frac{1}{2^{3k}} = \frac{1}{2} \times \frac{1}{2^{3n}} \times \frac{8}{7} = \frac{4}{7 \times 2^{3n}}. \end{array}$$

Par ailleurs $\ln \beta - u_n > 0$ et donc, en revenant à la suite d, on obtient l'encadrement

$$0 < \ln \frac{\beta^{2^n}}{d_n} < \frac{4}{7 \times 2^{2n}}$$

Soit en posant $N = 2^n$

$$\begin{array}{rrcccl} & 0 & < & \ln \frac{\beta^N}{c_N} & < & \frac{4}{7N^2} \\ \Leftrightarrow & 1 & < & \frac{\beta^N}{c_N} & < & e^{\frac{4}{7N^2}} \\ \Leftrightarrow & \beta^N e^{-\frac{4}{7N^2}} & < & c_N & < & \beta^N \end{array}$$

$\sqrt[N]{c_N}$ converge donc vers β et on a l'encadrement annoncé pour c_N.
Pour l'évaluation numérique de β, on utilise l'encadrement :

$$c_N^{1/N} < \beta < c_N^{1/N} e^{\frac{4}{7N^3}}$$

Pour $N = 2^{10}$, on obtient $c_N^{1/N} \simeq 1,502836801$ qui minore β au moins à $1,51 \times (e^{\frac{4}{7 \times 2^{30}}} - 1) < 10^{-9}$ près. ■

Le nombre de partitions contenues dans une hiérarchie binaire parfaitement équilibrée est donc exponentiel en fonction de la taille de l'ensemble sur lequel la hiérarchie est bâtie.

Hiérarchies binaires totalement déséquilibrées

À l'opposé des hiérarchies qui viennent d'être vues, les *arbres binaires totalement déséquilibrés*, notés BD_N, correspondent aux hiérarchies binaires qui admettent le moins de coupes possibles. Ces hiérarchies sont constituées d'une unique suite d'ensembles emboîtés (BD_8 est représenté ci-contre). Remarquons, que contrairement aux BE_N, les BD_N existent pour tout N. On voit alors que ces hiérarchies comportent exactement N niveaux et que chaque coupe possible est associée à un niveau de manière univoque. Pour ce type de hiérarchies, on a donc

$$|\mathbb{C}ut(BD_N)| = N.$$

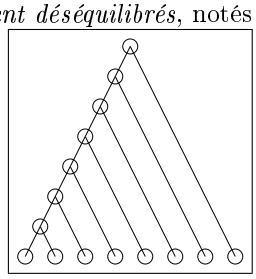


Fig. 7.7: Arbre binaire totalement déséquilibré BD_8

Dimensions d'autres espaces de partitions

Nous venons de voir que les espaces de coupes de hiérarchies sont potentiellement très riches : ils peuvent contenir un nombre de partitions exponentiel en fonction du nombre de *parties* différentes entrant dans la composition de ces partitions (e.g. pour un arbre BE_N, $2N-1$ parties permettent de former de l'ordre de $1,5^N$ partitions distinctes).

Comparons maintenant la richesse de ces espaces à celle d'autres espaces de partitions.

L'objectif de cette comparaison est simple : il est évident que, sans hypothèse supplémentaire sur l'énergie considérée, la minimisation *exacte* d'une énergie séparable dans un espace de partitions donné nécessite *au moins* l'évaluation de l'énergie individuelle de chacune des parties distinctes entrant dans la composition des partitions de l'espace. L'optimisation d'une énergie générique - même séparable - sur le treillis complet des partitions est donc intrinsèquement impraticable, cet espace comportant un nombre exponentiel de *parties*. Insistons sur le fait que dans ce cas, la combinatoire du problème ne réside pas dans la complexité de l'algorithme d'optimisation mais dans la combinatoire des données d'*entrée* de l'algorithme (notons que du point de vue de la complexité algorithmique, l'algorithme de coupe minimale par programmation dynamique réalise la borne minimale de complexité possible, en étant linéaire en fonction du nombre de parties).

Partitions – Les cardinaux de $\mathbb{P}(X)$ en fonction de $n = |X|$ sont donnés par les nombres de Bell, B_n. Leur expression la plus connue est due à Dobiński (1877), cf. e.g. Simon (1985) :

$$B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{n!}.$$

La rubrique Bell Numbers du CRC concise encyclopedia of mathematics (Weisstein 1998) indique la formule de Comtet (1974), qui utilise une somme finie :

$$B_n = \left\lceil \frac{1}{e}\sum_{k=1}^{2n}\frac{k^n}{n!} \right\rceil,$$

et, entre autres, l'expression asymptotique de Bruijn (1958) :

$$\frac{\ln B_n}{n} = \ln n - \ln\ln n - 1 + \frac{\ln\ln n}{\ln n} + \frac{1}{\ln n} + \frac{1}{2}\left(\frac{\ln\ln n}{\ln n}\right)^2 + \mathcal{O}\left(\frac{\ln\ln n}{(\ln n)^2}\right). \tag{7.11}$$

Appuyons-nous sur cette dernière pour trouver un encadrement simple de B_n.

Les deux premiers termes de 7.11 divergent et le reste

$$R_n = \frac{\ln\ln n}{\ln n} + \frac{1}{\ln n} + \frac{1}{2}\left(\frac{\ln\ln n}{\ln n}\right)^2$$

tend vers 0 en décroissant. En particulier, R_n devient vite inférieur à l'unité. Donc, pour n assez grand (en pratique dès les premiers rangs) :

$$\begin{array}{rrcccl} & \ln n - \ln\ln n - 1 & \leq & \frac{\ln B_n}{n} & \leq & \ln n - \ln\ln n \\ \Leftrightarrow & \left(\frac{n}{e\ln n}\right)^n & \leq & B_n & \leq & \left(\frac{n}{\ln n}\right)^n \end{array}$$

(B_n se rapproche de la borne inférieure de cet encadrement mais extrêmement lentement, e.g. $R_{10^{10}} \sim 0,2$, $R_{10^{100}} \sim 0,03$ (valeurs données par le logiciel Maple)).

Labellisations – De son côté le nombre d'images de labels possibles sur un domaine de taille N est N^N, ce qui est donc supérieur à la taille de l'ensemble des partitions possibles de l'image. En effet, chaque partition de taille M est associée à C_N^M labellisations équivalentes. L'approche par labellisation plonge donc le problème dans un espace de solutions plus grand que nécessaire.

Partitions en ensembles connexes – Désignons par $\mathbb{P}_C(X)$ l'ensemble des partitions en composantes connexes d'un ensemble X discret muni d'une topologie. Wang (1998) a alors judicieusement remarqué que si l'ensemble X possède une topologie plane, alors

$$|\mathbb{P}_C(X)| < 4^{|X|}.$$

L'argument provient du célèbre théorème du «coloriage en 4 couleurs d'un graphe planaire» [6]. En effet, si X est à deux dimensions alors le graphe d'adjacence topologique des régions d'une partition de X est planaire. On peut donc colorier ses noeuds en n'utilisant que 4 couleurs, de telle manière que deux noeuds voisins soient toujours coloriés avec deux couleurs différentes. Toute partition en ensembles connexes de X peut donc être représentée par un étiquetage (coloriage) des points de X qui n'utilise que 4 étiquettes (couleurs) différentes. Les régions de la partition sont alors définies comme les composantes connexes de points de même étiquette.

Notons que $4^{|X|}$ est un majorant : plusieurs 4-coloriages fournissent la même partition (au moins ceux obtenus par permutation de $(1,2,3,4)$, mais pas seulement. Les trous peuvent par exemple être coloriés avec n'importe laquelle des 3 couleurs complémentaires de la couleur de l'ensemble les entourant).

La restriction de l'espace de recherche de l'ensemble des partitions quelconques à celui des partitions en régions connexes s'accompagne donc d'un changement de classe de complexité combinatoire du problème (d'hyper-exponentiel (N^N) à exponentiel (a^N)). On voit donc tout l'intérêt pratique qu'il y a considérer le problème de partitionnement en régions connexes. C'est d'autant plus important quand on *sait* que la solution optimale du problème se trouve dans l'espace des partitions en régions connexes (e.g. le minimum global de la fonctionnelle constante par morceaux de Mumford et Shah).

Tronçonnement d'arbres recouvrants – Une manière de partitionner un graphe en composantes connexes est de considérer les *tronçonnements d'un arbre recouvrant* ce graphe. En effet, si $T = (X, V)$ est un arbre alors toute élimination d'une arête de T le *déconnecte* (il s'agit même d'une des multiples caractérisations d'un arbre, cf. p.ex. Jungnickel (1999, p.99)). Donc si T recouvre un graphe $G = (X, U)$ connexe, i.e. s'il est également connexe, alors toute sélection d'un sous-ensemble E d'arêtes de T partitionne X en $|E| + 1$ composantes connexes. Si $|X| = N$, T possède $N - 1$ arêtes, et les sous-ensembles d'arêtes sont en nombre 2^{N-1}. Comme deux sous-ensembles d'arêtes distincts définissent deux partitions distinctes, l'ensemble des partitions de G induites par les tronçonnements de T est de taille 2^{N-1}.

Si G est un graphe d'adjacence de régions d'une image, dont les arêtes sont attribuées par une dissemblance, un choix naturel pour T est l'arbre recouvrant de poids minimal de G, dont nous avons parlé à plusieurs reprises au chapitre 5.

[6] Wang l'appelle encore conjecture (la conjecture a été faite par Guthrie en 1853), ne sachant apparemment pas que le théorème a été démontré en 1977 - à grands renforts de moyens informatiques ce qui à l'époque a suscité de nombreuses controverses sur la validité de la preuve - voir par exemple (Jungnickel 1999, p.239).

Évaluons maintenant le nombre de *parties* entrant dans la composition de tels espaces de partitions. Pour cela, ordonnons T sous forme d'une arborescence en sélectionnant une racine arbitraire. L'ensemble des parties connexes qui contiennent le sommet de l'arborescence T correspond alors bijectivement à l'ensemble des coupes de T : chaque union des nœuds supérieurs à une coupe définit une partie connexe distincte. De la même manière, on construit l'ensemble des parties connexes que l'on peut obtenir par tronçonnement de T en considérant toutes les arborescences partielles de T et pour chacune d'elle toutes les parties connexes contenant le sommet et déconnectées de l'amont de la hiérarchie. Le nombre de parties connexes ainsi défini est donc égal à la somme des nombres de coupes sur l'ensemble des arborescences partielles de T.

On voit donc que le nombre de parties délimitées par tronçonnement de T est supérieur à $\frac{N(N+1)}{2}$ (cas où l'arbre est une chaîne), et devient rapidement exponentiel quand les ramifications de T augmentent. En particulier, le cas extrême où T est une arborescence binaire parfaitement équilibrée produit un nombre de parties supérieur à $\sqrt{2}^N$ (qui représente à lui seul le nombre de parties connexes contenant la racine).

Partitions d'hypergraphes – Sous l'angle des graphes, le problème de minimisation d'une énergie séparable se formalise comme un problème de partitionnement de poids minimal d'un hypergraphe (ensemble de parties d'un ensemble). Sous sa forme générale, ce problème est notoirement NP-difficile (Gondran et Minoux 1995, p.395-). Donc, même si le nombre de parties élémentaires considérées est raisonnable (polynomial en fonction de N), leur combinaison en une partition de coût minimal est dans le cas général infaisable dans la pratique (à moins que P=NP, ce qui semble peu probable à la plupart des spécialistes).

Hypergraphes d'intervalles - Pyramides – Il existe cependant une classe d'hypergraphes particuliers, appelés *hypergraphes d'intervalles*, pour lesquels les problèmes de partitionnement admettent des algorithmes de résolution praticables. «On appelle hypergraphe d'intervalles un hypergraphe [ensemble de parties] dont les sommets correspondent à des points sur une droite et les arêtes [parties] à des intervalles de cette droite» (Gondran et Minoux 1995, p.401). Cette correspondance entre arêtes et segments mono-dimensionnels permet de munir la structure d'une relation d'ordre partiel (l'ordre 1D entre segments disjoints), et le problème de partitionnement de poids minimal se transforme en un problème de plus court chemin dans un graphe orienté sans cycle (cf. fig. 7.8(b)). Le problème se résout alors par un algorithme de programmation dynamique de complexité linéaire en fonction du nombre de parties (algorithme de Michel Gondran 1977, reporté dans (Gondran et Minoux 1995, p.402-403)).

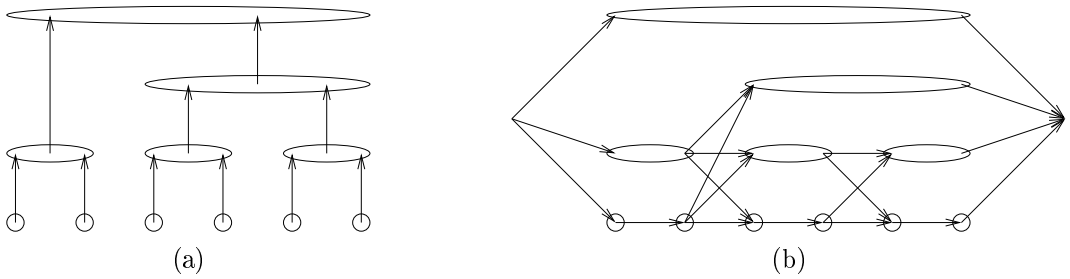


FIG. 7.8 – Graphes considérés pour la programmation dynamique hiérarchique et pour le partitionnement d'un hypergraphe d'intervalles.

Les ensembles hiérarchisés sont des cas particuliers d'hypergraphes d'intervalles (l'algorithme de représentation planaire d'un dendrogramme donné p. 65 injecte les nœuds de X dans $\langle 0, |X| \langle$ de telle manière que les parties de la hiérarchie correspondent à des intervalles). Notons toutefois que la méthode d'optimisation par programmation dynamique hiérarchique ne coïncide pas avec la méthode d'optimisation par plus court chemin de Michel Gondran. L'ordre partiel considéré par Gondran est en quelque sorte

«orthogonal» à l'ordre hiérarchique, comme l'illustre la figure 7.8. Le nombre de coupes de la hiérarchie correspond au nombre de chemins dans le «graphe de Gondran» correspondant.

Les hypergraphes d'intervalles sont appelés pyramides dans le domaine de la classification arborée (Batbedat 1990) et sont effectivement présentées comme des généralisations des structures hiérarchiques (le clustering pyramidal a été introduit par Diday (1984)). Une pyramide s'appuyant sur un ensemble de taille N comporte au plus $\frac{1}{2}N(N+1)$ éléments, ce qui est tout simplement réalisé quand les parties correspondent à tous les intervalles de $\langle 1, N\rangle$. On retrouve alors le cas du tronçonnement d'une chaîne qui a été vu ci-dessus.

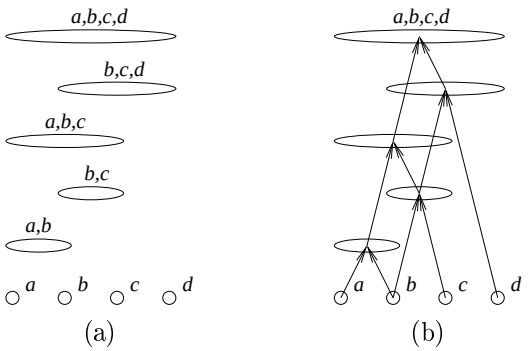


FIG. 7.9 – Une 2-3-hiérarchie et le graphe de couverture correspondant. Les ensembles $\{a, b\}$ et $\{b, c\}$ d'une part, et $\{a, b, c\}$ et $\{b, c, d\}$ d'autre part ont une intersection propre. Dans le graphe, les nœuds correspondant à $\{b\} = \{a, b\} \cap \{b, c\}$ et à $\{b, c\} = \{a, b, c\} \cap \{b, c, d\}$ ont deux pères (notons que les intersections propres ne sont pas nécessairement des éléments de la 2-3-hiérarchie, i.e. on pourrait supprimer $\{b, c\}$ de la structure ci-dessus, ce serait alors ses fils qui possèderaient deux pères).

2-3-hiérarchies – Les pyramides les plus simples sont appelées 2-3-hiérarchies par Bertrand (2002). Dans une telle structure, chaque ensemble possède au plus une intersection propre avec un autre ensemble (voir fig. 7.9). Une hiérarchie, elle, n'autorise aucune intersection propre. Bertrand (2002) a récemment montré qu'une 2-3-hiérarchie batie sur un ensemble de taille N comporte au plus $\frac{5}{2}(N-1)$ parties (et également que les 2-3-hiérarchies sont des pyramides, ce qui est loin d'être évident). Ainsi, si l'on met à part les $N+1$ parties triviales d'une (2-3)-hiérarchie (base et sommet), autoriser une intersection propre permet de multiplier par $1,5$ le nombre potentiel de régions non triviales par rapport au cas où l'on interdit les intersections propres. Une application de l'algorithme de plus court chemin de Gondran à une 2-3-hiérarchie permettrait donc d'en extraire une partition optimale en temps linéaire. Nous n'avons pas évalué le nombre de partitions que l'on peut composer à partir des ensembles d'une 2-3-hiérarchie.

Bilan comparatif

La table 7.1 résume les tailles des différents espaces de partitions que nous venons de passer en revue, et les met en regard des nombres de parties élémentaires composant ces espaces.

Remarquons que pour les domaines plans, les tailles des espaces de coupes hiérarchiques sont de l'ordre d'une racine $n^{\text{ième}}$ de la taille de l'espace complet des partitions en composantes connexes, e.g. $\sqrt[4]{|\mathbb{C}ut(BE_N)|} > |\mathbb{P}_C(N)|$. En outre, il semble - intuitivement - difficile d'obtenir des espaces de partitions significativement plus grand sans simultanément significativement augmenter le nombre de parties de l'espace (notons que le cas du tronçonnement d'une chaîne - ou de façon équivalente le partitionnement d'un hypergraphe d'intervalle maximal - correspond à une topologie 1D).

Évidemment, les structures des hiérarchies qui s'appuient sur le contenu d'images dépendent de ce contenu (et évidemment aussi de l'algorithme de construction). Nous avons constaté expérimentalement

Structures ($\|X\| = N$)	$\|$Parties$\|$		$\|$Partitions$\|$	
Labellisations	$=$	2^N	$=$	N^N
Partitions	$=$	2^N	$<$	$(N/\ln N)^N$
Partitions en régions connexes d'un domaine plan		...	$<$	4^N
Partitionnements d'hypergraphes quelconques		...		...
Tronçonnements d'arbres recouvrants binaires équilibrés	$>$	$\sqrt{2}^N$	$=$	2^{N-1}
Tronçonnements d'arbres recouvrants quelconques		...		...
Tronçonnements de chaînes recouvrantes	$=$	$\frac{1}{2}N(N+1)$	$=$	2^{N-1}
Partitionnement d'hypergraphes d'intervalles	$\leq$	$\frac{1}{2}N(N+1)$	$\leq$	2^{N-1}
Coupes de 2-3-hiérarchies	$\leq$	$\frac{5}{2}(N-1)$		...
Coupes de hiérarchies binaires totalement équilibrées	$=$	$2N-1$	$>$	$\sqrt{2}^N$
Coupes de hiérarchies quelconques		...		...
Coupes de hiérarchies binaires totalement déséquilibrées	$=$	$2N-1$	$=$	N

TAB. 7.1 – Tailles de différents espaces de partitions. Pour différentes représentations induisant des espaces de partitions d'un ensemble de taille N, la seconde colonne indique le nombre de parties élémentaires de l'espace, et la troisième le nombre de partitions que l'on peut former par composition de ces parties (à part pour les labellisations, où le nombre indique le nombre de labellisations possibles).

que les hiérarchies binaires construites sur des images usuelles par les algorithmes proposés dans ce mémoire induisent des espaces de partitions dont la taille varie entre $1,3^N$ et $1,4^N$. La figure 7.10 présente une telle hiérarchie. Elle possède 120 régions de base et admet de l'ordre de $2,67 \times 10^{16}$ coupes, ce qui vaut approximativement $1,37^{120}$.

7.1.4 Optimisation hiérarchique et descente de gradient

Pour conclure cette section, nous comparons la méthode de minimisation d'énergie par programmation dynamique à la stratégie de sélection de solution employée dans les méthodes de fusion de régions par descente de gradient (FRD, voir 6.1.5 p. 129).

Une méthode de fusion itérée de couples de régions s'appuyant sur un critère de descente balaye évidemment une hiérarchie binaire de régions. Soit H cette hiérarchie et N la taille de sa base (la sur-partition initiale). Comme nous venons de le voir, H possède en général un très grand nombre de coupes. Cependant, un algorithme de FRD ne considère pas toutes les coupes possibles de la hiérarchie mais seulement N : la solution de l'algorithme est choisie parmi les N partitions exhibées au cours de la fusion (si l'algorithme ne s'arête pas à la première partition 2-normale rencontrée). Ces partitions correspondent au coupes naturelles de H pour l'ordre de fusion, qui est par définition un critère croissant. Une FRD considère donc implicitement une hiérarchie indicée. Si l'on met de côté cet indice, on voit alors que de nombreuses autres coupes sont possibles et que l'on peut, par programmation dynamique, optimiser l'énergie sur la totalité de cet espace, et ce sans surcoût de calcul. Cela revient à clairement séparer une étape de construction hiérarchique d'une étape de coupe. L'étape de coupe est alors *indépendante de l'ordre de construction* de H.

Intuitivement, l'intérêt de s'affranchir de l'ordre de construction de H est le suivant : si, du fait de l'ordre particulier de fusion, une région pertinente à un certain endroit de l'image commence à être fusionnée à son environnement tandis que d'autres régions pertinentes situées ailleurs dans l'image sont encore sur-segmentées, alors aucune coupe naturelle pour l'ordre de fusion ne peut contenir simultanément ces différentes régions pertinentes. En bref, les coupes naturelles «ratent» la bonne segmentation dès lors que la fusion est *asynchrone*, i.e. dès lors qu'au cours du processus de fusion des zones sous-segmentées apparaissent quand des zones sur-segmentées existent encore.

Ce phénomène est particulièrement frappant quand on considère un critère de fusion différent du critère de sélection finale de la coupe. C'est illustré par la figure 7.10 dans laquelle la hiérarchie a été

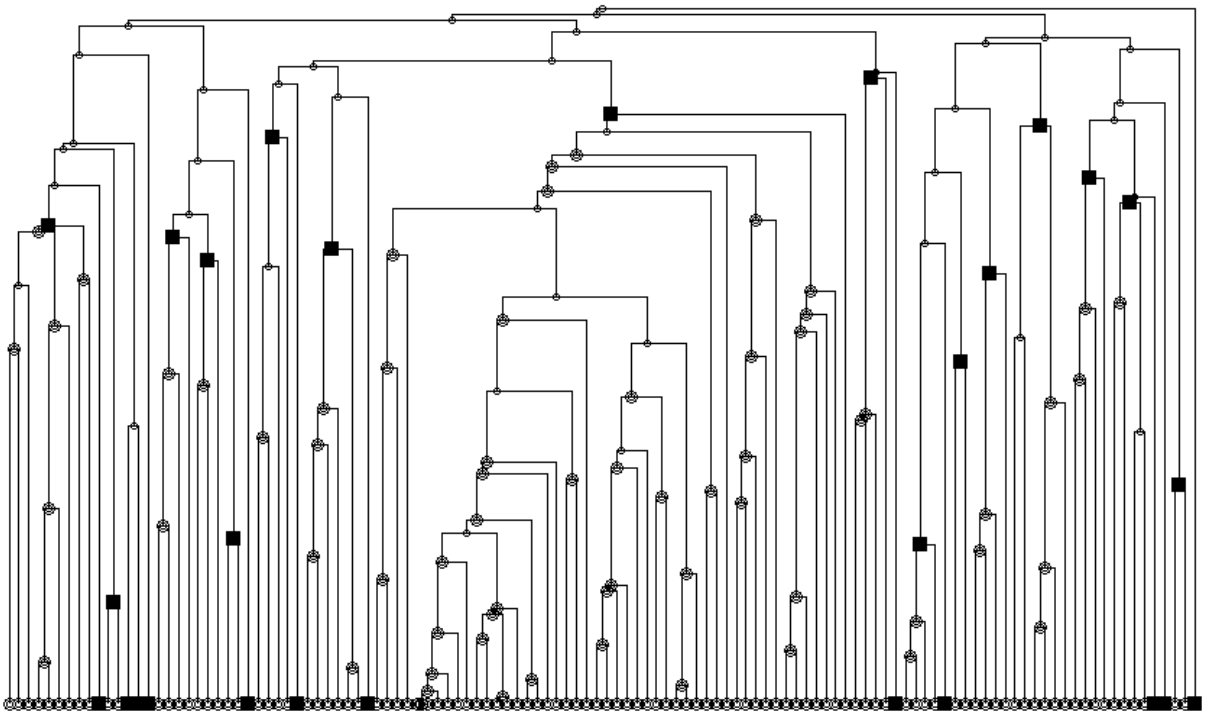


FIG. 7.10 – Séparation entre critère de fusion et critère de coupe. La hiérarchie dessinée a été obtenue sur une image 50×50 par complete linkage. Sa base contient 120 régions élémentaires. L'ordonnée des nœuds correspond à leur ordre de fusion. Les nœuds cerclés signalent les étapes de fusion qui font décroître une énergie séparable (MDL). Les carrés pleins figurent les nœuds appartenant à la coupe optimale pour cette énergie (i.e. les plus hauts nœuds cerclés).

construite par un algorithme de cocons durs (complete linkage) et la coupe sélectionnée par une énergie de type Minimum Description Length (MDL). Nous voyons que :

i) La coupe optimale pour l'énergie (carrés pleins) n'est pas une des 120 coupes naturelles pour l'ordre de fusion (horizontales dans la représentation 7.10). Rappelons qu'ici $\mathbb{C}ut(H) \simeq 2,67 \times 10^{16}$.

ii) L'énergie utilisée n'est pas monotone dans H : des étapes de décroissance locale alternent avec des étapes de croissance de l'énergie (sur la figure 7.10, les nœuds cerclés alternent avec les nœuds non cerclés).

Évidemment, quand le critère de fusion correspond à une descente sur la même énergie que celle qui sert à sélectionner la coupe, les coupes naturelles sont de «bonnes» coupes. Il ne faut toutefois pas s'arêter à la première segmentation 2-normale rencontrée. On peut en outre trouver des cas de fusion asynchrone. La coupe optimale est dans ce cas une coupe irrégulière pour l'ordre de fusion.

On a donc tout intérêt à concevoir la résolution de problèmes de partitionnement en deux étapes bien distinctes :

a) Construction d'une hiérarchie de régions.
b) Coupe dans cette hiérarchie s'appuyant soit sur un prédicat d'homogénéité (consistant ou non[7]), soit sur une énergie séparable.

[7] Salembier et Garrido (2000) proposent d'utiliser un algorithme de Viterbi pour sélectionner une coupe dans le cas où le prédicat considéré n'est pas consistant. Assimilons un prédicat «d'homogénéité» à une fonction booléene de H dans $\{0,1\}$. En deux mots, l'algorithme de Viterbi consiste alors à projeter le prédicat non consistant P sur le plus proche prédicat C consistant pour H, où la proximité de deux prédicats est mesurée par la distance de Hamming : $\sum_{x \in H} |P(x) - C(x)|$. La coupe retenue est alors la coupe homogène maximale pour C.

La première étape correspond à un élagage substantiel de l'ensemble des parties susceptibles d'entrer dans la solution : de 2^N parties on passe à un nombre linéaire, au pire $2N-1$. Ces parties sont structurées de telle manière que la deuxième étape puisse être résolue de manière exacte et efficacement.

Dans ce cadre, le problème de segmentation revient à définir un bon critère de groupement ou de division hiérarchique, qui exhibe - à un niveau quelconque - les régions pertinentes. Si les «bonnes» régions sont effectivement balayées durant le parcours hiérarchique de l'image, le succès de l'ensemble repose alors sur l'aptitude du critère de coupe à les sélectionner. Si l'on considère une coupe sur un critère de minimisation d'énergie, mises bout-à-bout, ces deux étapes correspondent à trouver l'optimum global de l'énergie sur un espace de partitions de taille honorable.

Dans (Guigues et al. 2001b) nous avons proposé d'utiliser un algorithme de lien moyen s'appuyant sur un test de Hotelling pour construire dans un premier temps une hiérarchie de cocons équilibrés (voir le chapitre 5), puis, dans un deuxième temps, d'en sélectionner une coupe minimale pour un critère MDL par l'algorithme de programmation dynamique.

Didier Boldo (2002) a également utilisé cette stratégie en deux étapes pour segmenter en régions planes un Modèle Numérique de Surface (MNS) obtenu par corrélation automatique[8]. La hiérarchie est construite par un critère de fusion ad hoc mélangeant un critère de gradient sur la frontière séparant les régions et un critère de colinéarité des vecteurs normaux à des plans de régression sur les valeurs des régions. L'énergie utilisée pour la coupe est une énergie MDL identique à celle précédemment proposée par Frank Fuchs (2001). Cette segmentation entre dans la chaîne proposée par Didier Boldo pour réaliser des ortho-images aériennes fines, convenablement redressées au niveau des discontinuités de la surface topographique. Cette chaîne est en cours d'évaluation pour la production industrielle d'ortho-images à l'IGN.

Nous citons ces propositions pour mémoire. Nous les considérons en effet aujourd'hui comme des étapes intermédiaires de notre cheminement, sortes d'hybrides entre les méthodes de cocons et la méthode purement énergétique que nous préconisons aujourd'hui. Lorsque nous avons fait ces propositions, nous concevions encore le problème de segmentation de bas niveau comme un problème de partitionnement et nous n'envisagions pas qu'un algorithme de segmentation s'arrête à la production d'une hiérarchie. Le fait de concevoir le problème de segmentation en deux étapes clairement distinctes - qui date de cette époque - constitue néanmoins le point clef autour duquel s'articule toute notre démarche actuelle.

Historiquement, après avoir proposé cette décomposition en deux étapes du problème de partitionnement, nous avons constaté empiriquement que

i) Les critères MDL que nous envisagions fournissaient globalement des sur-segmentations par rapport au niveau de découpage recherché (nous en donnerons un exemple au chapitre 9).

ii) En augmentant le poids des termes de complexité de modèle dans ces formulations MDL, on obtenait des partitions de plus en plus grossières.

Nous nous sommes par la suite rendus compte que le deuxième point correspond à un comportement relativement général, commun à la plupart des énergies employées en segmentation, comme nous proposons de le voir maintenant.

7.2 Familles de coupes minimales monotones

Intéressons-nous maintenant aux énergies affines séparables, i.e. aux énergies qui s'écrivent

$$\mathsf{E}_\lambda(P) = \sum_{R \in P} \lambda \mathsf{C}(R) + \mathsf{D}(R). \tag{7.12}$$

où C est une énergie de «régularisation» et D une énergie d'«attache aux données».

[8] Nous avons développé cette application avec Didier Boldo, en nous appuyant sur les programmes que nous avions écrit pour la segmentation d'images photométriques.

La donnée d'une telle énergie est équivalente à la donnée d'un couple (C, D) d'énergies portant sur les parties de l'espace, aussi noterons-nous souvent $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$.

Dans l'esprit de la section qui précède, nous nous intéressons à l'agencement des minima d'une énergie affine non pas sur l'espace complet des partitions mais sur l'espace des coupes d'une hiérarchie. Dans tout ce qui suit, nous considérons une hiérarchie H sur un ensemble X et une énergie affine E_λ données. Nous nous intéressons alors à la famille des coupes minimales de H pour E_λ, famille indexée par λ.

Définition 22
- *Pour tout $\lambda \in \mathbb{R}$, nous appelons **λ-coupe** la coupe qui minimise E_λ dans H. Nous la notons $C^*_\lambda(H)$.*
- *Nous notons $C^*(H) = \{C^*_\lambda(H)\}_{\lambda \in \mathbb{R}}$ la famille des λ-coupes de H pour E_λ.*
- *Une région $x \in H$ qui appartient à la coupe $C^*_\lambda(H)$ est dite **λ-optimale**.*

7.2.1 Condition de monotonie des coupes minimales

Optimalité partielle et maximalité

En examinant plus en détail la structure de l'algorithme de programmation dynamique pour l'optimisation d'énergies séparables décrit à la section précédente, on constate aisément que

Proposition 7 *$x \in H$ est λ-optimal si et seulement si il verifie les deux propriétés :*

*i) x est **partiellement optimal** dans H, i.e.*

$$\forall Y \in \mathbb{C}ut(H(x)) \qquad \mathsf{E}_\lambda(\{x\}) \leq \mathsf{E}_\lambda(Y).$$

*ii) x est **maximal** pour i) dans H : aucun $y \in H$ tel que $x \subset y$ n'est également partiellement λ-optimal.*

Rappelons que $H(x)$ représente la hiérarchie partielle issue de x et $\mathbb{C}ut(H(x))$ l'ensemble de ses coupes.

Remarquons l'inégalité *large* dans i) : si deux partitions ordonnées possèdent la même énergie, la plus grossière est considérée comme meilleure, ce qui est relativement naturel.

On notera la forte parenté entre les deux conditions d'optimalité données par la proposition 7 et les deux conditions d'optimalité pour les segmentations définies par un prédicat d'homogénéité (cf. p. 72).

Définition 23 *Nous notons $P^*_\lambda(H)$ l'ensemble des noeuds **partiellement** λ-optimaux de H, c'est-à-dire ceux qui vérifient la propriété i) de la proposition 7.*

Sous-additivité d'énergies de parties et décroissance d'énergies de partitions

Définition 24 *Une énergie de régions $\mathsf{E} : \mathcal{P}(X) \to \mathbb{R}^+$ qui vérifie*

$$\forall (x, y) \in \mathcal{P}^2(X) \quad x \cap y = \emptyset \quad \Rightarrow \quad \mathsf{E}(x \cup y) \leq \mathsf{E}(x) + \mathsf{E}(y). \tag{7.13}$$

*est appelée une énergie **sous-additive** (strictement quand l'inégalité est stricte).*

La sous-additivité d'une énergie de régions est équivalente à la **décroissance** dans $\mathbb{P}(X)$ de l'énergie séparable associée. Autrement dit,

Proposition 8 *$\mathsf{E} : \mathcal{P}(X) \to \mathbb{R}^+$ est sous-additive si et seulement si*

$$\forall (P, Q) \in \mathbb{P}^2(X) \qquad P \geq Q \quad \Rightarrow \quad \sum_{p \in P} \mathsf{E}(p) \leq \sum_{q \in Q} \mathsf{E}(q).$$

Preuve :

$\Rightarrow$ La propriété de sous-additivité 7.13 s'étend immédiatement à toute famille finie d'ensembles disjoints. Si E est sous-additive alors $\forall(P = \{P_1 \dots P_m\}, Q = \{Q_1 \dots Q_n\}) \in \mathbb{P}^2(X)$, $P \geq Q$ implique que les P_i s'écrivent $P_i = \bigcup_{j \in L_i} Q_j$, où $\{L_1 \dots L_m\}$ forme une partition de $\langle 1, n \rangle$. Et donc

$$\mathsf{E}(P) \quad = \quad \sum_{i=1}^{m} \mathsf{E}(\bigcup_{j \in L_i} Q_j) \quad \leq \quad \sum_{i=1}^{m} \sum_{j \in L_i} \mathsf{E}(Q_j) \quad = \quad \sum_{i=1}^{n} \mathsf{E}(Q_j) \quad = \quad \mathsf{E}(Q).$$

$\Leftarrow$ Soit E une énergie séparable décroissante sur $\mathbb{P}(X)$. Considérons alors pour tout $(x, y) \in \mathcal{P}^2(X)$ tels que $x \cap y = \emptyset$ les partitions $Q = \{x, y, Q_1 \dots Q_n\}$ et $P = \{x \cup y, Q_1 \dots Q_n\}$. On obtient

$$\begin{aligned} P \succ Q \quad \Rightarrow \quad P > Q \quad \Rightarrow \quad \mathsf{E}(x \cup y) + \sum_{i=1}^{n} \mathsf{E}(Q_i) \quad &\leq \quad \mathsf{E}(x) + \mathsf{E}(y) + \sum_{i=1}^{n} \mathsf{E}(Q_i) \\ \Leftrightarrow \quad \mathsf{E}(x \cup y) \quad &\leq \quad \mathsf{E}(x) + \mathsf{E}(y). \end{aligned}$$

■

Familles de solutions causales

L'ensemble de notre approche multi-échelles de la segmentation repose alors sur le théorème :

Théorème 3 ***Coupes minimales causales***
Soit H une hiérarchie et $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$ une énergie affine séparable sur un ensemble X. Si C est sous-additive alors l'ensemble $\{C^*_\lambda(H)\}_{\lambda \in \mathbb{R}}$ est causal, i.e.

$$\forall(\lambda_1, \lambda_2) \in \mathbb{R}^2 \qquad \lambda_1 \leq \lambda_2 \quad \Rightarrow \quad C^*_{\lambda_1}(H) \leq C^*_{\lambda_2}(H).$$

Preuve

Soit $\lambda_1 \in \mathbb{R}$. Explicitons la propriété de λ_1-optimalité partielle (proposition 7-i)) pour tout x de $P^*_{\lambda_1}(H)$:

$\forall x \in P^*_{\lambda_1}(H) \quad \forall Y \in \mathbb{C}ut(H(x))$:

$$\begin{aligned} \lambda_1 \mathsf{C}(x) + \mathsf{D}(x) \quad &\leq \quad \sum_{y \in Y} \lambda_1 \mathsf{C}(y) + \mathsf{D}(y) \\ \Leftrightarrow \quad \lambda_1 \left(\mathsf{C}(x) - \sum_{y \in Y} \mathsf{C}(y) \right) \quad &\leq \quad \sum_{y \in Y} \mathsf{D}(y) - \mathsf{D}(x). \end{aligned}$$

Comme C est sous-additive, et par définition Y est une partition de x, on a

$$\Delta\mathsf{C} \quad \triangleq \quad \mathsf{C}(x) - \sum_{y \in Y} \mathsf{C}(y) \quad = \quad \mathsf{C}(\bigcup_{y \in Y} y) - \sum_{y \in Y} \mathsf{C}(y) \quad \leq 0.$$

Donc, $\forall \lambda_2 \in \mathbb{R}$

$$\lambda_2 \geq \lambda_1 \qquad \Rightarrow \qquad \lambda_2 \Delta\mathsf{C} \quad \leq \quad \lambda_1 \Delta\mathsf{C} \quad \leq \quad \sum_{y \in Y} \mathsf{D}(y) - \mathsf{D}(x)$$

qui exprime la λ_2-optimalité partielle de x. On obtient donc une propriété de «causalité partielle» :

$$\forall(\lambda_1, \lambda_2) \in \mathbb{R}^2 \quad \lambda_2 \geq \lambda_1 \Rightarrow \left(\forall x \in H \quad x \in P^*_{\lambda_1}(H) \Rightarrow x \in P^*_{\lambda_2}(H)\right). \tag{7.14}$$

Examinons maintenant les conditions d'optimalité globale d'un noeud x.

La condition i) de la proposition 7 s'écrit également

$$\forall \lambda_1 \in \mathbb{R} \quad (\forall x \in H \quad x \in C^*_{\lambda_1}(H) \Rightarrow x \in P^*_{\lambda_1}(H)).$$

En utilisant alors l'équation de «causalité partielle» 7.14, on obtient

$$\forall (\lambda_1, \lambda_2) \in \mathbb{R}^2 \quad \lambda_2 \geq \lambda_1 \Rightarrow \left(\forall x \in H \quad x \in C^*_{\lambda_1}(H) \Rightarrow x \in P^*_{\lambda_2}(H)\right).$$

Or, si $x \in P^*_{\lambda_2}(H)$ alors aucun noeud $y \subset x$ partiellement λ_2-optimal n'est maximal (verifie la proposition 7ii)). En conclusion

$$\begin{array}{ll} & \forall (\lambda_1, \lambda_2) \in \mathbb{R}^2 \\ & \lambda_2 \geq \lambda_1 \Rightarrow \quad \forall x \in C^*_{\lambda_1}(H) \quad \forall y \in C^*_{\lambda_2}(H) \quad x \cap y \neq \emptyset \Rightarrow x \subset y \\ \Leftrightarrow & \lambda_2 \geq \lambda_1 \Rightarrow \quad C^*_{\lambda_2}(H) \geq C^*_{\lambda_1}(H). \end{array} \qquad \blacksquare$$

Réciproquement, on voit facilement que si C n'est pas sous-additive, on peut construire des hiérarchies et des énergies D pour lesquelles les coupes minimales ne sont pas monotones en fonction de λ. En conséquence,

Définition 25 *Une énergie affine $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$ possédant un terme C sous-additif est appelée une* ***énergie multi-échelles*** *et le paramètre λ est appelé le* ***paramètre d'échelle*** *de l'énergie.*

Notons que cet agencement remarquable des minima de E_λ n'était pas du tout acquis : comme nous l'avons montré précédemment les coupes possibles d'une hiérarchie sont très nombreuses et sont bien sûr loin d'être toutes ordonnées. En outre, la section 6.3.1 p. 144 a montré que les minima d'une énergie affine sur $\mathbb{P}(X)$ ne sont pas nécessairement emboîtés (notons que le terme de régularisation proportionnel à longueur totale de frontières qui est employé dans l'exemple 6.3.1 est sous-additif).

Remarque : le théorème 3 s'étend à d'autres types d'énergies que les énergies affines. En particulier, toutes les énergies de la forme $f(\lambda)\mathsf{C} + \mathsf{D}$ où f est une fonction croissante conduisent à des solutions monotones quand C est sous-additif. Le cas des énergies affines correspond à choisir l'identité pour f. Nous reparlerons ci-dessous du cas $f = \log$.

Examinons maintenant les conséquences de cette monotonie des coupes minimales.

Apparition, disparition et persistance de régions

Pour chaque région x de la hiérarchie, soit

$$\Lambda^*(x) \triangleq \{\lambda | x \in C^*_\lambda(H)\}.$$

$\Lambda^*(x)$ représente l'ensemble des échelles pour lesquelles x appartient à une coupe minimale.

Considérons également

$$\Lambda^*_\uparrow(x) \triangleq \{\lambda | x \in P^*_\lambda(H)\}$$

qui représente l'ensemble des échelles pour lesquelles x est partiellement optimal.

Sous les conditions du théorème 3, si une région x est partiellement optimale pour une échelle λ alors elle le reste pour toute échelle $\lambda' > \lambda$. Donc, $\forall x \in H$, l'ensemble $\Lambda^*_\uparrow(x)$ est un intervalle de la forme $[a, +\infty[$.

Définition 26 ***Échelle d'apparition***
La borne inférieure de l'intervalle $\Lambda^*_\uparrow(x)$ est appelée l'échelle d'apparition de la région x (dans une coupe optimale de H pour E_λ). Nous la notons $\lambda^+(x)$.

Par ailleurs, x est globalement optimal à l'échelle λ ssi il est également λ-maximal. Définissons donc

$$\Lambda^*_\downarrow(x) \triangleq \bigcup_{y \in H, x < y} \Lambda^*_\uparrow(y) \tag{7.15}$$

qui représente l'ensemble des échelles pour lesquelles x *ne peut pas être maximal* et donc ne peut pas être un optimum global. Comme $\Lambda^*_\downarrow(x)$ est une union d'intervalles de la forme $[a, +\infty[$, c'est aussi un intervalle de cette forme.

Définition 27 ***Échelle de disparition***
La borne inférieure de l'intervalle $\Lambda^*_\downarrow(x)$ est appelée l'échelle de disparition de la région x (d'une coupe optimale de H pour E_λ). Nous la notons $\lambda^-(x)$.

D'après l'équation 7.15, on a alors immédiatement :

Proposition 9 *Sous les conditions du théorème 3*

$$\lambda^-(x) = \min_{y \in H, x \subset y} \lambda^+(y). \tag{7.16}$$

Finallement, les deux conditions d'optimalité globale données par la proposition 7 montrent que

$$\Lambda^*(x) = \Lambda^*_\uparrow(x) \backslash \Lambda^*_\downarrow(x). \tag{7.17}$$

et donc

Proposition 10 *Sous les conditions du théorème 3, $\forall x \in H$ l'ensemble $\Lambda^*(x)$ est un intervalle de la forme*

$$\Lambda^*(x) = [\lambda^+(x), \lambda^-(x)[$$

*que nous appelons l'**intervalle de persistance** de x (dans une coupe optimale de H pour E_λ).*

Les intervalles de persistance caractérisent complètement l'ensemble des λ-coupes pour E_λ dans H. Les connaissant, on accède à n'importe quelle λ-coupe par :

$$C^*_\lambda(H) = \{x \in H | \lambda^+(x) \leq \lambda < \lambda^-(x)\}.$$

7.2.2 Calcul des intervalles de persistance des régions

Voyons maintenant comment calculer les intervalles Λ^*.

L'unique problème réside dans le calcul des échelles d'apparition des régions, car les connaissant, les échelles de disparition s'obtiennent en une traversée descendante de la hiérarchie en utilisant l'équation 7.16. La traversée peut s'effectuer indifféremment en largeur ou en profondeur d'abord. La solution la plus compacte (en longueur de description...) utilise une procédure récursive. Cette procédure est décrite en annexe C.

De son côté, le calcul des apparitions se fait par une généralisation de la méthode de coupe par programmation dynamique dans laquelle l'optimisation porte sur des *fonctions* de λ plutôt que sur une unique valeur d'énergie à échelle fixée. Détaillons cette procédure.

Programmation dynamique fonctionnelle

Associons à chaque région x de H la fonction qui donne l'évolution de son énergie par rapport à λ. Appelons cette fonction l'**énergie propre** de x et notons la E_x :

$$\mathsf{E}_x : \lambda \mapsto \mathsf{E}_x(\lambda) = \lambda \mathsf{C}(x) + \mathsf{D}(x).$$

E_x est affine en fonction de λ avec une pente et une ordonnée à l'origine positives.

Considérons également pour chaque hiérarchie partielle $H(x)$ l'énergie de ses λ-coupes comme une fonction de λ. Appelons cette fonction l'**énergie partielle** de x et notons la E^*_x :

$$\mathsf{E}^*_x : \lambda \mapsto \mathsf{E}^*_x(\lambda).$$

Remarquons que pour toute région de la base de H, dont la hiérarchie partielle est réduite à un seul élément, énergie propre et énergie partielle se confondent :

$$\forall b \in \mathcal{B}(H) \qquad \mathsf{E}_b^* = \mathsf{E}_b.$$

Plus généralement, la relation de programmation dynamique 7.2 devient une relation fonctionnelle :

Proposition 11 *Les fonctions d'énergie partielle E_x^* des nœuds de H sont reliées par la relation de programmation dynamique fonctionnelle*

$$\forall x \in H \qquad E_x^* = \inf\{E_x, \sum_{s \in \mathcal{F}(x)} E_s^*\}. \tag{7.18}$$

Preuve : L'équation 7.2 est valable pour toutes valeurs de λ, i.e. s'écrit

$$\forall x \in H \quad \forall \lambda \in \mathbb{R} \qquad \mathsf{E}_x^*(\lambda) = \min\{\mathsf{E}_x(\lambda), \sum_{s \in \mathcal{F}(x)} \mathsf{E}_s^*(\lambda)\}.$$

Les opérations de somme et d'infimum sur les fonctions se construisent point-à-point à partir des opérations sur les réels. Cette expression est donc équivalente à 7.18. ■

Cette équation fonctionnelle permet alors de calculer exactement l'échelle d'apparition des nœuds, et d'expliciter la fonction d'énergie des λ-coupes d'une hiérarchie. Ce calcul repose sur la forme caractéristique des fonction d'énergie partielle E_x^* quand E_λ est une énergie multi-échelles :

Proposition 12 *Soit H une hiérarchie et $E_\lambda = (C, D)$ une énergie multi-échelles. Alors, $\forall x \in H$*
i) E_x^* est affine par morceaux, non-décroissante, continue et concave.
ii) Elle s'écrit

$$E_x^*(\lambda) = \begin{cases} \sum_{s \in \mathcal{F}(x)} E_s^*(\lambda) & si \quad \lambda < \lambda^+(x) \\ E_x(\lambda) & sinon \end{cases} \tag{7.19}$$

*iii) Si C est **strictement** sous-additive alors $\lambda^+(x)$ appartient à $\mathbb{R}$ et c'est l'unique solution de*

$$E_x(\lambda) = \sum_{s \in \mathcal{F}(x)} E_s^*(\lambda).$$

Preuve :
i) La démonstration est faite par récurrence sur l'étape de programmation dynamique.

$\forall x \in H$, son énergie propre $\mathsf{E}_x = \lambda \mathsf{C}(x) + \mathsf{D}(x)$ est affine avec $\mathsf{C}(x) \geq 0$. E_x est donc affine par morceaux, non-décroissante, continue et concave, en abrégé AMC_3 (pour $AMCCC$). En particulier, comme $\forall b \in \mathcal{B}(H)$ $\mathsf{E}_b^* = \mathsf{E}_b$, les énergies partielles de la base sont AMC_3. Considérons donc $x \in H$ et supposons que $\forall s \in \mathcal{F}(x)$, E_s^* est AMC_3. Comme la somme et l'inf de fonctions AMC_3 est AMC_3, alors d'après l'équation 7.18, E_x^* est AMC_3.
ii) Posons $\mathsf{E}_{\mathcal{F}(x)}^* \triangleq \sum_{s \in \mathcal{F}(x)} \mathsf{E}_s^*$.

D'après l'équation de programmation dynamique, l'intervalle $\Lambda_\uparrow^*(x)$ de λ-optimalité partielle d'un nœud x est donné par

$$\Lambda_\uparrow^*(x) = \{\lambda | \mathsf{E}_x(\lambda) \leq \mathsf{E}_{\mathcal{F}(x)}^*(\lambda)\}.$$

Par ailleurs, la proposition 10 indique que cet intervalle vaut

$$\Lambda_\uparrow^*(x) = [\lambda^+(x), +\infty[.$$

Donc

$$\mathsf{E}_x(\lambda) \leq \mathsf{E}_{\mathcal{F}(x)}^*(\lambda) \quad \Leftrightarrow \quad \lambda \geq \lambda^+(x)$$

ce qui revient à la formulation en deux morceaux donnée par 7.19.

iii) Comme toutes les fonctions impliquées sont continues

$$\lambda^+(x) = \min\{\lambda | \mathsf{E}_x(\lambda) = \mathsf{E}^*_{\mathcal{F}(x)}(\lambda)\}.$$

Considérons donc $\Delta = \mathsf{E}_x - \mathsf{E}^*_{\mathcal{F}(x)}$ et étudions ses annulations.

$\forall \lambda \in \mathbb{R}$, $\mathsf{E}^*_{\mathcal{F}(x)}(\lambda)$ représente l'énergie d'une coupe de $H(x)$ qui correspond à une sur-partition stricte $P(\lambda)$ de $\{x\}$. Ainsi

$$\mathsf{E}^*_{\mathcal{F}(x)}(\lambda) = \lambda \sum_{R \in P(\lambda)} \mathsf{C}(R) + \sum_{R \in P(\lambda)} \mathsf{D}(R).$$

$\mathsf{E}^*_{\mathcal{F}(x)}$ est affine par morceaux et pour tous les λ pour lesquels elle est dérivable[9]

$$\frac{\partial \mathsf{E}^*_{\mathcal{F}(x)}}{\partial \lambda}(\lambda) = \sum_{R \in P(\lambda)} \mathsf{C}(R).$$

On obtient

$$\frac{\partial \Delta}{\partial \lambda}(\lambda) \quad = \quad \frac{\partial \mathsf{E}_x}{\partial \lambda}(\lambda) - \frac{\partial \mathsf{E}^*_{\mathcal{F}(x)}}{\partial \lambda}(\lambda) \quad = \quad \mathsf{C}(x) - \sum_{R \in P(\lambda)} \mathsf{C}(R).$$

Donc, si C est *strictement* sous-additive alors

$$\forall \lambda \in \mathbb{R} \qquad \frac{\partial \Delta}{\partial \lambda}(\lambda) < 0.$$

Δ est donc strictement décroissante. On vérifie par ailleurs aisément que dans ce cas $\Delta(-\infty) = +\infty$ et $\Delta(+\infty) = -\infty$. Δ s'annule donc en un point unique $\lambda \in \mathbb{R}$ qui est $\lambda^+(x)$. ■

D'après cette proposition, la fonction d'énergie E^*_x des coupes optimales d'une hiérarchie (partielle) est une fonction affine par morceaux. On peut donc *explicitement* la mémoriser pour chaque noeud de H, typiquement par la liste de ses portions affines. Ceci est vrai quelle que soit la forme de E_λ. Cependant, quand C est sous-additive, les E^*_x sont concaves (voir la figure 7.11(b)). Dans ce cas, pour chaque nœud x, le calcul de E^*_x par l'équation de programmation dynamique fonctionnelle 7.18 emploie uniquement deux opérations sur les fonctions affines par morceaux :

- Somme de deux affines par morceaux, pour calculer l'énergie des coupes de $H(x)$ strictement inférieures à $\{x\}$.
- Intersection entre une affine (E_x) et une affine par morceaux ($\mathsf{E}^*_{\mathcal{F}(x)}$), pour calculer l'échelle d'apparition de x.

Ce calcul est illustré par la figure 7.11.

Pour le calcul de l'apparition de x, on a plus précisément :

• Si C est strictement sous-additive, par la proposition 12iii), l'intersection entre $\mathsf{E}^*_{\mathcal{F}(x)}$ et E_x est unique et donne $\lambda^+(x)$.

• Si C est (au moins localement) additive trois cas se présentent. Notons C_d et D_d respectivement la pente et l'ordonnée à l'origine du *dernier* morceau de $\mathsf{E}^*_{\mathcal{F}(x)}$. Quand C est additive, $\mathsf{C}(x) = \mathsf{C}_d$. Alors

a) Si $\mathsf{D}(x) < \mathsf{D}_d$ l'intersection entre $\mathsf{E}^*_{\mathcal{F}(x)}$ et E_x est unique et donne $\lambda^+(x)$.

b) Si $\mathsf{D}(x) = \mathsf{D}_d$ x apparaît au début du dernier morceau de $\mathsf{E}^*_{\mathcal{F}(x)}$. En particulier, si x appartient au premier niveau de H alors $\lambda^+(x) = -\infty$.

c) Si $\mathsf{D}(x) > \mathsf{D}_d$ $\lambda^+(x) = +\infty$.

Connaissant $\lambda^+(x)$, la forme de E^*_x est alors donnée par 7.19.

[9] Pour faire une démonstration complète, y compris aux points de non dérivabilité des fonctions AMC_3 (ruptures de pente), il faut utiliser des dérivées à droite.

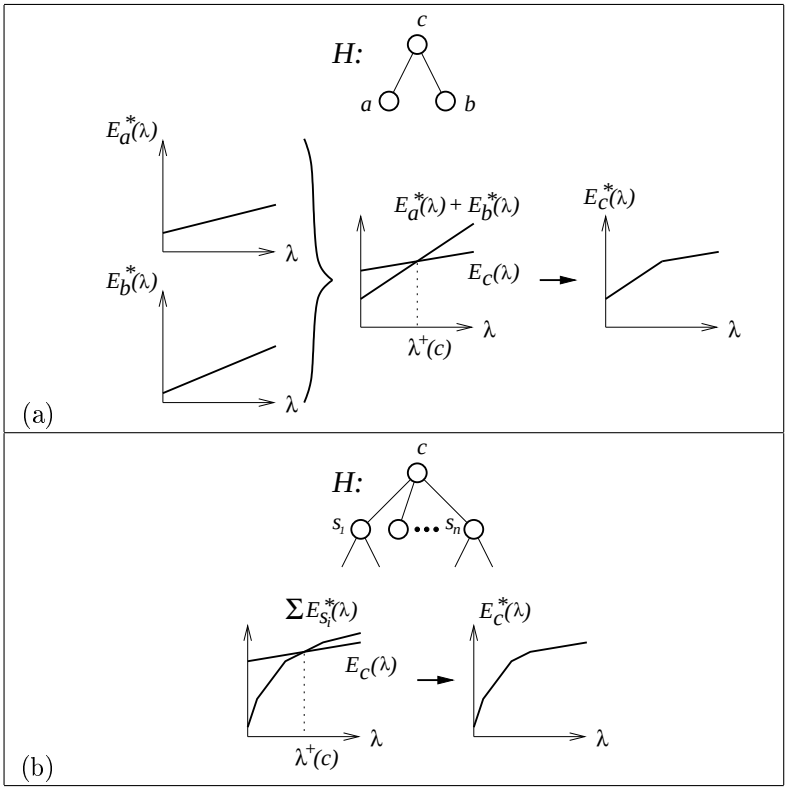


FIG. 7.11 – Calcul de l'énergie partielle E_c^* et de l'échelle d'apparition $\lambda^+(c)$ d'un nœud c. (a) Cas d'un nœud du premier niveau de H ayant deux fils. (b) Cas général.

Cette méthode permet donc de calculer λ^+ et E^* pour chaque région de H en une seule remontée ordonnée de H.

Dans un second temps, λ^- se calcule en une passe descendante dans H par l'équation 7.16.

Hiérarchie persistante

Après deux traversées de la hiérarchie initiale H, on obtient donc trois quantités pour chaque nœud : sa fonction d'énergie partielle et ses échelles d'apparition et de disparition.

Une région x peut très bien n'appartenir à aucune coupe minimale de H. x disparaît alors avant même d'apparaître, $\lambda^+(x) \leq \lambda^-(x)$ et donc $\Lambda^*(x) = \emptyset$, ce qui s'obtient quand

$$\exists y \in H \quad | \quad x \subset y \quad \text{et} \quad \lambda^+(y) < \lambda^+(x).$$

Nous disons d'une telle région qu'elle est **non persistante**. A contrario, une région dont l'intervalle de persistance n'est pas vide sera déclarée **persistante**. Les régions non persistantes n'ont pas d'intérêt du point de vue des minima de la fonction d'énergie, il s'agit d'étapes transitoires que les coupes optimales «sautent» lors de leur remontée. Il est alors naturel de les supprimer de la hiérarchie initiale.

Définition 28 *Si H est une hiérarchie et E_λ une énergie multi-échelles, alors le sous-ensemble de H*

$$H^* \triangleq \{x \in H \mid \Lambda^*(x) \neq \emptyset\}$$

*est une hiérarchie que nous appelons la **hiérarchie persistante** (pour E_λ).*

Le fait que H^* soit une hiérarchie est évident : on peut la définir comme l'union des coupes optimales de H pour E_λ. Notons toutefois que si certains nœuds de base de H ne sont pas persistants, alors la base de H^* peut être une sous-partition de celle de H.

Il est alors évident que dans H^* l'échelle de disparition d'un noeud correspond à l'échelle d'apparition de son père :

$$\forall x \in H^* \qquad \lambda^-(x) \;=\; \lambda^+(\mathsf{p}_{H^*}(x)).$$

L'échelle d'apparition est donc un critère croissant dans H^*.

La figure 7.12 illustre les étapes successives du calcul d'une hiérarchie persistante.

Nous retrouvons donc à ce point le cadre général des représentations ensembles-échelle qui a été vu en 6.2.2. Le couple (H^*, λ^+) est la représentation ensembles-échelle des coupes minimales de H, une hiérarchie indicée dont la famille de coupes naturelles est la famille des coupes minimales de H pour E_λ. En particulier, le prédicat d'homogénéité naturel associé à l'apparition est consistant. Notons le Q_λ. Il est défini par :

$$\forall x \in H^* \qquad Q_\lambda(x) \text{ est vrai } \Leftrightarrow \lambda^+(x) \leq \lambda.$$

Les coupes maximales[10] pour Q_λ représentent les coupes minimales pour E_λ ; nous sommes en quelque sorte passés d'une formulation énergétique à une formulation par prédicats. Dans la représentation naturelle de (H^*, λ^+), dans laquelle un nœud x se voit attribuer l'ordonnée $y(x) = \lambda^+(x)$, les coupes optimales correspondent à des droites horizontales. L'attribution de la hiérarchie persistante par l'indice λ^+ peut véritablement s'envisager comme le passage dans le repère «naturel» pour les coupes minimales, repère dans lequel elles sont horizontales (voir la figure 7.13).

Notons qu'après cette procédure d'optimisation hiérarchique, la fonction d'énergie partielle du sommet de la hiérarchie, $\mathsf{E}^*_{\vee H}$, représente l'énergie des coupes optimales de H en fonction de λ (voir la figure 7.14). Nous la noterons simplement E^*. Elle est affine par morceaux et pour λ donné la pente du morceau correspondant représente $\mathsf{C}(C^*_\lambda)$ et son ordonnée à l'origine donne $\mathsf{D}(C^*_\lambda)$.

[10] au sens de la relation de finesse

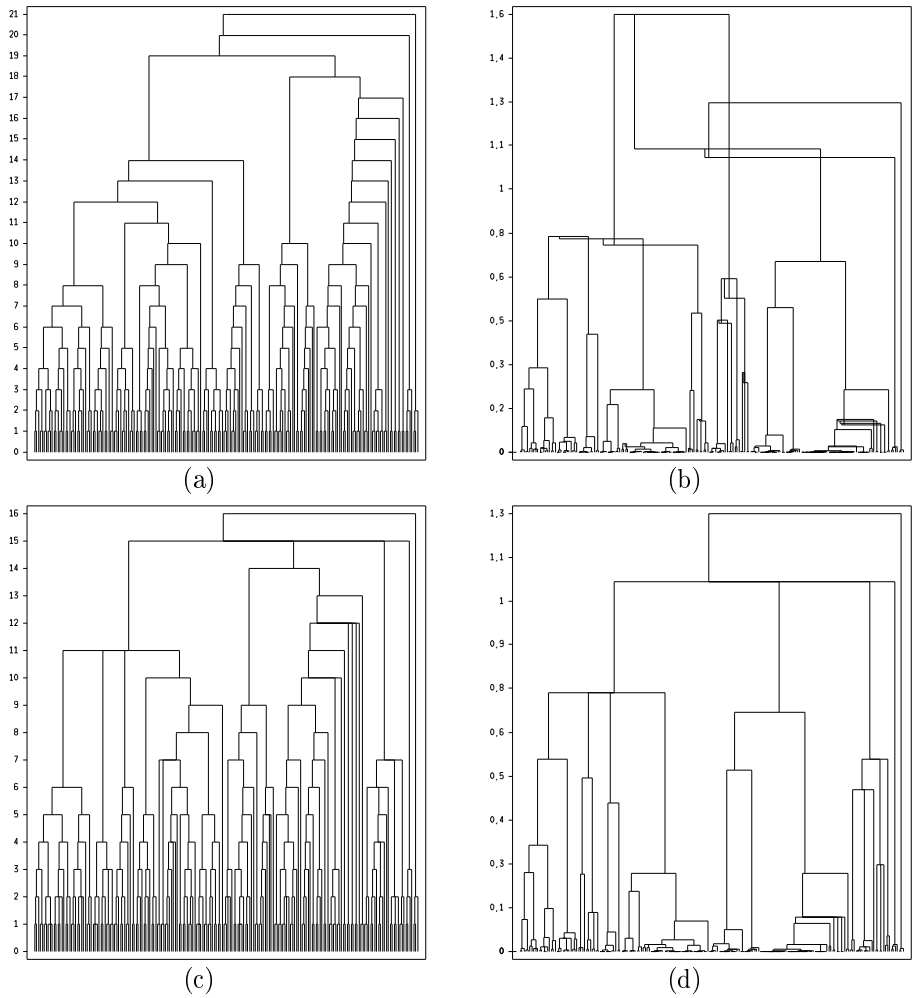


FIG. 7.12 – Les deux étapes de l'optimisation d'une énergie multi-échelles dans une hiérarchie. (a) Une hiérarchie binaire initiale H indicée par le niveau de ses noeuds. Après une traversée ascendante de H, chaque région se voit dotée de son échelle d'apparition dans une solution optimale pour l'énergie multi-échelles considérée. (b) Représente la hiérarchie H indicée par l'échelle d'apparition des noeuds. Dans cette représentation, les nœuds non persistants - qui n'appartiennent jamais à une solution optimale - se trouvent plus haut que certains de leurs ancêtres. Une traversée descendante de H permet de calculer l'échelle de disparition des régions, qui révèle les nœuds non persistants. (c)-(d) La hiérarchie persistante H^* obtenue après suppression des nœuds non persistants. Remarquons que la suppression d'étapes intermédiaires conduit à une hiérarchie n-aire. (d) Illustre le fait que l'apparition est un critère croissant pour H^*.

7.2.3 Discussion

Sous-additivité et causalité

Les résultats qui précèdent ne dépendent que d'une condition : la sous-additivité de C.

Il est remarquable que les solutions soient monotones *quel que soit le comportement de D*. Par contre, on conçoit bien que la sous-additivité de C est la condition minimale pour obtenir des séries causales : si le comportement de D est quelconque, il faut bien que le terme de régularisation guide les solutions dans la bonne direction... On pourrait donc exprimer le théorème 3 sous une forme nécessaire et suffisante : les optima d'une énergie affine $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$ sont monotones sur $\mathbb{C}ut(H)$ pour toute image I, pour toute hiérarchie H sur $\mathcal{D}(I)$ et *pour toute* énergie d'attache aux données D si et seulement si C est sous-additive.

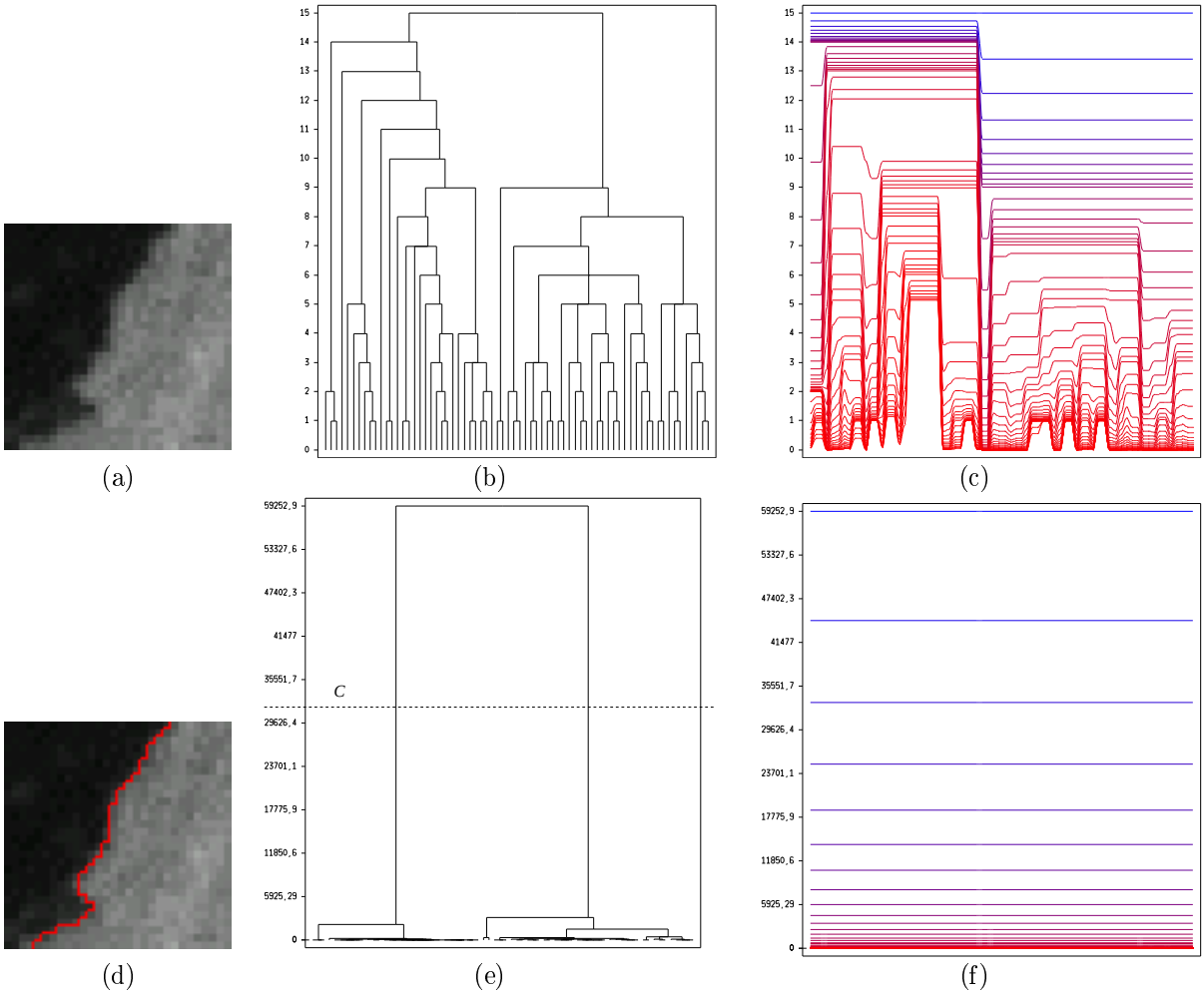


FIG. 7.13 – La représentation ensembles-échelle transforme la famille des coupes minimales d'une hiérarchie en une famille de droites horizontales. (a) Une image 30×30. (b) Hiérarchie persistante indicée par le niveau des nœuds. (c) Coupes optimales de la représentation (b). (d) Partition obtenue par coupe à échelle 1/2 de l'échelle d'apparition du sommet de la hiérarchie. (e) Hiérarchie persistante indicée par l'échelle d'apparition. C représente la coupe qui conduit à la partition (d). (f) Coupes optimales de la représentation (e). La hiérarchie est obtenue par escalade sur l'énergie constante par morceaux de Mumford-Shah à partir d'une ligne de partage des eaux (voir le chapitre suivant). Cette figure donne un premier exemple du fait que le saut d'échelle qu'il est nécessaire d'effectuer pour fusionner des régions est caractéristique de leur saillance visuelle : la représentation ensembles-échelle (e) montre clairement qu'il est nécessaire de fortement régulariser l'énergie pour réussir à grouper les deux régions de la partition (d).

Évaluons maintenant la validité de la condition de sous-additivité de C dans le cadre des énergies utilisées pour la segmentation d'image.

Régularisation non géométrique – C est évidemment sous additive pour la forme de régularisation la plus simple, qui attribue un coût identique à toutes les régions et évalue donc la complexité d'une partition à son cardinal. Ce type de régularisation a été utilisé par Fuchs (2001) pour collecter des solutions à différents niveaux de détail.

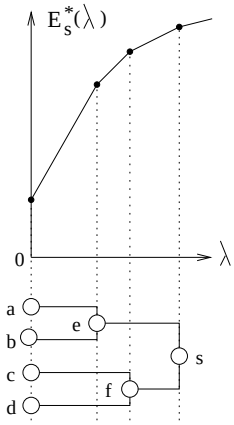


FIG. 7.14 – Fonction d'énergie des λ-coupes. La fonction d'énergie partielle E_s^* du sommet s d'une hiérarchie donne l'énergie de ses coupes optimales en fonction de λ. Elle est affine par morceaux et concave. Chaque rupture de pente traduit un événement de fusion de régions.

Énergies géométriques fortement séparables – C est systématiquement sous-additive s'il s'agit d'une énergie géométrique **fortement séparable** (voir la section 7.1.2 p. 154), typiquement si C est un terme de longueur de frontière à la Mumford-Shah/Potts/Freeman.

Dans le cas de régions connexes, ce type d'énergie géométrique décroit toujours *strictement* par union de régions voisines car l'union consiste à supprimer certaines portions de frontières, d'énergie positive[11].

Si l'on considère des régions non connexes, les énergies géométriques locales peuvent être additives. En effet, pour une union d'ensembles R_i non voisins on a

$$\delta(\bigcup R_i) = \bigcup(\delta R_i)$$

et l'énergie de frontière des régions séparées est identique à celle de leur union. Si l'attache aux données est systématiquement meilleure pour la solution divisée alors les ensembles non connexes n'appartiennent jamais à une solution optimale. C'est la raison pour laquelle l'optimum de la fonctionnelle BD de Mumford-Shah est une partition en région connexes. Dans notre cadre, si des régions non voisines et leur union appartiennent toutes à la hiérarchie alors l'union possède une échelle d'apparition de $+\infty$ (cela apparaîtra comme cas particulier de la proposition 13 ci-dessous).

Énergies géométriques faiblement séparables – Par contre, les énergies géométriques faiblement séparables - s'appuyant sur des caractéristiques **globales** des formes des régions - ne sont *a priori* pas sous-additives.

Prenons l'exemple simple d'un langage de description qui ne connait qu'un seul type de modèle : les cercles. Alors, la fusion de deux régions circulaires disjointes n'est pas un cercle et emploie une énergie de description supérieure à celle requise pour la description des deux cercles séparément. Cependant, dans l'esprit de compétition de modèles de l'inférence par codage minimal, la description par un seul cercle *ne devrait jamais être utilisée*, i.e. la meilleure description de l'union est précisément celle en tant que complexe formé de deux cercles.

Notons alors que l'on peut systématiquement *enrichir* un langage de description en lui donnant la possibilité de décrire des formes par parties. Il suffit d'ajouter en entête de la description d'une région le nombre de parties que comporte la description puis d'enchaîner les descriptions des parties (par des

[11] Notons que la sous-additivité stricte vis à vis des unions de *régions* voisines disjointes d'une énergie de *partitions* fortement séparable n'est pas contradictoire avec la condition d'additivité pour les unions de *courbes* disjointes de l'énergie de *frontières* sur laquelle elle est bâtie (equ. 7.7 p. 154).

codes auto-délimités). Le langage de description devient alors sous-additif. C'est le mécanisme qui rend sous-additives les deux notions principales de complexité «universelle» : la *complexité de Kolmogorov* et la *profondeur logique de Bennet* (Delahaye 1999, p. 101). Si S et T sont deux séquences binaires, un programme trivial produisant ST est la concaténation de deux programmes, un programme produisant S suivi d'un programme pour T. La longueur (complexité de Kolmogorov) comme le temps d'exécution (profondeur logique de Bennet) du programme minimal produisant ST est donc inférieure à la somme des longueurs ou temps d'exécution nécessaires à produire les deux séquences séparément. La sous-additivité est même jugée indispensable pour tout candidat à ce type de notion. Comme l'exprime Delahaye (1999, p.109) : «Deux moutons ne sont pas deux fois plus complexes qu'un mouton».

Notons que si l'on prend en compte des régions non connexes, alors contrairement aux critères locaux, les critères globaux peuvent produire des compressions de description pour les unions d'ensembles non voisins. Quand les différents ensembles possèdent des formes ou des modèles image proches, i.e. réalisent des Gestalts, une «factorisation» des descriptions individuelles peut être réalisée (emploi d'un dictionnaire, codage de Huffman...). Par exemple, si nos deux cercles qui précèdent ont le même rayon, celui-ci peut n'être stipulé qu'une seule fois. Et même plus, il peut être plus économique de modéliser deux cercles de rayons proches comme deux cercles de même rayon et décrire l'erreur commise. Les descriptions *orientées région* peuvent donc avoir un véritable comportement sous-additif, y compris pour les unions non connexes.

Techniquement, la condition de sous-additivité n'est pas nécessaire pour réaliser l'optimisation exhaustive d'une énergie affine en fonction de λ par la méthode de programmation dynamique fonctionnelle. Cependant, quand C n'est pas sous-additive, les ensembles de persistance $\Lambda^*(x)$ ne sont pas nécessairement des intervalles. En effet, dans ce cas, l'ensemble $\Lambda^*_\uparrow(x)$ des échelles d'optimalité partielle de x peut être un intervalle *borné*, de la forme $[a, b]$ (alors que pour les énergies sous-additives, il est toujours de la forme $[a, +\infty[$). Dans le cas non sous-additif, l'ensemble des échelles de persistance, qui est défini en général par l'équation 7.17 peut alors consister en une union d'intervalles disjoints. Les coupes optimales ne sont alors pas nécessairement monotones, une région peut apparaitre dans une solution à λ_1, disparaitre pour $\lambda_2 > \lambda_1$ pour réapparaitre pour $\lambda_3 > \lambda_2 \ldots$

L'échelle des échelles

Comme on a pu le remarquer, rien dans les développements techniques qui précèdent ne contraint le paramètre d'échelle à être *positif*.

Comparaison avec le vide – En particulier, toutes les régions de base apparaissent a priori en $-\infty$. De son côté, le sommet disparaît en $+\infty$. Ces deux cas correspondent à des effets de bord que nous avons déja rencontrés avec les cocons. En 5.7.3 nous avions appelé ce phénomène le problème de «comparaison avec le vide». L'origine du problème est exactement la même ici : le critère d'optimalité *globale* d'une région pour E_λ est également un critère relatif entre intérieur et extérieur de la région. L'échelle d'apparition dépend de l'énergie des sous-parties d'une région (passe ascendante) et l'échelle de disparition de l'énergie de sur-parties (passe descendante). Ces sous ou sur-parties n'existent pas aux limites du domaine perceptuel, au niveau des «atomes» ou de l'«Univers». Les conséquences sont les mêmes que pour les cocons : on ne peut pas qualifier la pertinence des parties élémentaires ou de l'image entière. Dans le cas des cocons, les nœuds isolés et le graphe entier étaient systématiquement des cocons, de contraste infini. Dans le cas présent, les feuilles et le sommet possèdent systématiquement des intervalles de persistance infinis.

Échelles d'apparition négatives – En dehors des extrêmes, si D est elle-même *sous-additive* (éventuellement seulement ponctuellement, i.e. si elle est non sur-additive) alors des régions qui n'appartiennent pas à la base peuvent également apparaître dans les échelles négatives.

Pour montrer cela, voyons une expression importante de l'échelle d'apparition :

Proposition 13 *L'échelle d'apparition d'une région x est donnée par*

$$\lambda^+(x) \quad = \quad -\frac{D(\{x\}) - D(P)}{C(\{x\}) - C(P)} \tag{7.20}$$

où P est la partition de $\mathbb{C}ut(H(x))\backslash\{x\}$ qui est optimale en $\lambda^+(x)$.

Preuve : $\lambda^+(x)$ représente l'échelle pour laquelle l'énergie de la coupe $\{x\}$ de $H(x)$ devient égale à l'énergie d'une coupe optimale P strictement inférieure à $\{x\}$. Donc

$$\begin{aligned} \lambda^+(x) &= \lambda \quad \text{tel que} \quad \lambda\mathsf{C}(\{x\}) + \mathsf{D}(\{x\}) = \lambda\mathsf{C}(P) + \mathsf{D}(P) \\ &= -\frac{\mathsf{D}(\{x\}) - \mathsf{D}(P)}{\mathsf{C}(\{x\}) - \mathsf{C}(P)}. \end{aligned}$$

■

Cette propriété relativement évidente a de nombreuses conséquences.

On voit tout d'abord que par la sous-additivité de C, le dénominateur de l'expression 7.20 est toujours négatif. Le signe de $\lambda^+(x)$ est donc celui de la variation d'attache aux données, $\mathsf{D}(\{x\}) - \mathsf{D}(P)$, quand P est fusionnée en $\{x\}$. L'échelle d'apparition d'une région peut donc être négative, si D est localement sous-additive.

L'idée d'échelles négatives est relativement déplaisante : qu'est-ce qu'une énergie de régularisation négative signifie ? Si aucune régularisation du tout n'est nécessaire pour qu'une région soit préférablement décrite comme un tout plutôt que par parties, cela signifie que le terme d'«attache aux données» D contient lui-même un facteur de «régularisation». Le terme $\lambda\mathsf{C}$ négatif que l'on trouve s'interprète alors comme une quantité d'énergie d'«anti-régularisation» nécessaire pour compenser la régularisation interne à D et parvenir à ... diviser la région !

À la limite, si D est parfaitement sous-additive, alors le sommet de H apparait à une échelle négative : la coupe optimale «saute» immédiatement en haut de la hiérarchie, l'image est complètement fusionnée pour toute échelle positive. Notons que d'un point de vue du codage minimal, des λ négatifs n'ont pas de sens : ils correspondent à des codes de taille négative.

Énergies multi-échelles pures – Inversement, si D est *sur-additive* alors d'après l'expression 7.20 toutes les échelles d'apparition sont positives. En conséquence,

Définition 29 *Une énergie affine $E_\lambda = (C, D)$ constituée d'un terme C sous-additif et d'un terme D sur-additif est appelée une* ***énergie multi-échelles pure****.*

Une telle énergie correspond à un modèle dans lequel les deux termes C et D sont véritablement en opposition du point de vue de l'ordre sur les parties, l'un croissant et l'autre décroissant.

Notons toutefois que l'on peut toujours ne s'intéresser qu'aux solutions pour les échelles positives, même si l'énergie n'est pas purement multi-échelles. *c'est ce que nous ferons en pratique* car la manipulation des échelles négatives est malaisée. Les régions de base pourront alors de jamais être optimales, quand une sur-partie possède une apparition négative, et être remplacées à la base de la hiérarchie persistante par leur sur-partie qui est optimale à $\lambda = 0$.

Version complémentaire pour les énergies sur-additives

Les résultats qui précèdent ont une version complémentaire dans le cas où l'énergie C est **sur-additive** plutôt que sous-additive. Bien sûr, cette situation ne correspond a priori pas aux problèmes de segmentation. Elle correspond cependant à certains problèmes de compression d'image que nous aborderons dans un instant.

On constate aisément que dans le cas où C est sur-additive, le théorème de monotonie est inversé : les coupes décroissent dans H avec λ. Le rôle des échelles d'apparition et de disparition est également inversé. L'intersection entre énergie propre et somme des énergies partielles des fils d'un nœud représente λ^-. λ^+ est alors donnée par :

$$\lambda^+(x) = \max_{y \in H, y > x} \{\lambda^-(y)\}.$$

Il suffit finalement que l'une des deux énergies C ou D soit monotone sur $\mathbb{P}(X)$ (sur ou sous additive) pour obtenir une série causale de solutions. On doit juste placer le facteur d'échelle devant le terme adéquat.

Extension multidimensionnelle

La méthode s'étend naturellement à des régularisations «multi-dimensionnelles».

Considérons un «paramètre d'échelle» vectoriel $\lambda \in \mathbb{R}^n$ et un vecteur d'énergies de régularisation

$$\mathsf{C} : \mathcal{P}(X) \to \mathbb{R}^n.$$

Une énergie multi-échelles multi-dimensionnelle est alors définie sur chaque région par

$$\mathsf{E}_\lambda(R) =< \lambda, \mathsf{C}(R) > + \mathsf{D}(R)$$

où $< \cdot, \cdot >$ désigne le produit scalaire de $\mathbb{R}^n$. L'énergie d'une partition est toujours définie comme la somme des énergies de ses parties.

Les énergies propres des régions en fonction de λ sont alors des hyperplans affines de $\mathbb{R}^{n+1}$ ($\mathbb{R}^n \times$ l'axe d'énergie) et les énergies partielles sont des hyperplans affines par morceaux. La figure 7.15 illustre le cas en 2D.

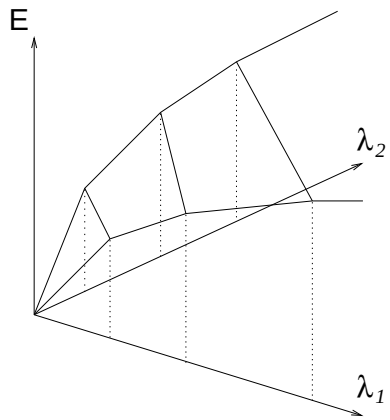


FIG. 7.15 – Extension multidimensionnelle de l'optimisation hiérarchique. Illustration en 2D.

Si toutes les composantes de C sont sous-additives alors tous les résultats qui précédent sont toujours valides.

Considérons en effet une droite D de $\mathbb{R}^n$ de vecteur directeur v ayant toutes ses coordonnées positives. Un point d'abcisse x le long de D se trouve alors en $x.v$ et la valeur de l'énergie d'une région en ce point s'écrit

$$\mathsf{E}_x^v(R) = x < v, \mathsf{C}(R) > + \mathsf{D}(R).$$

La sous-additivité de chaque composante de C assure alors que l'énergie projetée $< v, \mathsf{C} >$ est également sous-additive (car toutes les coordonnées de v pointent dans la direction positive). On retrouve

donc le cas à une dimension qui vient d'être étudié, pour toute direction de l'espace (ayant toutes ses coordonnées positives, i.e. pointant dans le cadran principal de l'espace).

Les fusions de régions interviennent alors le long de droites de $\mathbb{R}^{n+1}$ qui sont totalement déterminées par leurs intersections avec les plans verticaux s'appuyant sur chacun des axes dirigés par une composante de λ. Autrement dit, l'apparition d'une région est déterminée par le vecteur de ses apparitions pour l'énergie projetée sur chaque direction principale de l'espace (chaque terme de régularisation).

L'optimisation dans $\mathbb{R}^n$ revient donc à conduire simultanément n optimisations mono-dimensionnelles. Les nœuds de la hiérarchie portent n fonctions affines par morceaux et des vecteurs n-dimensionnels d'apparition.

7.2.4 Compression avec perte et minimisation sous contrainte

Caricature visuelle optimale

Salembier et Garrido (2000) décrivent une application possible des hiérarchies de régions : la création d'aperçus d'images, ou d'imagettes (thumbnails) pour permettre l'examen visuel rapide d'images (visual browsing), par exemple pour sélectionner les résultats d'une requête dans une base de données image. Salembier et Garrido envisagent le problème de construire une image simplifiée d'une image d'origine et considèrent pour cette construction deux problèmes duaux d'optimisation sous contrainte qui correspondent aux problèmes 6.18 et 6.19 vus page 133. Nous montrons ici que la méthode d'optimisation multi-échelles proposée permet de résoudre exactement et simultanément tous les problèmes de caricature envisagés par ces auteurs.

Salembier et Garrido envisagent des caricatures par plages constantes d'une image. Le codage des régions de la caricature est indépendant, chaque région étant déterminée par sa moyenne (codée sur 24 bits en couleur) et le code de Freeman de sa frontière, ce qui conduit pour une région R à un coût :

$$\mathsf{C}(R) = 24 + \frac{k}{2}|\delta R|.$$

où k est le coût de codage d'un élément de frontière (entre $\log 3$ et 2 bits en 4-connexité ou pour des frontières inter-pixellaires).

La distorsion D de l'image d'origine I par rapport à ce modèle est alors mesurée en norme L_2, qui est la façon la plus courante de mesurer la distorsion visuelle dans les problèmes de compression (Guillois 1996) :

$$\mathsf{D}(R) = \sum_{x \in R} ||I(x) - \overline{I}_R||^2.$$

où $\overline{I}_R$ représente la moyenne de l'image sur la région R.

Si l'on dispose d'une hiérarchie de régions H sur l'image, le problème de *meilleure caricature de volume borné* est posé comme celui de trouver une coupe P de H solution de

$$\min \quad \mathsf{D}(P) \qquad \text{sous la contrainte} \quad \mathsf{C}(P) \leq \gamma \tag{7.21}$$

(Salembier et Garrido envisagent également le problème primal de compression optimale, qui échange le rôle de D et C). Les auteurs passent alors au lagrangien $\mathsf{E}_\lambda = \lambda\mathsf{C} + \mathsf{D}$ et emploient l'algorithme de programmation dynamique simple pour trouver une coupe optimale *pour un λ donné*. Ils couplent alors cet algorithme à une descente de gradient sur λ pour trouver une coupe P correspondant à un λ tel que $\mathsf{C}(P) \approx \gamma$. Remarquons qu'*a priori* le problème peut être non convexe en λ et une descente de gradient ne fournir qu'un optimum local. En outre, cette itération de coupes minimales ne fournit qu'une seule solution, pour une valeur donnée de γ.

Résolution complète du problème de débit/distorsion

Le problème lagrangien correspond à un problème d'optimisation d'énergie multi-échelles, tel que nous venons de l'étudier.

Il faut d'abord s'assurer que la fonction C est bien sous-additive, ce qui est évident : le problème n'a pas de sens si la fusion de régions ne produit pas de compression (c'est effectivement le cas pour la fonction C de Salembier et Garrido qui combine une énergie fortement séparable et une énergie constante par région qui sont deux énergies sous-additives). L'algorithme d'optimisation multi-échelles est donc applicable, et après une remontée hiérarchique, le sommet de la hiérarchie porte une fonction affine par morceaux $\lambda \mapsto \mathsf{E}^*(\lambda)$ qui est la forme explicite du lagrangien du problème.

Comme nous l'avons vu, le coût de codage des solutions est alors donné par la pente de ce Lagrangien :

$$\mathsf{C}^* = \frac{\partial \mathsf{E}^*}{\partial \lambda}. \tag{7.22}$$

De son côté, la distorsion de la solution à λ_0, $\mathsf{D}^*(\lambda_0)$, est donnée par l'ordonnée à l'origine du morceau affine qui correspond à λ_0 :

$$\mathsf{D}^*(\lambda_0) \quad = \quad \mathsf{E}^*(\lambda_0) - \lambda_0 \mathsf{C}^*(\lambda_0) \quad = \quad \mathsf{E}^*(\lambda_0) - \lambda_0 \frac{\partial \mathsf{E}^*}{\partial \lambda}(\lambda_0). \tag{7.23}$$

D'après la proposition 12, E^* est non décroissante et concave (voir figure 7.16(a)). C^* et D^* sont donc deux fonctions constantes par morceaux, la première décroissante et la seconde croissante (cf. figure 7.16(b) et (c)). D^* est donc inversible et la fonction $\mathsf{C}^*(\mathsf{D}^*)$ est alors la courbe de débit/distorsion du problème, qui est bien décroissante (cf. figure 7.16(d)).

Pour chaque contrainte γ, la formulation approximative de l'échelle de coupe optimale donnée ci-dessus ($\mathsf{C} \approx \gamma$) pour le problème de caricature optimale peut alors être précisée : l'échelle (ou paramètre de Lagrange) optimale λ^* est

$$\lambda^*(\gamma) \quad = \quad \min\{\lambda \mid \mathsf{C}^*(\lambda) \leq \gamma\}. \tag{7.24}$$

Cette échelle s'obtient instantanément pour toute valeur de γ, ainsi que la coupe naturelle qui lui correspond dans la représentation ensembles-échelle.

La méthode d'optimisation multi-échelles permet donc de déterminer en deux passes hiérarchiques toutes les solutions (pour toutes les valeurs de contrainte γ) du problème de caricature optimale proposé par Salembier et Garrido. En outre, elle montre que si C est strictement sous-additive, alors

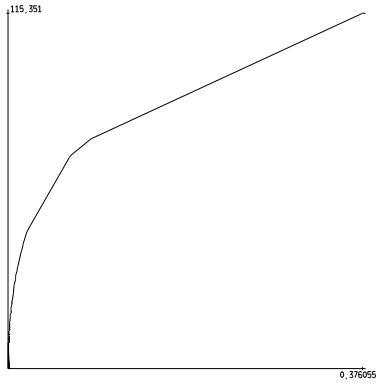
- Une solution existe si et seulement si $\gamma > \mathsf{C}(\vee H)$.
- Si une solution existe, elle est unique.
- Les partitions optimales sont monotones en fonctions de γ.

Le problème de compression optimale à distorsion bornée échange le rôle de C et D. Dans les problèmes de compression l'énergie est une énergie multi-échelles «pure» et comme D est sur-additive, il faut utiliser la version complémentaire de l'algorithme d'optimisation.

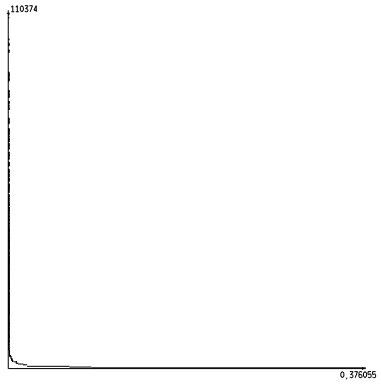
7.3 Conclusion

Ce chapitre s'est intéressé à une importante famille d'énergies employées en partitionnement d'image, les énergies *séparables* sur les régions d'une partition qui s'écrivent comme une combinaison *linéaire* de deux termes antagonistes, un terme d'attache aux données D et un terme de régularisation C.

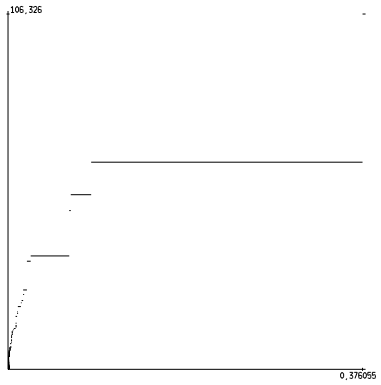
Nous avons vu que la minimisation de telles énergies, bien qu'elle soit intrinsèquement combinatoire lorsqu'elle est envisagée sur le treillis complet des partitions, peut être résolue exactement lorsqu'on la considère sur l'espace des partitions qui correspondent à des coupes dans une hiérarchie de régions. Nous avons en outre montré la richesse de ces espaces de partitions et la force du principe de programmation dynamique, qui conduit à un algorithme de complexité linéaire en fonction du nombre de régions de la hiérarchie et donc du nombre de pixels d'une image.



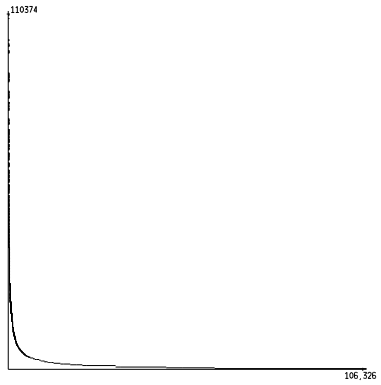
(a) $\mathsf{E}^*(\lambda)$ – Lagrangien



(b) $\mathsf{C}^*(\lambda)$ – Débit en fonction du paramètre de Lagrange



(c) $\mathsf{D}^*(\lambda)$ – Distorsion en fonction du paramètre de Lagrange



(d) $\mathsf{C}^*(\mathsf{D}^*)$ – Débit en fonction de la distorsion

FIG. 7.16 – Résolution exacte du problème de débit/distorsion dans une hiérarchie. Cet exemple correspond aux courbes d'énergie associées à une hiérarchie d'escalade bâtie sur la fonction d'énergie de Salembier et Garrido à partir des 4096 pixels d'une image 64×64 (voir le chapitre suivant). Les fonctions comportent seulement 3340 morceaux car 756 étapes de la fusion ne sont pas persistantes. L'origine de chaque courbe correspond à la sur-segmentation absolue de l'image et le point d'abcisse maximale à sa sous-segmentation absolue.

Nous avons ensuite étendu ce résultat en nous intéressant au comportement des coupes minimales en fonction du paramètre d'échelle λ, qui règle le poids relatif des deux termes énergétiques. Nous avons montré que si le terme de régularisation est sous ou sur-additif, c'est-à-dire si l'énergie de régularisation d'une partition est monotone dans le treillis des partitions, alors les coupes minimales pour E_λ sont également monotones.

Ce résultat représente le pendant pour les formulations énergétiques affines de la propriété de croissance des coupes naturelles associées à un critère croissant, qui intervient dans le cadre des formulations par prédicats d'homogénéité.

Nous avons décrit comment complètement inverser le problème par rapport à λ, ce qui conduit à une sous-hiérarchie de la hiérarchie d'origine - la hiérarchie persistante - indicée par un critère croissant - l'échelle d'apparition - dont les coupes naturelles sont les coupes minimales pour E_λ.

Revenons sur la philosophie de l'approche, qui consiste en un renversement de point de vue :

On considère habituellement λ comme un paramètre de l'algorithme de partitionnement A et le problème est posé comme celui de trouver la partition $A(\lambda)$ qui minimise E_λ. Nous adoptons le point de vue inverse en voyant λ comme une variable et en considérant la question : pour quelles valeurs de λ une région donnée entre-elle dans une solution optimale ? Cela revient à s'intéresser à A^{-1} plutôt qu'à

A, c'est-à-dire à s'intéresser aux échelles «propres» des objets plutôt qu'aux objets «vus» à une échelle d'analyse donnée. Bien entendu, une telle approche n'est praticable que parce que le nombre de régions d'une hiérarchie est raisonnable.

Le déroulement de l'optimisation elle-même consiste à manipuler de manière formelle les fonctions d'énergie en fonction de λ. Dans notre cas, il s'agit de simples fonctions affines par morceaux. On voit cependant que ce type de méthode est généralisable à tout autre type de calcul fonctionnel que l'on peut conduire de manière formelle. Le principe d'expliciter les fonctions d'énergie par rapport à leurs paramètres et de considérer le problème inverse d'exhiber les valeurs de paramètre pour lesquelles un élément devient solution mérite certainement d'être approfondi pour d'autres problèmes d'analyse de données. Cette approche permet en effet de s'interroger sur la stabilité d'une solution (dans notre cas la persistance des régions), une question certainement aussi importante que le fait d'en obtenir une. L'exploitation des échelles de vie des régions sera développée aux chapitres suivants.

Comme nous l'avons vu au chapitre précédent, les minima d'une énergie affine correspondent également aux solutions d'un problème d'optimisation sous contrainte. Cette équivalence Lagrangienne classique fournit finalement une interprétation relativement satisfaisante des solutions trouvées. En effet, intuitivement, une diminution d'échelle de description (augmentation de λ) doit être associée à une simplification de description. Le modèle qui vient d'être développé se comporte selon cette intuition : augmenter λ correspond à contraindre les modèles d'image recherchés à être de plus en plus simples (borne sur la complexité de modèle C). Pour chaque contrainte de complexité, le modèle exhibé est le plus proche de l'image d'origine, au sens d'une mesure D librement choisie. D'un point de vue compression d'images orientée régions, la méthode donne les solutions exactes d'une famille de problèmes de débit/distorsion opérationnels envisagés dans un sous-espace de parties hiérarchisées.

Chapitre 8

Escalade

Le chapitre qui précède a montré comment - pour une hiérarchie H *donnée* - on pouvait calculer et représenter exactement l'ensemble de ses coupes optimales pour une énergie multi-échelles E_λ.

La méthode d'optimisation est relativement générale, elle peut être utilisée en coopération avec n'importe quelle méthode de fusion ou division récursive d'une image. Notons en particulier que si la hiérarchie est un quad-tree, cette méthode constitue le pendant énergétique de la méthode de Split and Merge d'Horowitz et Pavlidis, qui s'appuie sur une formulation par prédicats d'homogénéité consistants. Il est cependant naturel de chercher à obtenir une hiérarchie irrégulière - dont les formes s'adaptent au contenu de l'image - et qui soit associée au critère d'optimalité qui est contenu dans l'énergie considérée.

Ce chapitre propose une façon naturelle de coupler la construction de la hiérarchie avec la technique d'optimisation multi-échelles, afin d'obtenir une méthodologie de segmentation multi-échelles complète s'appuyant sur un principe d'optimisation d'énergie.

Il aboutit à une formalisation correcte - en terme de recherche d'une *hiérarchie* optimale - du problème de segmentation multi-échelles, qui n'avait été formulé qu'imprécisément au chapitre 6 (nous y avons parlé de déterminer un algorithme multi-échelles qui fournisse des partitions «aussi proches que possible» des optima d'une énergie affine, dans un sens non défini).

8.1 Le principe d'escalade

Nous avons vu que différents auteurs (Koepfler et al. 1994; Ballester et al. 1994; Fuchs 2001) proposent de construire des pyramides de segmentations en itérant des fusions par descente d'énergie selon une suite croissante de paramètres d'échelle. Les résultats du chapitre précédent suggèrent une autre stratégie, qui relève d'un principe véritablement multi-échelles. Présentons cette stratégie sous un angle purement hiérarchique.

8.1.1 Complétions croissantes d'une pré-hiérarchie

Remarquons avant tout que la méthode de programmation dynamique pour l'optimisation multi-échelles est intrinsèquement ascendante : la résolution du problème d'optimisation globale dans une hiérarchie se fait par composition de résolutions partielles, pour des sous-parties d'image. Du point de vue de l'optimisation d'une énergie séparable, l'ordre dans une hiérarchie possède un *sens* privilégié, qui correspond à l'orientation de l'axe des échelles donnée par le principe de causalité. Ce cadre suggère donc une profonde dissymétrie entre mécanismes de groupement et de division. Il est alors naturel de s'en remettre à un processus ascendant pour la construction de la hiérarchie.

Supposons donc que l'on ait déja partiellement construit - par le bas - une hiérarchie H sur un ensemble X (e.g. les pixels d'une image). Cette construction ne sera pas à proprement parler une hiérarchie mais un

ensemble de hiérarchies disjointes, une forêt en terme de graphes (cf. figure (8.1)). En effet, il manque à H son sommet (ou sa racine) unique X pour être strictement une hiérarchie. Appelons une telle structure une **pré-hiérarchie**. Formellement

Définition 30 *Une **pré-hiérarchie** sur X est un ensemble de parties hiérarchisées de X qui contient $\{x\}$ pour tout $x \in X$, mais pas nécessairement X lui-même.*

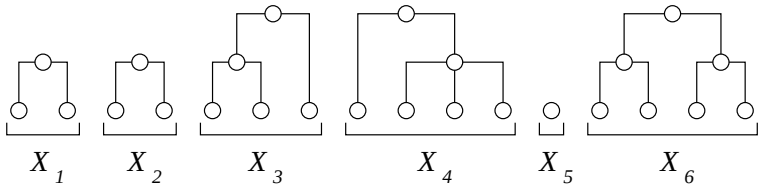


FIG. 8.1 – Une pré-hiérarchie.

Soit H une pré-hiérarchie sur X. H se décompose de manière unique en une union de hiérarchies disjointes, qui correspondent aux composantes connexes de la forêt qui représente H. Nous appelons les éléments de cette décomposition les **composantes hiérarchiques** de H et nous notons $\mathsf{CC}(H)$ leur ensemble. Les composantes hiérarchiques partitionnent H et leurs sommets partitionnent X.

Si H est une pré-hiérarchie sur X, nous disons qu'un ensemble $x \in \mathcal{P}(X)$ est une **complétion** possible de H si $x \notin H$ et $H \cup \{x\}$ est également une pré-hiérarchie (nous dirons aussi de $H \cup \{x\}$ est la complétion de H par x). Si une pré-hiérarchie n'est pas une hiérarchie (ne contient pas X), une manière de la compléter est alors de lui ajouter un ensemble x formé par réunion de certains sommets de ses composantes hiérarchiques, i.e. qui s'écrit

$$x = \bigcup_{C \in K} \vee C$$

où $K \subset \mathsf{CC}(H)$ est un sous-ensemble des composantes hiérarchiques de H. Nous disons qu'une telle partie est une **complétion croissante** de H et nous notons H^+ l'ensemble des complétions croissantes de H. On peut également définir une complétion croissante de H comme une complétion telle que $\forall y \in H \;\; x \not\subset y$. A contrario, nous dirons que les complétions non croissantes sont des **complétions intérieures**, elles correspondent en effet à des ensembles inclus dans un sommet d'une composante hiérarchique.

Ajouter une partie de H^+ à H revient à «réunir» un certain nombre de composantes hiérarchiques en une seule composante, par la construction d'un nouveau sommet. On peut voir cette opération comme une opération d'«union interne» de hiérarchies disjointes, définie par

$$\underbrace{\biguplus H_i}_{\text{(opération sur les hiérarchies)}} \triangleq \underbrace{\left(\bigcup H_i\right) \cup \left\{\bigcup \vee H_i\right\}}_{\text{(opérations sur les ensembles)}}.$$

Toute pré-hiérarchie résultant d'une complétion croissante de H s'obtient donc sous la forme

$$\left(\biguplus_{C \in K} C\right) \cup \left(\bigcup_{C \in \mathsf{CC}(H) \setminus K} C\right).$$

où $K \subset \mathsf{CC}(H)$ désigne l'ensemble des composantes hiérarchiques de H qui sont réunies. Notons qu'une complétion croissante diminue systématiquement le nombre de composantes de H alors qu'une complétion intérieure le laisse inchangé.

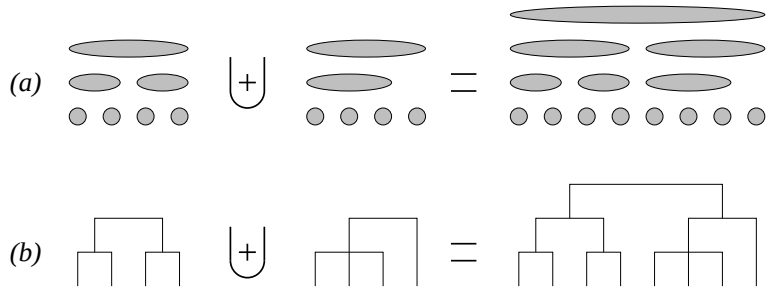


FIG. 8.2 – Union interne de deux hiérarchies disjointes. (a) vue sous l'angle ensembliste. (b) vue sous l'angle des graphes.

8.1.2 Escalade pure

Supposons alors que l'on dispose d'une énergie multi-échelles E_λ. Cette énergie attribue une échelle d'apparition aux élément d'une hiérarchie. En particulier, quand on réalise l'union interne d'un certain nombre de composantes hiérarchiques, le sommet de l'union acquiert une échelle d'apparition pour E_λ.

Définissons donc la stratégie «pure» d'escalade (cf. figure 8.3) :

Définition 31 ***Escalade pure***
Si X est un ensemble et E_λ une énergie multi-échelles,
• *Pour toute pré-hiérarchie H sur X, un opération d'**escalade pure** sur H pour E_λ consiste à réaliser la complétion croissante de H qui possède la plus faible échelle d'apparition pour E_λ, i.e. à ajouter à H*

$$\operatorname*{argmin}_{x \in H^+} \lambda^+(x).$$

• *La **hiérarchie d'escalade pure** sur X pour E_λ est la hiérarchie obtenue par itération d'opérations d'escalade pure à partir de la pré-hiérarchie minimale $H = \{\{x\}\}_{x \in X}$ (la sur-partition absolue de X).*

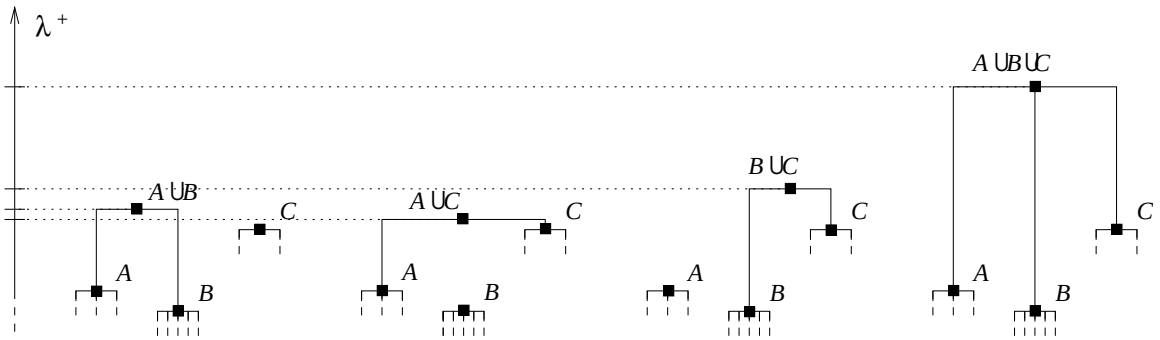


FIG. 8.3 – Stratégie d'escalade pure. Pour une pré-hiérarchie H qui comporte trois composantes hiérarchiques de sommets A, B et C, la stratégie d'escalade pure consiste à évaluer toutes ses complétions croissantes possibles - $A \cup B$, $A \cup C$, $B \cup C$ ou $A \cup B \cup C$ - et à retenir celle qui possède la plus faible échelle d'apparition λ^+ pour l'énergie considérée. Dans l'exemple ci-dessus, on retiendrait $A \cup C$.

Comme nous l'avons vu, si E_λ est une énergie multi-échelles alors les coupes optimales d'une hiérarchie remontent quand λ croit. Elles «sautent» à des éléments supérieurs pour certaines valeurs caractéristiques du paramètre d'échelle. Disposant d'une hiérarchie incomplète, le critère d'escalade pure choisit alors de la compléter par l'ensemble qui serait le premier atteint par la remontée des coupes optimales au cours d'une augmentation progressive du paramètre d'échelle λ.

Notons que toutes les étapes d'une escalade pure sont persistantes. En effet, si après une complétion par un ensemble $Y = \bigcup_{k \in K} X_k$, où les X_k sont des sommets de composantes, un des ensembles réunis X_p devenait non persistant, i.e. tel que $\lambda^+(X_p) > \lambda^+(Y)$, alors l'escalade n'aurait pas choisi d'ajouter X_p à une étape précédente, mais aurait directement complété la hiérarchie par Y. L'escalade pure «shunte» naturellement les étapes non pertinentes pour l'énergie, qui n'apparaîtraient jamais dans une coupe optimale. Autrement dit, au cours d'une escalade pure, il n'est pas nécessaire de considérer les complétions intérieures : elles ne seraient jamais sélectionnées.

Le critère d'escalade pure considère à chaque étape toutes les complétions possibles, y compris par X lui-même. Si l'image contient effectivement des formes, les ensembles intermédiaires qui délimitent ces formes apparaissent *avant* la fusion de l'image en un seul ensemble, qui nécessite une plus forte régularisation pour devenir solution. L'escalade pure sélectionne donc naturellement les étapes intermédiaires de fusion qui interviennent avant l'apparition de l'image entière, circonscrite en un ensemble unique.

8.1.3 Escalade binaire et escalade binaire connexe

Évidemment, l'escalade pure est impraticable : si à une étape donnée, H est constituée de N composantes hiérarchiques, ce critère considère 2^N complétions possibles. En pratique, l'escalade s'effectue donc comme tous les algorithmes de fusion de régions, en restreignant l'ensemble des complétions possibles aux réunions de paires de régions voisines.

Dans le formalisme qui vient d'être présenté, on définit l'ensemble $H^{2+} \subset H^+$ des unions $X_i \cup X_j$ de seulement deux sommets de composantes hiérarchiques de H. Et si X est muni d'une topologie $\mathcal{T}$, on définit l'ensemble $H_c^{2+} \subset H^{2+}$ des unions de deux sommets voisins pour $\mathcal{T}$. Si les sommets de H sont tous connexes, H_c^{2+} ne comporte que des régions connexes pour $\mathcal{T}$.

On obtient alors des restrictions praticables de l'escalade pure en remplaçant H^+ par H^{2+} ou H_c^{2+} dans la définition 31 qui précède.

Définition 32 *L'escalade selon H^{2+} est appelée* ***escalade binaire*** *et celle selon H_c^{2+}* ***escalade binaire connexe****.*

En pratique, nous n'avons considéré que des hiérarchies obtenues par escalade binaire connexe. Nous verrons cependant que l'idée importante est celle d'escalade pure et que sa restriction à des unions de régions voisines n'en est qu'une approximation praticable. On pourrait bien sûr imaginer des versions de complexité calculatoire intermédiaire, considérant des unions de n-uplets de régions voisines.

Notons que contrairement à l'escalade pure, les escalades binaires peuvent produire des étapes intermédiaires non persistantes (cf. figure 8.4). L'interprétation de l'escalade comme un processus d'augmentation continue de λ reste cependant valide : après élimination des régions non persistantes, tout se passe comme si l'on avait directement considéré des fusions n-aires et continuement augmenté λ pour les atteindre. Bien entendu, l'escalade binaire est une restriction locale de la stratégie d'escalade pure, et ne conduit pas nécessairement aux mêmes solutions : pour une pré-hiérarchie donnée, sa complétion «pure» de plus faible échelle d'apparition peut très bien nécessiter de passer par une chaîne d'étapes binaires qui ne sont pas toutes d'apparition minimale.

Une escalade réalise simultanément l'optimisation par programmation dynamique de l'énergie multi-échelles et la construction de la hiérarchie. À chaque étape, une région est ajoutée en fonction du résultat d'un pas de programmation dynamique. À la fin d'une escalade, la passe ascendante de l'optimisation multi-échelles a donc été faite. La passe descendante de calcul des disparitions des régions permet d'aboutir à la hiérarchie persistante pour l'énergie. C'est cette hiérarchie persistante - indicée par l'échelle d'apparition de ses éléments - que nous appelons finalement **hiérarchie d'escalade** (HE).

Un algorithme de construction d'une hiérarchie d'escalade binaire connexe est détaillé en annexe C.

Notons que cet algorithme est *totalement exempt de paramètre* étant donnée une énergie multi-échelles et une partition initiale fine, qui peut être la partition de l'image en pixels individuels.

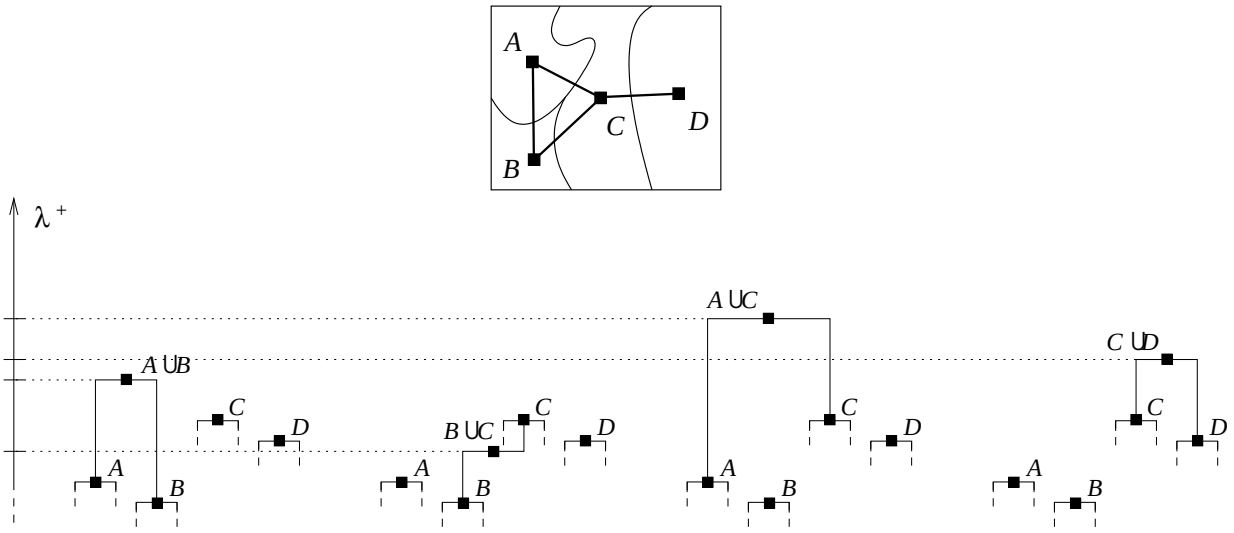


FIG. 8.4 – Escalade binaire connexe. L'escalade binaire connexe ne considère que les complétions par des unions de paires de sommets voisins. Pour la segmentation d'images (2D dans l'exemple), ces paires correspondent aux arêtes du graphe d'adjacence des sommets. Dans l'exemple ci-dessus, on ajouterait l'ensemble $B \cup C$ à la pré-hiérarchie courante, qui est celui qui possède la plus faible échelle d'apparition parmi les quatre complétions possibles. En l'occurrence, l'étape C ne serait pas persistante.

8.2 Propriétés des hiérarchies d'escalade

8.2.1 Invariance structurelle et covariance d'échelle

Homothétie sur les énergies

Théorème 4 *Si (H^*, λ^+) est la hiérarchie d'escalade sur X associée à l'énergie $\mathsf{E}_\lambda = (\mathsf{C}, \mathsf{D})$, alors la hiérarchie d'escalade sur X associée à $\mathsf{E}'_\lambda = (\mu\mathsf{C}, \nu\mathsf{D})$ est $(H^*, \frac{\nu}{\mu}\lambda^+)$.*

Preuve : Notons T la tranformation qui à une fonction affine $f(x) = ax + b$ associe la fonction affine $T(f)(x) = \mu ax + \nu b$. Si f et g sont deux affines, notons $I_x(f, g)$ l'abcisse de leur point d'intersection. On vérifie alors immédiatement que

$$I_x(T(f), T(g)) = \frac{\nu}{\mu} I_x(f, g). \tag{8.1}$$

Considérons maintenant un ensemble de fonctions affines $\{f_1 \dots f_n\}$ de pentes positives. Leur infimum $\inf\{f_i\}$ est une affine par morceaux composé de portions connexes des f_i (que nous supposons ordonnées pour simplifier l'écriture) :

$$\inf\{f_i\}(x) = \begin{cases} f_1(x) & \text{si} \quad x < x_1 \\ f_2(x) & \text{si} \quad x_1 \leq x < x_2 \\ \dots \\ f_n(x) & \text{si} \quad x \geq x_{n-1} \end{cases}. \tag{8.2}$$

Où les $x_i = I_x(f_i, f_{i+1})$ sont les abcisses de points d'intersection entre morceaux consécutifs. Alors, d'après 8.1, et comme $\inf\{f, g, h\} = \inf\{f, \inf\{g, h\}\}$

$$\inf\{T(f_i)\}(x) = \begin{cases} T(f_1)(x) & \text{si} \quad x < \frac{\nu}{\mu} x_1 \\ T(f_2)(x) & \text{si} \quad \frac{\nu}{\mu} x_1 \leq x < \frac{\nu}{\mu} x_2 \\ \dots \\ T(f_n)(x) & \text{si} \quad x \geq \frac{\nu}{\mu} x_{n-1} \end{cases}. \tag{8.3}$$

Par convention, si $F = \inf\{f_i\}$, nous notons $T(F) = \inf\{T(f_i)\}$.

On vérifie également aisément qu'une somme d'affines par morceaux ayant été construites par infimum de fonctions affines se comporte de manière analogue par rapport à la transformation T sur les fonctions affines qui l'engendrent : si $F = \inf\{f_i\}$ et $G = \inf\{g_i\}$ alors $F + G$ est également de la forme 8.2 et $T(F) + T(G) = T(F + G)$ est de la forme 8.3.

La preuve du théorème s'effectue alors par récurrence sur l'étape de programmation dynamique.

Soit H_k et H'_k les pré-hiérarchies obtenues après k pas d'escalade utilisant respectivement E_λ et E'_λ. Notons E^*_x et E'^*_y respectivement l'énergie partielle d'un élément x dans H_k et y dans H'_k. De même, $\lambda^+(x)$ et $\lambda'^+(y)$ représentent les apparitions respectives de x et y.

Formulons alors l'hypothèse de récurrence R_k :

i. $H_k = H'_k$.

ii. $\forall x \in H'_k \quad \mathsf{E}'^*_x = T(\mathsf{E}^*_x)$.

iii. $\forall x \in H'_k \quad \lambda'^+(x) = \frac{\nu}{\mu}\lambda^+(x)$.

• Évidemment $H_0 = H'_0 = \{\{x\}\}_{x\in X}$. En outre, les énergies partielles des éléments de base sont leurs énergies propres qui sont affines, et toutes leurs apparitions vallent $-\infty$ (ou 0 si l'on se restreint au demi axe positif des échelles). R_0 est donc vraie.

• Supposons que R_k est vraie. Comme $H_k = H'_k$, leurs complétions possibles sont identiques, i.e. $H^+_k = H'^+_k = H^+$ (et de même pour les restrictions binaire et binaire connexe).

Considérons $x \in H^+$. Son énergie propre E'_x pour E'_λ vaut $T(\mathsf{E}_x)$, où E_x est l'énergie propre de x pour E_λ. Si $\mathcal{F}(x)$ est l'ensemble des fils de x alors

$$\begin{aligned}
\mathsf{E}'^*_x &= \inf\{\mathsf{E}'_x, \sum_{s\in\mathcal{F}(x)} \mathsf{E}'^*_x\} = \inf\{T(\mathsf{E}_x), \sum_{s\in\mathcal{F}(x)} T(\mathsf{E}^*_x)\} \\
&= \inf\{T(\mathsf{E}_x), T(\sum_{s\in\mathcal{F}(x)} \mathsf{E}^*_x)\} = T(\inf\{\mathsf{E}_x, \sum_{s\in\mathcal{F}(x)} \mathsf{E}^*_x\}) = T(\mathsf{E}^*_x).
\end{aligned}$$

Par ailleurs, aussi bien pour les opérations d'inf et de sommation, les points de rupture de pente d'un affine par morceaux (les x_i ci-dessus) se trouvent multipliés par ν/μ après transformation par T. Le morceau de $\sum_{s\in\mathcal{F}(x)} \mathsf{E}'^*_x$ qu'intersecte E'_x est donc le transformé par T du morceau de $\sum_{s\in\mathcal{F}(x)} \mathsf{E}^*_x$ qu'intersecte E_x. Comme $\mathsf{E}'_x = T(\mathsf{E}_x)$, d'après 8.1

$$\lambda'^+(x) = \frac{\nu}{\mu}\lambda^+(x).$$

L'ordre entre les différentes hypothèses d'escalade est donc le même pour E_λ et E'_λ. La $k+1^{\text{ème}}$ région ajoutée à H_k et H'_k est donc identique - $H_{k+1} = H'_{k+1}$ - et son énergie partielle et son échelle d'apparition vérifient R_{k+1}i. et ii. ■

La *structure* d'une analyse ensembles-échelle obtenue par escalade - i.e. la hiérarchie - est donc invariante par transformation linéaire des énergies. L'ajustement du poids *relatif* des énergies a seulement un effet d'homothétie sur l'échelle finale. La description est donc covariante d'échelle. Ainsi, **le réglage du poids relatif des deux énergies est totalement transféré au paramètre d'échelle de la représentation finale**. Ajuster le poids des énergies *a priori* ou *a posteriori* a exactement le même effet. Il suffit finalement de spécifier les énergies à un facteur multiplicatif près, c'est-à-dire de spécifier *qualitativement* à quelles quantités elles doivent être proportionnelles.

Les analyses ensembles-échelle obtenues par escalade réalisent donc l'objectif poursuivi : extraire le paramètre d'échelle du bas niveau d'analyse. L'échelle devient un paramètre de réglage au sens fort : c'est un potentiomètre dont disposent les algorithmes ultérieurs d'interprétation d'image pour naviguer dans le contenu d'une scène.

Transposition d'invariances

Cette propriété permet également de convertir certaines qualités d'invariance des énergies en qualités d'invariance de la représentation multi-échelles finale.

Par exemple, si l'attache aux données D est telle que pour toute transformation T des valeurs d'une image appartenant à une classe $\mathcal{T}$ donnée

$$\mathsf{D}(T(R)) = k(T)\mathsf{D}(R)$$

où $k(T)$ est une constante qui ne dépend que de T, alors une HE bâtie sur D est structurellement invariante par application d'une transformation de la classe $\mathcal{T}$ à l'image, et voit simplement son axe des échelles dilaté du facteur $k(T)$.

Un exemple typique est fourni par une attache aux données exprimée sous forme de déviations quadratiques. Pour toute multiplication par une constante c des valeurs de l'image, on a

$$\sum_{x\in R}(cI(x) - \overline{cI_R})^2 = c^2 \sum_{x\in R}(I(x) - \overline{I_R})^2.$$

Si D est de ce type, une transformation linéaire de l'image provoque donc une transformation linéaire de la HE (d'un facteur c^2), ce que l'on peut noter symboliquement

$$\mathsf{HE}(cI) = c^2\mathsf{HE}(I).$$

Notons que ce type de propriété reste valable pour des écarts mesuré en norme L_p par rapport à n'importe quelle surface de régression du moment que l'estimateur des paramètres de la surface est lui-même invariant par tranformation linéaire des radiométries.

Des résultats similaires s'obtiennent pour des transformations de l'énergie de «régularisation» C.

Considérons par exemple un terme de régularisation à la Mumford-Shah/Potts/Freeman, proportionnel à la longueur des frontières de la partition. Associons-le à un terme D de déviations quadratiques. Envisageons alors une opération d'homothétie géométrique de l'image traitée, par laquelle toutes les longueurs sont multipliées par un facteur h et toutes les surfaces par h^2. Supposons que cette homothétie n'affecte pas les statistiques internes aux régions (moyenne, variance), ce qui, pour les images discrètes, revient à négliger les effets d'échantillon. Alors, pour toute région R, nous pouvons écrire (symboliquement)

$$\mathsf{C}(hR) = h\mathsf{C}(R) \quad \text{et} \quad \mathsf{D}(hR) = h^2\mathsf{D}(R).$$

Et donc

$$\mathsf{HE}(hI) = h\mathsf{HE}(I).$$

La hiérarchie d'escalade associée au modèle constant par morceaux de Mumford-Shah est donc invariante par homothétie *géométrique* de l'image.

8.2.2 Une solution unique aux deux problèmes duaux de débit/distorsion

Les problèmes de caricature/compression optimale s'appuyant sur des descriptions en régions d'une image proposés par Salembier et Garrido supposent que l'on dispose au préalable d'une hiérarchie de régions. Salembier et Garrido utilisent un algorithme «quelconque» de fusion pour produire cette hiérarchie, i.e. un algorithme qui n'a pas d'affinité particulière avec les problèmes d'optimisation posés. A contrario, la hiérarchie d'escalade se présente comme une solution particulièrement intéressante à ces problèmes.

Nous avons vu avec la proposition 13 que l'échelle d'apparition d'une région s'écrivait

$$\lambda^+(x) \quad = \quad -\frac{\Delta\mathsf{D}}{\Delta\mathsf{C}} \quad = \quad -\frac{\mathsf{D}(\{x\}) - \mathsf{D}(P)}{\mathsf{C}(\{x\}) - \mathsf{C}(P)} \tag{8.4}$$

où P est la partition de la pré-hiérarchie partielle $H(\{x\})\backslash x$ qui est optimale en $\lambda^+(x)$.

Il est important de noter que même dans un algorithme de fusion de paires de régions, la partition P qui correspond à une hypothèse de fusion $x \cup y$ n'est pas nécessairement $\{x, y\}$ (ce sera illustré par la figure 8.5 ci-dessous).

On a toujours $\Delta\mathsf{C} \leq 0$, par sous-additivité de C. $-\Delta\mathsf{C}$ représente la simplification de modèle quand la coupe optimale passe de P à $\{x\}$ et $\Delta\mathsf{D}$ représente la diminution de fidelité du modèle à l'image. Le critère d'escalade sélectionne la région de plus faible apparition. Il choisit donc la région qui réalise le meilleur compromis entre simplification du modèle et diminution de fidélité, au sens de leur ratio (notons à nouveau que multiplier les énergies par des constantes change les valeurs absolues des ratios mais par leur ordre).

Dans le cadre débit/distorsion, C représente le volume de codage d'une représentation simplifiée de l'image et D quantifie la distorsion visuelle entre la représentation et l'image. Pour le problème de moindre distorsion à volume borné ($\min \mathsf{D} \quad \text{avec} \quad \mathsf{C} < \mathsf{C}_0$), l'apparition d'une région s'interprète donc comme le ratio

$$-\frac{\text{Variation de distorsion}}{\text{Variation de volume}}.$$

Minimiser l'apparition semble alors une heuristique locale raisonnable pour tenter d'obtenir la plus faible distorsion quand le quota de volume sera atteint : on a intérêt à faire diminuer le volume le plus rapidement possible pour vite atteindre C_0, mais on a simultanément intérêt à ne pas trop augmenter la distorsion[1].

Considérons maintenant le problème dual : $\min \mathsf{C} \quad \text{avec} \quad \mathsf{D} < \mathsf{D}_0$. Dans ce cas, D est sur-additive et nous avons vu que l'optimisation du problème Lagrangien devait se faire par une version complémentaire de l'algorithme d'optimisation multi-échelles. Les coupes décroissent en fonction de λ et l'échelle calculée par intersection entre les énergies propres et partielles correspond à *l'échelle de disparition* d'une région. L'axe des échelles pointe donc en sens inverse de la construction de la hiérarchie. Considérons alors la reparamétrisation de l'énergie par $\mu = 1/\lambda$ afin de voir défiler les coupes dans le sens ascendant. La stratégie d'escalade consiste alors à choisir l'élément x d'échelle d'apparition $\mu_{app}(x)$ pour E_μ minimale, soit le x d'échelle de disparition $\lambda_{dis}(x)$ maximale pour E_λ. Il faut donc utiliser une stratégie d'escalade *duale* :

Définition 33 *Si E_λ est une énergie affine séparable bâtie sur un terme C sur-additif, la stratégie d'**escalade duale** considère la complétion croissante qui possède la plus forte échelle de disparition.*

Or, pour le problème dual, l'échelle de disparition d'un nœud s'écrit :

$$-\frac{\Delta\mathsf{C}}{\Delta\mathsf{D}}$$

et itérer le choix du nœud qui maximise cette quantité équivaut à itérer le choix de celui qui minimise son inverse, donc

Théorème 5 *Si E_λ est une énergie multi-échelles pure alors les hiérarchies d'escalade correspondant aux deux problèmes duaux de débit/distorsion associés à E_λ **coïncident**.*

[1] Ce problème peut s'envisager sous un angle ludique. Imaginons un jeu dissymétrique dans lequel un homme H est confronté à une machine M de la manière suivante :

Le jeu se présente en coups successifs. À chaque coup, l'ordinateur M tire un ensemble de couples $\{(c_i, d_i)\}$ de nombres aléatoires (ou dont H est incapable de modéliser la loi d'apparition) et H doit choisir l'un d'entre eux. Si le choix de H est (c_k, d_k) alors c_k s'ajoute à la perte globale C de M et d_k s'ajoute à la perte globale D de H. Mais M est une machine fatiguable, et quand le désagrément qu'elle a subi excède une valeur C_0 donnée par avance, i.e. quand $\mathsf{C} > \mathsf{C}_0$, le jeu s'arrête. Malheureusement H ne connait pas C_0 et souhaite également avoir subi le moindre désagrément à l'issue du jeu (minimiser D). Quel choix H doit-il faire à chaque coup ?

Nous n'avons pas trouvé trace de ce problème dans les ouvrages standards de théorie des jeux. Si une solution est connue des spécialistes nous sommes curieux de l'apprendre. Si elle n'est pas encore connue, peut-être qu'un Von Neumann comtemporain s'y intéressera ...

Nous voyons désormais qu'une hiérarchie d'escalade bâtie sur une énergie multi-échelles pure établit une symétrie complète entre les deux termes en opposition. Du point de vue de l'escalade, les deux problèmes duaux d'optimisation sous contrainte (perte/distorsion bornée) ont la même solution.

Remarquons que cette propriété - comme celle d'invariance par homothétie - reste valide pour toutes les formes d'escalade, pure ou binaire, connexe ou non. Elle est uniquement liée au critère de choix du mouvement dans l'espace des pré-hiérarchies, indépendamment de toute contrainte imposée sur les différents choix possibles.

8.3 Interprétations du principe d'escalade

8.3.1 Escalade et descentes

La stratégie d'escalade correspond à un point de vue radicalement différent de celui adopté lorsque l'on réalise une fusion de régions par descente (voir la figure 8.5).

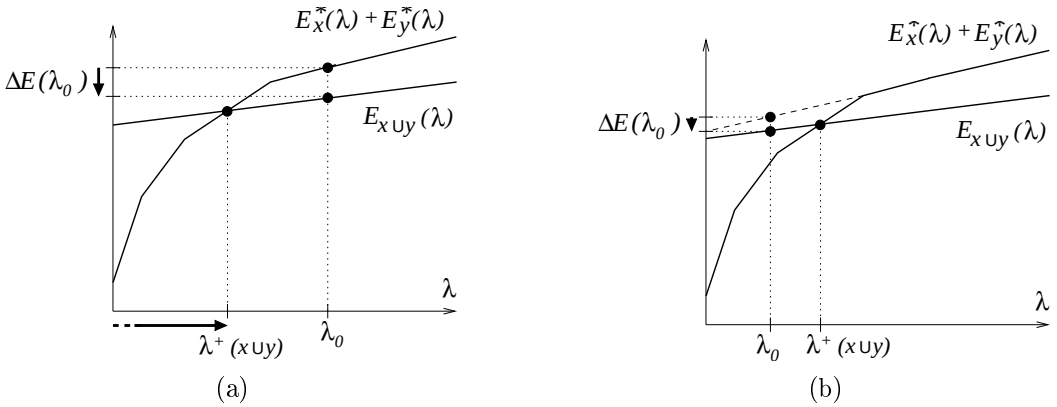


FIG. 8.5 – Escalade et descente. Les courbes représentent l'énergie propre et partielle d'une hypothèse de fusion $x \cup y$ au cours d'un algorithme de groupement ascendant. Une fusion sur une critère de descente d'énergie considère une échelle *fixe* λ_0 et attribue une priorité de fusion à $x \cup y$ proportionnelle à la variation d'énergie $\Delta \mathsf{E}(\lambda_0)$ entre $\{x, y\}$ et $\{x \cup y\}$ à λ_0. Le critère d'escalade considère lui la plus faible valeur de λ pour laquelle l'énergie décroit, qui correspond à l'échelle d'apparition $\lambda^+(x \cup y)$. Dans une descente, la priorité accordée aux différentes fusions possibles est dépendante de λ_0. En outre, la figure (b) montre que si λ_0 est fixé à une valeur plus faible que $\lambda^+(x \cup y)$, la fusion se fait toujours en comparant l'énergie de $x \cup y$ à celle de la partition $\{x, y\}$. C'est pourtant une autre partition - inférieure dans la hiérarchie partielle $H(x \cup y)$ - qui est dans ce cas optimale à l'échelle λ_0. En substance, une escalade, en mémorisant l'historique complet des fusions antérieures sous forme de l'énergie des coupes optimales de $H(x)$ et $H(y)$, compare l'énergie de $\{x \cup y\}$ à celle de toutes les partitions optimales de $H(x \cup y) \backslash (x \cup y)$.

S'effectuant à échelle donnée, une descente est intrinsèquement mono-échelle, et si l'on désire parvenir à des taux de régularisation forts, il est nécessaire d'enchaîner les descentes selon une suite croissante de valeurs de λ. En effet, la minimisation d'une énergie fortement régularisée par une unique descente conduit à de très mauvaises solutions. En début d'aggrégation, la fusion est trop fortement guidée le modèle a priori et ne tient pas compte du contenu image. L'algorithme étant sans retour, ces erreurs initiales ne peuvent pas être corrigées par la suite.

De son côté, le critère d'escalade est indépendant de toute valeur ou suite de valeurs d'échelles fixée au préalable, en préconisant précisément l'augmentation la plus progressive possible de λ : si l'on cherche à effectuer une fusion par descente en fixant le paramètre d'échelle en dessous de la plus faible apparition des complétions possibles, on ne trouve aucun candidat à la fusion. On peut donc voir la stratégie d'escalade

comme la limite de la stratégie d'itération de descentes quand l'échantillonnage de l'axe des échelles tend vers un échantillonnage continu.

Toutefois, il n'est pas possible de réaliser physiquement ce passage à la limite par une approche directe. L'algorithme d'escalade consiste alors à calculer explicitement l'échelle à laquelle survient le prochain événement de fusion et à retenir cet événement. La figure 8.5 montre qu'il s'agit en quelque sorte de deux points de vue orthogonaux : une descente s'appuie sur une quantité située sur l'axe des énergies alors que l'escalade considère l'axe des échelles.

Remarque : On pourrait considérer, sur l'axe des énergies, une priorité dépendant de l'énergie d'une union à son échelle d'apparition, i.e. de $\mathsf{E}(\lambda^+(x \cup y))$. Cette stratégie ne donne cependant pas de bons résultats. La raison fondamentale est que l'énergie est une grandeur absolue, qui dépend de façon prépondérante de la taille des régions, ce qui a tendance à homogénéiser la granularité des descriptions, indépendamment des contrastes. En outre, aucune des bonnes propriétés des hiérarchies d'escalade qui sont discutées dans ce chapitre ne sont vérifiées par fusion sur le critère de moindre énergie à échelle d'apparition.

8.3.2 Escalade et continuation

On peut également interpréter l'escalade comme une méthode de continuation (cf. 6.1.5 p.130) contrainte à produire une séquence de solutions monotones.

En effet, pour une attache aux données réaliste, l'optimum global à échelle nulle est la sur-partition absolue. Par définition, la première fusion par escalade *pure*, disons qui se produit à λ_0, correspond au minimum global de E_{λ_0} sur l'ensemble des partitions possible de l'image. La seconde fusion atteint alors un optimum à $\lambda_1 > \lambda_0$ qui est global dans l'espace des sous-partitions de la première partition, etc. L'escalade pure suit donc une ligne *monotone* d'optima *globaux*. On voit en particulier que si pour une énergie particulière, les solutions globales sur $\mathbb{P}(X)$ sont monotones, alors l'escalade pure les suit.

Par contre, les escalades binaires suivent un chemin monotone à pas serrés, en ne considérant à chaque étape que les partitions couvrantes. Elles peuvent donc s'engager dans une voie locale non optimale à long terme. Notons que dans l'esprit multi-échelles qui a été développé, l'optimisation ne vise pas à atteindre graduellement une solution pour une échelle spécifique. L'escalade «continue» jusqu'à fusion complète de l'image et s'intéresse à toutes les étapes de la continuation.

8.3.3 Escalade et prédicats naturels

Les stratégies de fusion s'appuyant sur des prédicats d'homogénéité naturels peuvent également s'interpréter comme des stratégies d'escalade.

Considérons en effet un critère croissant D (une mesure d'hétérogénéité de région) et le prédicat booléen naturel associé $\mathcal{B}^\lambda_{\mathsf{D}}$ qui est vrai si et seulement si l'hétérogénéité d'une région est bornée par λ. On sait alors que les coupes maximales pour $\mathcal{B}^\lambda_{\mathsf{D}}$ dans une hiérarchie H sont croissantes en fonction de λ. Les nœuds de H peuvent donc être caractérisés par leur échelle d'apparition dans une coupe. L'échelle d'apparition d'un nœud x est alors donnée par

$$\begin{aligned}
\lambda^+(x) &= \min \ \{ \ \lambda \ | \ \mathcal{B}^\lambda_{\mathsf{D}}(x) \ \text{ est vrai } \ \} \\
&= \min \ \{ \ \lambda \ | \ \mathsf{D}(x) \leq \lambda \ \} \\
&= \mathsf{D}(x).
\end{aligned}$$

Le critère d'escalade préconise donc de fusionner les régions les plus homogènes d'abord...

Notons que si le critère D n'est pas croissant, la stratégie d'escalade reste néanmoins valable et conduit à la définition d'une hiérarchie persistante : certaines étapes se présentent comme des maxima locaux d'hétérogénéité, vérifiant $\lambda^-(x) < \lambda^+(x)$, et doivent être shuntées pour obtenir une structure multi-échelles.

Les approches par prédicats naturels peuvent également se formuler comme des problèmes d'optimisation. En effet, une partition est λ-homogène si et seulement si toutes ses régions le sont. Considérons donc l'énergie de «distorsion» définie par

$$\mathsf{D}(P) \triangleq \max_{R \in P} \mathsf{D}(R). \tag{8.5}$$

Par ailleurs, utilisons la fonction de débit la plus simple possible en la matière de

$$\mathsf{C}(P) \triangleq |P| = \sum_{R \in P} 1.$$

Trouver une partition λ-homogène maximale (au sens de l'ordre sur les partitions) est alors équivalent à trouver une partition P qui minimise $\mathsf{C}(P)$ sous la contrainte $\mathsf{D}(P) \leq \lambda$.

À cause du max dans D, la solution n'est pas nécessairement unique dans $\mathbb{P}(X)$. Elle le devient dans une hiérarchie si le le critère D d'hétérogénéité sur les *parties* est croissant dans le treillis $\mathcal{P}(X)$ des parties. L'énergie D définie sur les *partitions* est alors également croissante dans le treillis $\mathbb{P}(X)$ des partitions. De son côté, C est sous-additive et conduit à une énergie de partitions décroissante. La formulation est donc une formulation multi-échelles pure (notons que l'expression sous contrainte correspond au problème de débit/distorsion primal).

Finalement, la condition importante est la monotonie des énergies définies sur les partitions. Dans le cas d'une formulation additive (séparable) de ces énergies (équ. 7.1), l'énergie de parties correspondante doit être sous ou sur-additive. Dans le cas d'une formulation utilisant un max (équ. 8.5), l'énergie de parties doit être monotone.

Remarque : ici le Langrangien n'est pas une énergie séparable : le max dans le terme D ne l'est pas. La résolution Lagrangienne ne peut donc pas se faire par programmation dynamique. En particulier, la formule $-\frac{\Delta D}{\Delta C}$ pour l'échelle d'apparition n'est pas valable. Il faut absolument regarder le problème sous sa forme de problème d'optimisation sous contrainte.

8.3.4 Interprétation continue

L'expression $-\frac{\Delta \mathsf{D}}{\Delta \mathsf{C}}$ de l'échelle d'apparition est frappante : le rapport des accroissements fait penser à une approximation discrète du gradient de D par rapport à C, si cette quantité a un sens. Sélectionner le mouvement élémentaire qui minimise l'apparition s'apparente alors formellement à une descente selon ce gradient. Examinons de plus près cette analogie, en supposant que l'espace de recherche est continu et non discret.

Soit $\mathcal{S}$ un espace vectoriel (pendant de l'espace $\mathbb{P}(X)$ des partitions de l'image). Soit C et D deux fonctions réelles sur $\mathcal{S}$ (volume/distorsion) supposées différentiables. Étant donné $\alpha \in \mathbb{R}$, on pose le problème de trouver $S^* \in \mathcal{S}$ tel que

$$\begin{aligned} S^* &= \operatorname*{argmin}_{S \in \mathcal{S}} \ \mathsf{D}(S) \\ \text{avec} \quad \mathsf{C}(S^*) &\leq \alpha \end{aligned}$$

Soit S_0 une solution initiale (la sur-partition initiale). On suppose que $\mathsf{C}(S_0) > \alpha$ et que le problème admet une solution. Pour tout $S \in \mathcal{S}$, soit $\mathcal{V}(S)$ le voisinage de S, l'ensemble des éléments accessibles en un pas depuis S (l'ensemble des partitions couvrant S). L'heuristique d'escalade correspond à suivre un chemin dans $\mathcal{S}$ en sélectionnant chaque pas S_{i+1} en fonction de S_i par :

$$\begin{aligned} S_{i+1} &= \operatorname*{argmin}_{S \in \mathcal{V}(S_i)} \ \Delta(S_i, S) \\ \text{avec} \quad \Delta(S_i, S) &= -\frac{\mathsf{D}(S) - \mathsf{D}(S_i)}{\mathsf{C}(S) - \mathsf{C}(S_i)} = -\frac{\Delta \mathsf{D}}{\Delta \mathsf{C}}. \end{aligned}$$

Supposons maintenant que l'on effectue à chaque étape un mouvement infinitésimal ϵ dans une direction d de la sphère unitaire $SU(\mathcal{S})$ de $\mathcal{S}$, et que l'on sélectionne à chaque pas la direction qui minimise

$$\begin{aligned}\Delta_d(S_i,\epsilon) &= -\frac{\mathsf{D}(S_i+\epsilon d)-\mathsf{D}(S_i)}{\mathsf{C}(S_i+\epsilon d)-\mathsf{C}(S_i)}\\ &= -\frac{\mathsf{D}(S_i+\epsilon d)-\mathsf{D}(S_i)}{\epsilon}\frac{\epsilon}{\mathsf{C}(S_i+\epsilon d)-\mathsf{C}(S_i)}\end{aligned}$$

En passant à la limite sur ϵ on obtient

$$\Delta_d(S_i) = \lim_{\epsilon\to 0^+}\Delta_d(S_i,\epsilon) = -\frac{\frac{\partial \mathsf{D}}{\partial d}}{\frac{\partial \mathsf{C}}{\partial d}}(S_i).$$

Voyons maintenant comment s'interprète cette quantité.

Posons pour commencer la condition de sous-additivité de C, c'est-à-dire supposons que *tous les mouvements élémentaires autorisés provoquent une décroissance de la fonction de contrainte* C, i.e. que

$$\forall x\in\mathcal{S}\quad \forall y\in\mathcal{V}(x)\qquad \mathsf{C}(y)<\mathsf{C}(x).$$

Notons que cette condition est irréaliste pour les problèmes énergétiques continus : elle suppose que l'on se trouve à chaque instant en un maximum local de C, ce qui est évidemment impossible. Cette analogie continue doit simplement être vue comme une *transcription formelle* de la situation *locale* dans laquelle l'optimisation discrète se trouve au moment de l'application du critère d'escalade.

Plaçons-nous en $x\in\mathcal{S}$ et supposons que l'on veuille faire décroître d'une quantité α la valeur de C, et donc passer de $\mathsf{C}(x)$ à $\mathsf{C}(x)-\alpha$. Cette opération peut s'interpréter comme un passage de x à un point de l'ensemble de niveau défini par $\mathsf{C}=\mathsf{C}(x)-\alpha$ (voir figure 8.6).

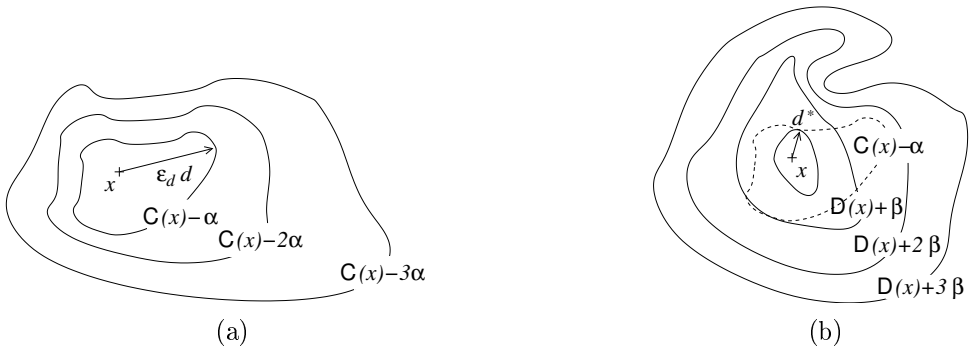


FIG. 8.6 – Escalade continue. (a) Courbes de niveau locales de C. (b) Courbes de niveau locales de D superposées à la courbe à $\mathsf{C}(x)-\alpha$ de C (en pointillés). La direction d^* sélectionnée par le critère d'escalade est celle qui entraine la plus faible variation de D pour une variation de constante de C.

Autour de x, on a

$$\mathsf{C}(x+\epsilon d)=\mathsf{C}(x)+\epsilon\frac{\partial\mathsf{C}}{\partial d}(x)+o(\epsilon).$$

Dans la direction d, le pas ϵ_d à effectuer pour obtenir $C(x+\epsilon_d d)=C(x)-\alpha$ vaut donc

$$\epsilon_d=-\frac{\alpha+o(\epsilon_d)}{\frac{\partial\mathsf{C}}{\partial d}(x)}. \tag{8.6}$$

Par un tel pas ϵ_d, la variation d'énergie d'attache aux données D est alors

$$\begin{aligned}\Delta\mathsf{D}_\alpha(x,d) &= \mathsf{D}(x+\epsilon_d d)-\mathsf{D}(x)\\ &= \epsilon_d\frac{\partial\mathsf{D}}{\partial d}(x)+o(\epsilon_d)\\ &= -(\alpha+o(\epsilon_d))\frac{\frac{\partial\mathsf{D}}{\partial d}(x)}{\frac{\partial\mathsf{C}}{\partial d}(x)}+o(\alpha).\end{aligned}$$

D'après l'équation (8.6), un $o(\epsilon_d)$ est également un $o(\alpha)$, donc

$$\Delta\mathsf{D}_\alpha(x,d)=-\alpha\frac{\frac{\partial\mathsf{D}}{\partial d}(x)}{\frac{\partial\mathsf{C}}{\partial d}(x)}+o(\alpha).$$

Intéressons-nous alors à la direction d qui minimize -pour α fixé- la variation de D, et suivons l'évolution de cette direction quand α tend vers 0. C'est-à-dire considérons

$$d^*(x)=\lim_{\alpha\to 0^+}\operatorname*{argmin}_{d\in SU(\mathcal{S})}\ \Delta\mathsf{D}_\alpha(x,d).$$

On voit alors que pour α assez petit

$$\begin{aligned}\operatorname*{argmin}_{d\in SU(\mathcal{S})}\ \Delta\mathsf{D}_\alpha(x,d) &= \operatorname*{argmin}_{d\in SU(\mathcal{S})}\ -\alpha\frac{\frac{\partial\mathsf{D}}{\partial d}(x)}{\frac{\partial\mathsf{C}}{\partial d}(x)}+o(\alpha)\\ &= \operatorname*{argmin}_{d\in SU(\mathcal{S})}\ -\frac{\frac{\partial\mathsf{D}}{\partial d}(x)}{\frac{\partial\mathsf{C}}{\partial d}(x)}\end{aligned}$$

qui est indépendant de α, et donc

$$d^*(x)=\operatorname*{argmin}_{d\in SU(\mathcal{S})}\ -\frac{\frac{\partial\mathsf{D}}{\partial d}(x)}{\frac{\partial\mathsf{C}}{\partial d}(x)}.$$

En continu, le critère d'escalade s'interprète donc comme le choix local de la direction qui minimise l'augmentation de distorsion D pour une variation infinitésimale donnée du volume C.

On peut véritablement le voir comme une descente du gradient de la fonction qui associe localement C à D. Considérons en effet une demi-droite $\mathcal{D}$ issue de x et paramétrons-la par la distance σ à x. Comme $\mathsf{C}(\sigma)$ est strictement décroissante au voisinage de x, elle est localement inversible. Notons $\Sigma:\mathsf{C}\mapsto\Sigma(\mathsf{C})$ son inverse. La fonction $\mathsf{C}\mapsto\mathsf{D}o\Sigma(\mathsf{C})$ qui aux valeurs de C le long de $\mathcal{D}$ associe les valeurs de D est donc bien définie et dérivable. En x, la quantité

$$-\frac{\partial\mathsf{D}o\Sigma}{\partial\mathsf{C}}(\mathsf{C}(x))$$

possède donc un sens et représente l'analogue continu du critère sur lequel s'appuie l'escalade discrète.

8.4 Formulation finale du problème de segmentation multi-échelles

Une stratégie de descente d'énergie à λ_0 est une heuristique qui vise à minimiser E_{λ_0} et assure de trouver un minimum local (au sens de la 2-normalité). De même, une continuation cherche à minimiser *une* énergie donnée. Ces stratégies sont donc définies par un *objectif* clairement identifié. L'escalade, elle, n'est pas définie de cette manière, mais comme un *processus* de construction. Sélectionner à chaque étape la solution qui nécessite la moindre régularisation est un critère relativement intuitif, certes, tout comme la fusion des régions les plus homogènes d'abord dans les approches par prédicats, mais la structure globale que l'algorithme produit, ou tente d'approximer, n'est pas identifiée.

Nous nous sommes alors interrogés sur la manière de formuler le problème de description d'images par parties en termes véritablement multi-échelles. Dans un cadre énergétique, cela revient à définir un critère d'optimalité globale pour les structures multi-échelles : les hiérarchies.

Nous commençons par décrire la première idée qui vient naturellement, pour voir qu'elle conduit à une solution triviale. Cette première idée nous conduit néanmoins à une deuxième définition qui s'avère cette fois pertinente. Nous montrons alors que la stratégie d'escalade correspond à une heuristique gloutonne localement optimale pour tenter d'exhiber cette description multi-échelles optimale.

8.4.1 Principe d'indifférence d'échelle

Une idée naturelle pour définir l'optimalité globale d'une hiérarchie H est de considérer non pas la minimisation de E_λ dans H pour une seule valeur d'échelle mais à toutes les échelles. Cette idée est conforme au principe d'indifférence d'échelle : si aucune échelle n'est privilégiée, il faut rechercher une solution qui soit simultanément bonne pour toutes les échelles.

Considérons une hiérarchie H sur X. Notons comme précédemment E^*_H la fonction d'énergie des coupes minimales de H en fonction de λ (après programmation dynamique, E^*_H est l'énergie partielle du sommet X de H). Optimiser E_λ pour toutes les échelles peut alors s'envisager comme la minimisation de la somme des énergies des solutions mono-échelle, ce qui revient à définir l'énergie totale d'une hiérarchie comme l'intégrale de E^*_H et rechercher la hiérarchie qui minimise cette intégrale. Cette intégrale diverge cependant car E^*_H diverge en $\mp\infty$. Il y néanmoins certaines bornes naturelles d'intégration : 0 inférieurement pour les énergies «pures» et l'échelle d'apparition de X supérieurement, qui correspond à l'échelle pour laquelle la coupe optimale atteint le sommet de la hiérarchie.

On peut alors définir l'énergie totale d'une hiérarchie par

$$\mathsf{E}(H) \triangleq \int_0^{\lambda^+(X)} \mathsf{E}^*_H(\lambda)d\lambda \tag{8.7}$$

et poser le problème de segmentation multi-échelles comme celui de trouver la hiérarchie H de moindre énergie totale (cf. figure 8.7).

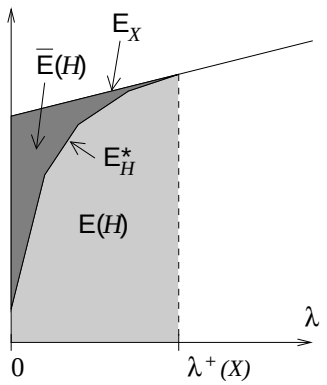


FIG. 8.7 – Énergie et co-énergie totale d'une hiérarchie.

Cette définition n'est cependant pas satisfaisante : son optimum est systématiquement la hiérarchie «creuse», qui ne comporte aucun ensemble intermédiaire entre la base et le sommet. En effet, la borne supérieure d'intégration n'est pas fixe mais dépend de H. X est présent dans toutes les solutions et a toujours la même énergie propre, cependant son échelle d'apparition dépend de la structure interne de la hiérarchie. On constate alors que toute complétion intérieure d'une hiérarchie par un ensemble dont l'apparition est située entre celle du sommet et celle de ses fils : i) repousse systématiquement l'échelle

d'apparition de X et ii) augmente systématiquement l'énergie totale de la hiérarchie (cf. fig. 8.8(a)). Toute hiérarchie pouvant se construire par ajout d'ensembles intérieurs depuis la hiérarchie creuse, dans l'ordre de leurs apparitions, la hiérarchie creuse possède une énergie totale inférieure à toutes les autres (on voit également que le critère d'escalade ne tend pas non plus à minimiser l'échelle d'apparition de l'image entière).

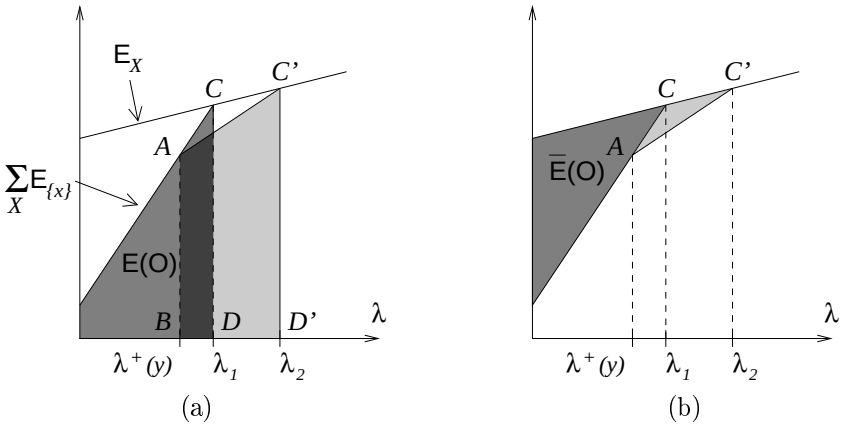


FIG. 8.8 – La hiérarchie d'énergie minimale est toujours creuse (a) mais celle de co-énergie maximale ne l'est généralement pas (b). Soit O la hiérarchie creuse et y un ensemble qui n'est ni un singleton ni l'ensemble entier et dont l'échelle d'apparition se situe avant celle λ_1 du sommet de O. Soit $Y = O \cup y$ et λ_2 la nouvelle apparition de X dans cette hiérarchie. (a) illustre que $\mathsf{E}(Y) \geq \mathsf{E}(O)$: quand on passe de O à Y, on perd l'énergie correspondant à la surface $ACDB$ et l'on gagne celle correspondant à $AC'D'B$, qui est supérieure (par homothétie). (b) illustre que $\overline{\mathsf{E}}(Y) \geq \overline{\mathsf{E}}(O)$: la co-énergie totale augmente également toujours, de la surface du triangle ACC'.

8.4.2 Minimiser la co-énergie totale d'une hiérarchie

Faire reposer la définition de l'énergie totale d'une hiérarchie sur une borne variable d'intégration conduit à une solution triviale. Définissons alors une intégrale «complémentaire» de celle de E^*_H, qui converge, elle.

X est systématiquement dans H. On vérifie par ailleurs facilement que E^*_H est l'infimum d'au plus $|X|$ fonctions affines, et que l'une d'elles est toujours la fonction d'énergie propre de X, E_X. Intéressons-nous donc à la différence entre l'énergie propre d'une image (de X) et celle de ses coupes hiérarchiques :

Définition 34 *La fonction*

$$\overline{E}^*_H \triangleq E_X - E^*_H \tag{8.8}$$

est appelée la ***co-énergie*** *des coupes optimales de H pour E_λ.*

Si C est *strictement* sous-additive, alors l'apparition de X est finie et $\overline{\mathsf{E}}^*_H$ est nulle à partir de $\lambda^+(X)$. L'intégrale de $\overline{\mathsf{E}}^*_H$ sur $\mathbb{R}^+$ est donc finie et

Définition 35 *La quantité*

$$\overline{E}(H) \triangleq \int_0^{+\infty} \overline{E}^*_H(\lambda) d\lambda \tag{8.9}$$

est appelée la ***co-énergie totale*** *de la hiérarchie H (pour une énergie multi-échelles donnée).*

Nous proposons alors de poser le problème de segmentation multi-échelles comme celui de **trouver la hiérarchie de co-énergie totale maximale**.

Cette fois, le problème n'est pas trivial (voir figure 8.8(b)). Toutefois, présentée sous un angle purement énergétique, la solution de ce problème n'est pas unique : une hiérarchie contenant des ensembles non

persistants (jamais optimaux pour E_λ) a la même co-énergie que sa hiérarchie persistante associée. Nous choisissons naturellement le représentant persistant de chaque classe. Finalement, une bonne formulation du problème de segmentation multi-échelles semble être :

Problème de segmentation multi-échelles

Si X est un ensemble et E_λ une énergie multi-échelles sur X, une hiérarchie H sur X qui est

i. **persistante** pour E_λ, et

ii. **de co-énergie totale maximale**

est appelée une **segmentation multi-échelles optimale** de X pour E_λ.

En supposant qu'une telle segmentation est unique, nous la notons $\overline{H}^*(\mathsf{E}_\lambda)$. Le problème proposé consiste alors à trouver :

$$\overline{H}^*(\mathsf{E}_\lambda) \quad \triangleq \quad \arg \max_{H \in \mathbb{H}(X) | H^* = H} \overline{\mathsf{E}}(H). \tag{8.10}$$

Cette formulation du problème est apparue très tardivement dans nos recherches et nous n'avons pas eu le temps de beaucoup approfondir la question. Le problème semble toutefois bien posé et hautement non trivial car il y a une interdépendance entre toutes les échelles : trouver l'optimum global de 8.10 est vraisemblablement bien plus complexe que trouver une partition minimale sur $\mathbb{P}(X)$.

Nous expliquerons en conclusion en quoi introduire une dépendance entre les échelles est susceptible de modéliser de manière unifiée plusieurs phénomènes visuels importants : textures, groupements perceptuels, influences contextuelles. Voyons pour l'instant que les segmentations multi-échelles optimales vérifient à nouveau une propriété d'invariance par homothétie sur les énergies et que la stratégie d'escalade *pure* se présente comme une stratégie d'approximation gloutonne localement optimale pour ce problème de segmentation multi-échelles.

8.4.3 Invariance par homothétie des solutions

Proposition 14 *Si $\mathsf{E}_\lambda = (C, D)$ est une énergie multi-échelles et H une hiérarchie sur X, notons $\overline{\mathsf{E}}(H, \mathsf{C}, \mathsf{D})$ la co-énergie totale de H pour (C, D). Alors*

$$\forall (\mu, \nu) \in \mathbb{R}^{+*^2} \qquad \overline{\mathsf{E}}(H, \mu \mathsf{C}, \nu \mathsf{D}) \quad = \quad \frac{\nu^2}{\mu} \overline{\mathsf{E}}(H, \mathsf{C}, \mathsf{D}).$$

Preuve : Notons $(\lambda_0^+ \leq \cdots \leq \lambda_k^+)$ la séquence ordonnée des échelles d'apparition des ensembles de H pour l'énergie (C, D) et P_i la coupe optimale sur $[\lambda_i^+, \lambda_{i+1}^+[$. Évidemment, $P_k = \{X\}$. Comme $\overline{\mathsf{E}}_H^*$ est nulle après λ_k^+, il vient :

$$\begin{aligned}
\overline{\mathsf{E}}(H, \mathsf{C}, \mathsf{D}) &= \int_0^{+\infty} \overline{\mathsf{E}}_H^*(\lambda) d\lambda \\
&= \sum_{i=0}^{k-1} \int_{\lambda_i^+}^{\lambda_{i+1}^+} \overline{\mathsf{E}}_H^*(\lambda) d\lambda \\
&= \sum_{i=0}^{k-1} \int_{\lambda_i^+}^{\lambda_{i+1}^+} \lambda \mathsf{C}(X) + \mathsf{D}(X) - \lambda \mathsf{C}(P_i) - \mathsf{D}(P_i) d\lambda \\
&= \sum_{i=0}^{k-1} \overline{\mathsf{D}}_i \Delta\lambda_i^+ + \frac{1}{2} \overline{\mathsf{C}}_i (\Delta\lambda_i^+)^2
\end{aligned}$$

avec $\quad \overline{\mathsf{C}}_i \triangleq \mathsf{C}(X) - \mathsf{C}(P_i), \quad \overline{\mathsf{D}}_i \triangleq \mathsf{D}(X) - \mathsf{D}(P_i) \quad$ et $\quad \Delta\lambda_i^+ \triangleq \lambda_{i+1}^+ - \lambda_i^+.$

Transformons alors (C, D) en $(\mu\mathsf{C}, \nu\mathsf{D})$. Comme toutes les échelles d'apparition sont multipliées par $\frac{\nu}{\mu}$, on obtient

$$\overline{\mathsf{E}}(H, \mu\mathsf{C}, \nu\mathsf{D}) \quad = \quad \sum_{i=0}^{k-1} \nu\overline{\mathsf{D}}_i \frac{\nu}{\mu}\Delta\lambda_i^+ + \frac{1}{2}\mu\overline{\mathsf{C}}_i \left(\frac{\nu}{\mu}\Delta\lambda_i^+\right)^2 \quad = \quad \frac{\nu^2}{\mu}\overline{\mathsf{E}}(H, \mathsf{C}, \mathsf{D}).$$

■

En conséquence, les segmentations multi-échelles optimales d'une image sont invariantes par homothétie sur les énergies. Si l'on transforme (C, D) en $(\mu\mathsf{C}, \nu\mathsf{D})$, l'axe des échelles de l'analyse ensemble-échelles finale est à nouveau dilaté d'un facteur $\frac{\nu}{\mu}$.

8.4.4 L'escalade pure : une stratégie d'approximation localement optimale

Si, pour trouver une segmentation multi-échelles optimale, on s'en remet à une heuristique d'optimisation gloutonne, qui complète itérativement une pré-hiérarchie H_k sans remettre en question l'acquis, alors l'escalade pure se présente comme une stratégie localement optimale :

Soit H_k une pré-hiérarchie sur X et E_λ une énergie multi-échelles.

Supposons que

$$\forall x \in H_k \quad \forall y \in H_k^+ \qquad x \subset y \Rightarrow \lambda^+(y) > \lambda^+(x).$$

Autrement dit, supposons qu'aucun choix de complétion de H_k ne permet de rendre non persistant un de ses éléments : l'algorithme qui a construit H_k n'a pas fait de choix inutile, qui serait non persistant.

Notons x^* l'élément de H^+ d'échelle d'apparition minimale (supposons pour simplifier qu'il est unique) et définissons

$$\begin{aligned} \mathbb{H}_x &\triangleq \{G \in \mathbb{H}(X) \mid H_k \subset G \quad \text{et} \quad x^* \in G\}, \\ \mathbb{H}_{\not x} &\triangleq \{G \in \mathbb{H}(X) \mid H_k \subset G \quad \text{et} \quad x^* \notin G\}. \end{aligned}$$

$\mathbb{H}_x$ représente l'ensemble des hiérarchies que l'on peut obtenir par complétion de H_k et qui contiennent x^*, et $\mathbb{H}_{\not x}$ est l'ensemble des hiérarchies qui contiennent H_k mais pas x^*. Comme x^* est l'élément choisi par escalade à partir de H_k, $\mathbb{H}_x$ représente l'ensemble des solutions potentiellement accessibles après un pas d'escalade depuis H_k et $\mathbb{H}_{\not x}$ est son complémentaire.

Notons y^* l'élément de H^+ de deuxième plus petite échelle d'apparition après x^*.

Avec ces hypothèses et notations, on obtient alors

Proposition 15

$$\forall H \in \mathbb{H}_x \quad \forall G \in \mathbb{H}_{\not x} \qquad \int_0^{\lambda^+(y^*)} \overline{\mathsf{E}}_H^*(\lambda)d\lambda \quad \geq \quad \int_0^{\lambda^+(y^*)} \overline{\mathsf{E}}_G^*(\lambda)d\lambda.$$

En outre, si C est strictement sous-additive, l'inégalité est stricte.

Preuve :

On remarquera avant tout qu'au cours d'une escalade pure, l'ensemble des complétions croissantes possibles H^+ diminue, mais ne peut pas s'enrichir. L'ajout à H_k d'un ensemble x de H_k^+ pour donner H_{k+1} conduit à un ensemble H_{k+1}^+ qui consiste en H_k^+ privé de tous les ensemble inclus dans x ou intersectant proprement x.

En second lieu, l'ajout à H_k d'un élément x de H_k^+ provoque nécessairement une *augmentation* des échelles d'apparition des éléments subsistant dans H_{k+1}^+. En effet, d'une part les ensembles disjoints de x ne sont pas affectés par l'ajout de x à la pré-hiérarchie. D'autre part, C étant sous-additive, tout sur-ensemble y de x qui apparaissait en $\lambda^+(y) > \lambda^+(x)$ voit en $\lambda^+(y)$ son énergie propre comparée à une

énergie partielle moindre qu'à l'étape précédente, et son apparition repoussée. Enfin, les sur-ensembles de x qui apparaissaient avant x ont une échelle d'apparition inchangée.

Si H est une pré-hiérarchie de composantes hiérarchiques C_i, la fonction d'énergie E^*_H de ses λ-coupes est la somme des fonctions d'énergie $\mathsf{E}^*_{C_i}$ des C_i. On peut donc doter H_k d'une co-énergie, exactement comme s'il s'agissait d'une hiérarchie, comportant X.

Si l'on ajoute à H_k un ensemble x de H_k^+ alors la fonction $\mathsf{E}^*_{H_{k+1}}$ d'énergie des coupes de H_{k+1} s'écrit en fonction de $\mathsf{E}^*_{H_k}$

$$\mathsf{E}^*_{H_{k+1}} = \inf\{\mathsf{E}^*_{H_k}, \mathsf{E}_x + \sum_{\substack{y \in \vee H_{k+1} \\ y \neq x}} \mathsf{E}_y\} \tag{8.11}$$

où $\vee H_{k+1}$ désigne l'ensemble des sommets des composantes hiérarchiques de H_{k+1}. Le deuxième terme de l'inf correspond à l'énergie de la coupe la plus haute de H_{k+1}, qui comporte tous les sommets des composantes hiérarchiques de H_{k+1}, dont x (la seule différence entre cette relation et la relation de programmation dynamique fonctionnelle déja vue est que l'on a sommé les énergies de toutes les composantes hiérarchiques). Notons F_x ce deuxième terme de l'inf, dépendant de x. F_x est une fonction affine, somme d'énergies propres.

Soit x^* l'élément de H_k^+ de plus faible échelle d'apparition, et y^* celui de deuxième plus faible échelle d'apparition.

Décomposons alors la co-énergie totale d'une (pré-)hiérarchie en trois parties (cf. figure 8.9) :

$$\overline{\mathsf{E}}(H) = A(H) + B(H) + C(H) \tag{8.12}$$

$$= \int_0^{\lambda^+(x^*)} \overline{\mathsf{E}}^*_H(\lambda)d\lambda + \int_{\lambda^+(x^*)}^{\lambda^+(y^*)} \overline{\mathsf{E}}^*_H(\lambda)d\lambda + \int_{\lambda^+(y^*)}^{+\infty} \overline{\mathsf{E}}^*_H(\lambda)d\lambda. \tag{8.13}$$

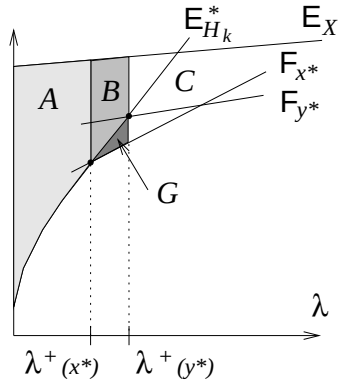


FIG. 8.9 – Gain local net de co-énergie par escalade.

D'après 8.11, avant $\lambda^+(x^*)$, $\mathsf{E}^*_{H_k} = \mathsf{E}^*_{H_{k+1}}$. Ainsi, quelle que soit la complétion de H_k choisie, $A(H_{k+1}) = A(H_k)$. En outre, comme H_k^+ s'apauvrit systématiquement et que les apparitions ne peuvent qu'augmenter par complétion, quelle que soit la hiérarchie H obtenue par un nombre arbitraire de complétions à partir de H_k, elle vérifie $A(H) = A(H_k)$.

Un raisonnement similaire s'applique au terme $B(H_{k+1})$:

Pour une complétion par x^*, $B(H_{k+1}) = B(H_k) + G$, où G est un terme positif (cf. figure 8.9). Si C est srictement sous-additive alors G est strictement positif. Par contre, pour tout autre choix que x^*, $B(H_{k+1}) = B(H_k)$, car les apparitions des autres ensembles sont toutes supérieures à celle de y^*. En outre, pour un choix différent de x^*, comme les apparitions ne peuvent que croître, aucune complétion ultérieure autre que par x^* ne peut augmenter la part B de la co-énergie totale de la hiérarchie finale. Remarquons finalement que si après une complétion par un ensemble y, x^* est toujours présent dans les

complétions possibles, alors x^* est nécessairement soit disjoint, soit supérieur à y. Si x^* et y sont disjoints alors les complétions par x^* et y commutent. Si $y \subset x^*$, alors $\lambda^+(x^*)$ est inchangée après complétion par y et compléter ultérieurement par x^* rend y non persistant dans la hiérarchie finale, ce qui revient à avoir directement complété par x^* en shuntant l'étape y. ■

On retrouve l'idée d'augmentation progressive de λ : l'ensemble ajouté par escalade pure est celui qui maximise la co-énergie cumulée jusqu'à une borne qui augmente petit à petit. Le point important est que le gain local de co-énergie est un gain *net* dans la solution finale : tout autre choix conduit à des solutions dont la co-énergie partielle, cumulée jusqu'à $\lambda^+(y^*)$, est *strictement* plus faible (si C est strictement sous-additive). Évidemment, cette stratégie ne conduit pas à l'optimum global car chaque choix local est irréversible et élimine de nombreux chemins possibles, qui auraient pu se révéler globalement meilleurs (en gagnant sur la part C de la co-énergie dans l'expression 8.12).

On retrouve également que la stratégie d'escalade pure correspond à un choix optimal du point de vue des coupes locales. À partir d'une solution courante H_k, le choix de l'ensemble x^* de plus faible apparition conduit aux coupes d'énergie minimale sur l'intervalle $[\lambda^+(x^*), \lambda^+(y^*)[$. L'escalade pure emprunte donc le chemin qui conduit aux meilleures solutions à court terme (idée de continuation).

8.5 Persistance des régions

Avant de conclure ce chapitre, interessons-nous à la persistance des régions déterminées par escalade.

L'intervalle de persistance $\Lambda^*(x)$ d'une région x représente la gamme d'échelles sur laquelle la région «vit» dans une description de l'image. La taille de cet intervalle traduit donc une notion de *stabilité* des régions dans une solution optimale pour E_λ. Plus une solution s'appuie sur des régions persistantes, moins une faible perturbation du paramètre d'échelle a de chance de la modifier. Nous renvoyons à Leclerc (1989a) qui discute longuement l'importance de l'idée de stabilité en vision.

On pourrait envisager de quantifier la stabilité d'une région par la longueur de $\Lambda^*(x)$, i.e. par $\lambda^-(x) - \lambda^+(x)$. Cette expression n'est cependant pas invariante par homothétie : elle est multipliée par ν/μ si l'énergie (C, D) est transformée en $(\mu\mathsf{C}, \nu\mathsf{D})$. *Si l'on restreint l'axe des échelles aux valeurs positives*, il semble donc préférable de considérer un ratio, qui est invariant par homothétie. En conséquence,

Définition 36 *Pour toute région x d'une hiérarchie persistante, le ratio*

$$\pi(x) \quad \triangleq \quad \frac{\lambda^-}{\lambda^+}(x) \tag{8.14}$$

est appelé la ***persistance*** *de x.*

Nous retrouvons donc l'intéret - tout au moins théorique - qu'il y a à considérer des énergies multi-échelles pures, qui conduisent à des solutions appartenant au demi-axe positif des échelles[2].

Remarque : nous avons vu en 7.2.1 que les formulations comme minimisation de $f(\lambda)\mathsf{C} + \mathsf{D}$ conduisaient également à des solutions causales en λ pour toute fonction f non décroissante. Posons $\mu = f(\lambda)$. Il est alors évident que les hiérarchies d'escalade associées à la minimisation de $\mu\mathsf{C} + \mathsf{D}$ et de $f(\lambda)\mathsf{C} + \mathsf{D}$ sont structurellement identiques. Seule l'échelle finale de la représentation change, avec $\lambda^+ = f^{-1}(\mu^+)$. L'échelle finale peut donc être arbitrairement étirée par une fonction monotone. Encore une fois, appliquer le changement d'échelle a priori ou a posteriori est équivalent. *Ceci ne signifie pas que les intervalles de persistance n'ont pas de signification.* Il serait en effet absurde de considérer une modulation de λ par

[2] Relevons à nouveau le problème de «comparaison avec le vide» : pour une définition additive de la persistance $(\lambda^- - \lambda^+)$, les régions de base apparaissent en $-\infty$ et pour une définition multiplicative (λ^-/λ^+), elle ont une apparition nulle. Dans les deux cas, leur persistance est infinie. Par ailleurs, le sommet de la hiérarchie a systématiquement une disparition infinie qui donne dans les deux cas une persistance infinie.

une fonction f arbitrairement contractante en certains points de l'axe des échelles et dilatante à d'autres (i.e. de dérivée non monotone). Par contre, on voit que l'on peut passer d'une formulation multiplicative à une formulation additive de la persistance en considérant $f = \log$. Une persistance

$$\pi_\mu(x) = \frac{\mu^-}{\mu^+}(x)$$

associée à l'énergie $\mu\mathsf{C} + \mathsf{D}$ correspond en effet à une persistance

$$\pi_\lambda(x) = \frac{\lambda^-}{\lambda^+}(x) = \frac{e^{\mu^-}}{e^{\mu^+}}(x) = e^{\mu^- - \mu^+}(x)$$

pour l'énergie $\log(\lambda)\mathsf{C} + \mathsf{D}$.

8.5.1 Interprétation : cas de la fonctionnelle de Mumford-Shah

Développons l'expression de la persistance d'une région dans le cas de l'énergie constante par morceaux de Mumford-Shah.

Considérons un nœud o d'une hiérarchie persistante, que nous appellerons l'objet d'étude. Soit g son père - l'objet plus global dans lequel l'objet d'étude s'incrit. Soit E l'ensemble des fils de g - les détails internes de g - dont o lui-même. Nous dirons aussi que E représente l'environnement de o, au sens large. Soit I l'ensemble des fils de o - ses propres détails internes (voir la figure 8.10).

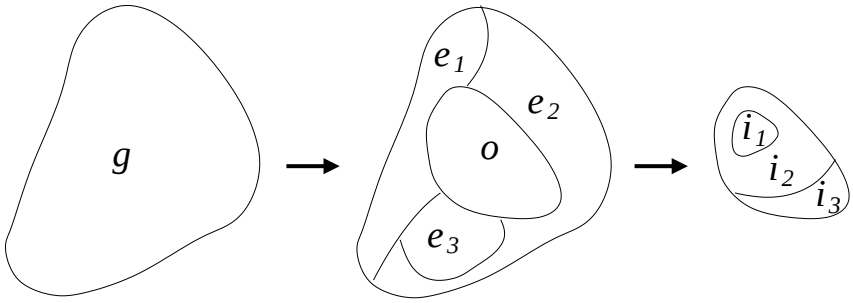


FIG. 8.10 – Un objet o s'inscrit dans un objet plus global g, possède un environnement o, e_1, e_2, e_3 et des détails internes i_1, i_2, i_3.

Considérons une image en noir et blanc. L'attache aux données de Mumford-Shah s'écrit alors pour une région x : $\mathsf{D}(x) = |x|\sigma_x^2$, où σ_x^2 représente la variance de l'image sur la région x. De son côté, le terme de régularisation est $\mathsf{C}(x) = |\delta x|$.

Développons alors l'échelle d'apparition de g :

$$\begin{aligned}\lambda^+(g) &= -\frac{\mathsf{D}(g) - \sum_{e\in E}\mathsf{D}(e)}{\mathsf{C}(g) - \sum_{e\in E}\mathsf{C}(x)} = -\frac{|g|\left(\sigma_g^2 - \frac{1}{|g|}\sum_{e\in E}|e|\sigma_e^2\right)}{|\delta g| - \sum_{e\in E}|\delta e|} \\ &= |g|\frac{\left(\sigma_g^2 - \frac{1}{|g|}\sum_{e\in E}|e|\sigma_e^2\right)}{2|\bigcup_{e\in E}\delta e\backslash\delta g|}. \end{aligned} \tag{8.15}$$

$\lambda^+(g)$ est proportionnelle à la taille de g.

Par ailleurs, le numérateur de 8.15 compare la variance du tout à la moyenne des variances des détails, moyenne pondérée par leurs tailles. La variance est une mesure d'hétérogénéité. Le numérateur mesure donc l'augmentation moyenne d'hétérogénéité quand on passe de la description détaillée à la description en un seul objet. Il est important de voir que ce numérateur n'est pas un terme d'hétérogénéité mais plutôt un terme de *non stationnarité* de g (du point du vue du morcellement considéré et des statistiques

d'ordre deux). En effet, une région g de très forte variance mais sans organisation interne, i.e. telle que les variances des sous-parties soient les mêmes que celle du tout, a une échelle d'apparition nulle car le numérateur de 8.15 est nul.

De son côté, le dénominateur de $\lambda^+(g)$ mesure la complexité de la géométrie des structures internes à g, ici à leur longueur (cf. figure 8.11(a)).

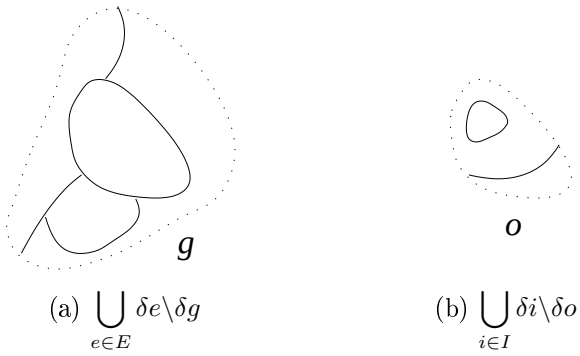


FIG. 8.11 – Sur et sous-structures d'un objet.

La persistance de l'objet o s'écrit alors

$$\pi(o) \quad = \quad \frac{\lambda^+(g)}{\lambda^+(o)} \quad = \quad \frac{|g|}{|o|} \times \frac{\sigma_g^2 - \frac{1}{|g|}\sum_{e\in E}|e|\sigma_e^2}{\sigma_o^2 - \frac{1}{|o|}\sum_{i\in I}|i|\sigma_i^2} \times \frac{|\cup_{i\in I}\delta i\backslash\delta o|}{|\cup_{e\in E}\delta e\backslash\delta g|}. \tag{8.16}$$

Bertolino (1995, pp.109-110) indique que «la capacité de notre système visuel à distinguer un objet homogène en niveaux de gris dans une scène statique observée à une distance variable dépend essentiellement de quatre facteurs classés par ordre d'importance décroissante : la taille de l'objet par rapport au champ visuel, son contraste par rapport à son environnement visuel, son homogénéité propre et sa forme».

L'ensemble de ces facteurs se retrouve dans l'expression 8.16.

Le «champ visuel» est pour nous relatif à l'objet d'étude et correspond à la structure plus globale g dans lequel l'objet o s'inscrit. Le terme $|g|/|o|$ dans 8.16 mesure le rapport de taille entre le «champ visuel» et l'objet. Le deuxième terme est le ratio entre la non stationnarité du champ visuel et la non stationnarité de l'objet. Un objet est d'autant plus persistant qu'il se distingue de son environnement et que ses détails internes se distinguent mal (on retrouve une idée de contraste proche de celle employée dans le modèle de cocons). Le troisième terme de 8.16 compare la complexité des sous-structures de l'objet à celle de ses super-structures. Pour le modèle de Mumford-Shah, un objet est d'autant plus pertinent que ses limites internes sont longues, ce qui augmente avec le nombre de détails internes et la sinuosité de leurs frontières, et d'autant plus pertinent que les limites des objets de son environnement - dont les siennes propres - sont courtes.

Paradoxalement, le terme $|g|/|o|$ indique que plus l'objet est petit dans le champ visuel, plus il est significatif. En y réfléchissant, c'est toutefois naturel. N'oublions pas que dans notre cas, le «champ visuel» g est relatif à l'objet o, il s'agit de la première sur-structure rencontrée au dessus de l'objet.

Pensons à un ciel de nuit dans lequel brille une étoile unique. Cet objet est alors particulièrement remarquable. Si une multitude d'autres étoiles luisent également, deux cas se présentent. Si l'on ne discerne aucune structure intermédiaire entre le ciel entier et les étoiles individuelles, ni constellation, ni voie lactée, alors chaque étoile est - du seul point de vue du terme $|g|/|o|$ - aussi pertinente que si elle brillait seule dans le ciel (du point de vue du ratio des non stationnarités, augmenter le nombre d'étoiles diminue la pertinence de chacune d'elle). Si des structures intermédiaires existent entre l'étoile et le ciel, l'étoile individuelle doit être comparée à la première structure dans laquelle elle s'inscrit et sa pertinence visuelle est - du point de vue du terme $|g|/|o|$ - plus faible que si elle brillait seule dans le ciel.

8.6 Conclusion

Ce chapitre a proposé une stratégie relativement intuitive pour la construction d'une description multi-échelles d'une image s'appuyant sur un principe de minimisation d'énergie. Le processus d'escalade simule à proprement parler une simplification progressive de description, en collectant les ensembles optimaux au fur et à mesure qu'ils apparaissent au cours d'une croissance d'échelle, ou de manière équivalente au cours du renforcement d'une contrainte sur la complexité de description. L'algorithme ne procède cependant pas par augmentation progressive de λ, ce qui conduirait nécessairement à un échantillonnage de l'axe des échelles, mais par calcul exact des échelles auxquelles interviennent des événements d'apparition de régions lors de la remontée de la coupe optimale. Ce calcul est rendu possible par la méthode de programmation dynamique fonctionnelle pour l'optimisation d'une énergie affine. Nous voyons maintenant tout l'intérêt qu'il y a à adopter un point de vue inverse sur le paramètre d'échelle : les quantités pertinentes sont les échelles caractéristiques auxquelles l'optimum *change* et non l'optimum trouvé à une échelle donnée, qui correspondra en probabilité 1 à un milieu de vie des structures.

Notre étude de la stratégie d'escalade pure nous a conduit à proposer une formulation énergétique globale du problème de segmentation multi-échelles - maximiser la somme sur toutes les échelles des co-énergies des coupes optimales d'une hiérarchie. L'escalade se présente alors comme une stratégie gloutonne pour atteindre cet objectif, qui maximise le gain immédiat de co-énergie. Une transposition continue de la stratégie d'escalade a également montré qu'il s'agissait d'une stratégie localement optimale du point de vue débit/distorsion, qui correspond à suivre la ligne de plus grande pente dans l'espace débit/distorsion du problème.

En pratique, la stratégie d'escalade pure est approximée par escalade binaire connexe et conduit à un algorithme *exempt de paramètre* étant donné un couple d'énergies antagonistes. Cet algorithme est par ailleurs relativement efficace (voir l'analyse de complexité de l'annexe C et les temps de calcul donnés au chapitre suivant). Les représentations multi-échelles obtenues vérifient finalement d'importantes propriétés : invariance structurelle par homothétie et covariance d'échelle, identité de solution pour les deux problèmes duaux de débit/distorsion.

Nous sommes également parvenus à une synthèse des comportements multi-échelles pour deux grandes familles de formulations du problème de segmentation : par prédicats d'homogénéité et par minimisation d'une énergie affine séparable. La tableau qui suit résume ces résultats.

Prédicats d'homogénéité			
		Condition de causalité	Escalade
Mesure d'hétérogénéité Prédicat Coupe Formulation sous contrainte	$\mathsf{D} \quad : \quad \mathcal{P}(X) \to \mathbb{R}^+$ Parties : $B^{\varepsilon}_{\mathsf{D}}(R)$ est vrai $\Leftrightarrow$ $\mathsf{D}(R) \leq \varepsilon$ Partitions : $B^{\varepsilon}_{\mathsf{D}}(P) = \bigwedge_{R \in P} B^{\varepsilon}_{\mathsf{D}}(R)$ $\max\{P \in \mathbb{C}ut(H) \mid B^{\varepsilon}_{\mathsf{D}}(P)$ est vrai$\}^a$ $\min_{P \in \mathbb{C}ut(H)} \lvert P \rvert$ avec $\max_{R \in P} \mathsf{D}(R) \leq \varepsilon$	Croissance de D $R \subset S \Rightarrow \mathsf{D}(R) \leq \mathsf{D}(S)$ $\Updownarrow$ Consistance de $B^{\varepsilon}_{\mathsf{D}}$ $R \subset S \Rightarrow (B^{\varepsilon}_{\mathsf{D}}(S) \Rightarrow B^{\varepsilon}_{\mathsf{D}}(R))$	$\min \mathsf{D}$
Minimisation d'une énergie affine séparable $\mathsf{E} = (\mathsf{C}, \mathsf{D})$ $\mathsf{C}(P) = \sum_{R \in P} \mathsf{C}(R) \qquad \mathsf{D}(P) = \sum_{R \in P} \mathsf{D}(R)$			
	Problème sous contrainte	Formulation Lagrangienne	Condition de causalité	Escalade
Complexité bornée	$\begin{cases} \min_{P \in \mathbb{C}ut(H)} & \mathsf{D}(P) \\ \text{avec} & \mathsf{C}(P) \leq \gamma \end{cases}$	$\Leftrightarrow \begin{cases} \min_{P \in \mathbb{C}ut(H)} & \lambda^* \mathsf{C}(P) + \mathsf{D}(P) \\ \text{avec} & \lambda^* = \min\{\lambda \mid \mathsf{C}(P^*_\lambda) \leq \gamma\} \end{cases}$	Sous-additivité de C sur $\mathcal{P}(X)$ $\mathsf{C}(R \cup S) \leq \mathsf{C}(R) + \mathsf{C}(S)$ $\Leftrightarrow$ Décroissance sur $\mathbb{P}(X)$ $P \geq Q \Rightarrow \mathsf{C}(P) \leq \mathsf{C}(Q)$	$\min -\frac{\Delta \mathsf{D}}{\Delta \mathsf{C}}$ $\Updownarrow$
Distorsion bornée	$\begin{cases} \min_{P \in \mathbb{C}ut(H)} & \mathsf{C}(P) \\ \text{avec} & \mathsf{D}(P) \leq \varepsilon \end{cases}$	$\Leftrightarrow \begin{cases} \min_{P \in \mathbb{C}ut(H)} & \mu^* \mathsf{D}(P) + \mathsf{C}(P) \\ \text{avec} & \mu^* = \min\{\mu \mid \mathsf{D}(P^*_\mu) \leq \varepsilon\} \end{cases}$	Sur-additivité de D sur $\mathcal{P}(X)$ $\mathsf{D}(R \cup S) \geq \mathsf{D}(R) + \mathsf{D}(S)$ $\Leftrightarrow$ Croissance sur $\mathbb{P}(X)$ $P \geq Q \Rightarrow \mathsf{D}(P) \geq \mathsf{D}(Q)$	$\max -\frac{\Delta \mathsf{C}}{\Delta \mathsf{D}}$

TAB. 8.1 – Résumé des modèles multi-échelles étudiés.

[a] le max porte sur la relation de finesse entre partitions.

Chapitre 9

Résultats expérimentaux

Ce chapitre met finalement en application la méthode de segmentation multi-échelles qui vient d'être étudiée et présente des résultats expérimentaux.

La section 9.1 s'intéresse dans un premier temps aux escalades s'appuyant sur le modèle constant par morceaux de Mumford-Shah, dont nous avons beaucoup parlé ci-dessus. Nous analysons en détail le comportement de l'algorithme pour cette énergie : influence de la sur-segmentation initiale, comportement en fonction de l'importance d'un bruit et de changements de résolution.

Dans un deuxième temps, les sections 9.2 et 9.3 envisagent d'autres formes d'énergie.

La section 9.2 étudie différentes énergies de complexité C, en particulier des énergies qui s'appuient sur des approximations polygonales des frontières des régions.

La section 9.3 s'intéresse à l'énergie de fidélité D. Nous voyons dans un premier temps une énergie qui dérive d'une approche par codage minimal selon un modèle gaussien. La forme de cette énergie nous conduit alors à proposer plusieurs variantes. Nous comparons finalement les résultats obtenus pour différentes combinaisons de critères énergétiques.

Pour conclure, la section 9.4 présente succintement un exemple de segmentation de Modèles Numériques d'Élévation (MNE) en facettes planes, c'est-à-dire utilisant un modèle de surface affine par morceaux.

9.1 Modèle constant par morceaux de Mumford-Shah

Soit I une image de $\mathcal{D}$ dans $\mathcal{V} = \mathbb{R}^d$. I peut être l'image d'origine - d est alors le nombre de canaux de I - ou une image de caractéristiques («feature image»), par exemple texturales, calculées sur l'image d'origine - d est dans ce cas le nombre de caractéristiques considérées.

Soit R une région de I comportant N observations $(X_1 \dots X_N)$ de $\mathbb{R}^d$. Nous notons X_i^j la $j^{\text{ième}}$ coordonnée de l'observation X_i, $\overline{X}$ la moyenne empirique des X_i et Σ_X leur matrice de variance/covariance, dont le terme général est

$$\Sigma_X(i,j) = \frac{1}{N}\sum_{k=1}^{N}(X_k^i - \overline{X^i})(X_k^j - \overline{X^j}).$$

Mumford et Shah mesurent la distance entre un modèle constant par région et l'image en norme L_2. Pour une image discrète, cette distance, que nous noterons Q, s'exprime pour une région R :

$$\mathsf{Q}(R) \triangleq \sum_{i=1}^{N}||X_i - \overline{X}||^2 = \sum_{i=1}^{N}\sum_{j=1}^{d}(X_i^j - \overline{X^j})^2 = N\mathrm{Tr}(\Sigma_X) = N\sum_{j=1}^{d}\lambda_i \tag{9.1}$$

où $\mathrm{Tr}(\Sigma_X)$ désigne la trace de Σ_X et les λ_i sont ses valeurs propres.

La trace étant invariante par changement de base orthonormée, dans le cas d'analyse d'images en couleurs, la mesure est invariante par translation et rotation du repère colorimétrique (la même remarque vaut bien étendu pour une approximation constante par morceaux d'une image de caractéristiques).

Koepfler et al. (1994) ont suggéré d'introduire des paramètres de pondération entre les difféents canaux, afin, par exemple, de privilégier l'information de chrominance par rapport celle de luminance (la pondération intervient dans ce cas après rotation du repère de manière à faire coïncider son axe principal avec l'axe d'intensité R+V+B). Nous n'avons pas exploré cette possibilité. Différents essais après transformation des observations RVB dans d'autres systèmes de représentation colorimétrique ($L^*a^*b^*$, $L^*u^*v^*$) n'ont pas permis de mettre en évidence de choix privilégié pour la segmentation. Nous préférons finalement travailler sur les observations RVB brutes.

De son côté, l'énergie de complexité C d'une région est choisie proportionnelle à la longueur de sa frontière. Nous noterons cette énergie $\mathsf{L}(R) \triangleq |\delta R|$.

Comme nous l'avons vu, nous considérons des frontières discrètes interpixellaires. La mesure L n'est donc pas invariante par rotation : les segments discrets diagonaux ont une longueur sur-évaluée d'un facteur $\sqrt{2}$ par rapport à la longueur euclidienne. Dans le cas où la sur-segmentation initiale n'est pas la segmentation en pixels, on peut obtenir une mesure moins biaisée par la discrétisation de l'image en considérant une approximation polygonale des frontières des régions initiales, tel que c'est illustré par la figure 4.6(b), p.57. L'algorithme d'approximation polygonale que nous avons utilisé est décrit en annexe D. On peut le voir comme un algorithme de segmentation 1D s'appuyant sur un prédicat d'homogénéité, l'homogénéité d'un segment étant donnée par la variance des points autour d'une droite de régression. La stratégie de construction de l'approximation est alors une escalade. Le degré d'approximation est réglé par un seuil σ sur la variance. Nous choisissons ici $\sigma = \sqrt{2}$, ce qui est juste suffisant pour supprimer les effets de discrétisation des frontières interpixellaires. L'énergie géométrique d'une région est alors égale à la longueur de son approximation polygonale.

En dehors de l'énergie, l'algorithme d'escalade dépend de la sur-segmentation initiale. Nous avons considéré deux types de sur-segmentations :

1) Les sur-segmentations en dalles régulières $k \times k$ que nous notons R_k. R_1 est donc la sur-segmentation absolue, en pixels.

2) Une sur-segmentation par ligne de partage des eaux sur le module du gradient de l'image. L'algorithme est le même que celui employé dans le cadre des cocons (cf. annexe A p. 279). Son résultat dépend donc du paramètre α du filtre de Deriche. Nous notons W_α les partitions obtenues.

Nous notons finalement les algorithmes correspondants LQR_k et LQW_α où les préfixes L et Q indiquent respectivement une énergie de longueur et des déviations quadratiques. Le suffixe indique le type de sur-segmentation considérée. Nous réemploierons ci-dessous le même type de nomenclature pour désigner d'autres algorithmes s'appuyant sur des énergies différentes.

9.1.1 Représentation des résultats

Dichotomie d'une description ensembles-échelle

Les descriptions obtenues sont volumiques. Nous sommes donc obligés de les explorer par des coupes. Si l'on doit fournir une séquence de coupes, les progressions naturelles sont les progressions géométriques, de la forme $\lambda_0 a^k$, où λ_0 et a sont deux constantes positives.

Comme nous l'avons vu, l'axe des échelles est défini à un facteur près (lié à l'arbitraire des unités utilisées pour chaque énergie). Tout choix *a priori* de l'échelle de référence λ_0 est donc arbitraire.

Pour une image I donnée, on trouve toutefois une échelle de référence naturelle *a posteriori* - après analyse de l'image - qui est l'échelle d'apparition du sommet de la hiérarchie, i.e. la plus petite échelle pour laquelle l'image est décrite comme une seule région. Dans ce chapitre, nous noterons cette échelle $\lambda^+(I)$. La façon la plus naturelle d'explorer un scale-sets par des coupes est finalement de fixer $\lambda_0 = \lambda^+(I)$ et de considérer des suites de raison $0 < a < 1$. Nous privilégierons les suites de raison $1/2$ et nous dirons

que les séquences de coupes obtenues sont des **dichotomies** des scale-sets.

Notons que les dichotomies héritent des propriétés d'invariance des scale-sets d'escalade. Par exemple, si le scale-sets est invariant par multiplication des valeurs de l'image par une constante alors les dichotomies le sont également : l'analyse s'adapte au contraste de l'image.

Ultramétrique et cartes de contours

Une hiérarchie indicée sur un ensemble X est totalement déterminée par la distance ultramétrique δ sur X qui lui est associée. Dans notre cadre, la distance entre deux points x et y est la plus petite échelle à laquelle x et y sont groupées au sein d'une région minimale. Formellement, il s'agit de l'échelle d'apparition de l'infimum de x et y dans la hiérarchie :

$$\delta(x, y) = \lambda^+(x \vee y).$$

Remarquons alors que si les boules de l'espace sont connexes alors l'ultramétrique est totalement spécifiée par la donnée des distances entre points voisins de X. Dans notre cas, les points de l'espace sont les régions de la sur-segmentation initiale (la base de la hiérarchie). L'ultramétrique est donc déterminée par la donnée d'une distance entre couples de régions voisines de la sur-segmentation initiale. On peut alors envisager cette distance comme un attribut des *frontières* de la sur-segmentation, attribut qui représente l'échelle à laquelle elles *disparaissent* (dans une représentation ensembles-échelle, la dualité régions/contours s'exprime par le fait que l'apparition d'une région coïncide avec la disparition des contours qui correspondent à ses limites internes). On peut finalement représenter une segmentation multi-échelles en régions connexes par une **carte de contours**, dans laquelle l'intensité d'un contour correspond à son échelle de disparition.

Soit f une frontière interpixellaire séparant deux pixels 4-voisins x et y. Nous définissons alors l'échelle de disparition de f par

$$\lambda^-(f) \quad \triangleq \quad \lambda^+(B_x \vee B_y)$$

où B_x et B_y sont les régions de base auxquelles x et y appartiennent.

Pour représenter les disparitions des frontières de l'image nous adoptons une approche similaire à l'approche par dichotomie. Les progressions géométriques en λ sont associées à une représentation logarithmique de l'axe des échelles, qui transforme les ratios d'échelles (les persistances) en différences. Nous affectons donc une intensité au contour f proportionnelle à $\log \frac{\lambda^-(f)}{\lambda_0}$, où λ_0 fixe l'origine des intensités. Comme ci-dessus, nous spécifions l'origine λ_0 relativement à l'échelle $\lambda^+(I)$ d'apparition de l'image, par

$$\lambda_0 \quad \triangleq \quad \frac{\lambda^+(I)}{\gamma_0}$$

où γ_0 est un réel supérieur à 1.

Les contours qui disparaissent en dessous de $\lambda^+(I)/\gamma_0$ ont alors une intensité négative et nous choisissons de ne pas les représenter dans la carte. Nous définissons finalement l'intensité normalisée $C(f) \in [0, 1[$ d'une frontière f par :

$$C(f) \quad \triangleq \quad \begin{cases} 0 & \text{si} \quad \lambda^-(f) \leq \lambda_0 \\ \frac{\log \frac{\lambda^-(f)}{\lambda_0}}{\log \frac{\lambda^+(I)}{\lambda_0}} & \text{sinon.} \end{cases} \tag{9.2}$$

γ_0 règle alors le niveau de détail visible dans la carte, c'est pourquoi appelons γ_0 la **précision** de la carte de contours. En développant $C(f)$ pour un contour f visible à la précision γ_0, i.e. tel que $\lambda^-(f) > \lambda^+(I)/\gamma_0$, on obtient finalement

$$C(f) = 1 - \frac{\log \frac{\lambda^+(I)}{\lambda^-(f)}}{\log \gamma_0}. \tag{9.3}$$

Pour visualiser cette carte de contours nous affectons alors un niveau de gris à la frontière f qui vaut

$$G(f) \quad \triangleq \quad 256 \times (1 - C(f)) \quad = \quad 256 \times \frac{\log \frac{\lambda^+(I)}{\lambda^-(f)}}{\log \gamma_0}.$$

Deux points voisins qui appartiennent à une région qui apparaît avant λ_0 sont donc séparés par une frontière blanche et une frontière est d'autant plus sombre qu'elle disparaît haut dans la hiérarchie[1]. Si l'on seuille cette image de contours à un niveau G_0 donné, on obtient les frontières de la coupe du scale-sets à l'échelle $\lambda(G_0) = \frac{\lambda^+(I)}{\gamma_0^{G_0/256}}$. Dans ce qui suit nous considérerons des cartes de contours de précision $\gamma_0 = 256$. Dans ce cas, $\lambda(G_0) = \frac{\lambda^+(I)}{2^{G_0/32}}$. Les 8 seuils aux 32èmes de l'échelle des gris correspondent alors aux 8 premières étapes de dichotomie du scale-sets.

Notons que cette vision «orientée contours» suggère une méthode de stockage très compacte d'une description ensembles-échelles. Il suffit de stocker la partition correspondant à la coupe à $\lambda = 0$ de la hiérarchie persistante et la liste des échelles de disparition des frontières de cette partition, par exemple dans l'ordre où l'on rencontre les frontières au cours d'un balayage vidéo. Une manière de diminuer encore la taille de la description est de construire un arbre de poids minimal (MST) recouvrant le graphe d'adjacence de régions attribué par les échelles de disparition des frontières et de n'indiquer les échelles que pour les arêtes de cet arbre (on retrouve les mêmes ensembles par tronçonnements du MST car l'ultramétrique sous-dominante d'une mesure qui est déja une ultramétrique est la mesure elle-même). On retrouve alors le type de représentation multi-échelles considéré par Fernand Meyer (1999a).

9.1.2 Résultats expérimentaux

La figure 9.1 montre le résultat de l'algorithme LQR_1 sur 6 images différentes. Chaque colonne comporte les 8 premières étapes de dichotomie des scale-sets. Les nombres situés à gauche de la figure indiquent le rang de la dichotomie, i.e. la valeur de k correspondant à l'échelle de coupe $\lambda_k = \frac{\lambda^+(I)}{2^k}$. La figure 9.1 présente les modèles constants par morceaux trouvés par l'algorithme, i.e. les images obtenues en moyennant les radiométries sur chaque région.

Le tableau 9.1 indique pour chaque image ses dimensions et son nombre de pixels - la taille de la base de la hiérarchie initiale - ainsi que l'échelle d'apparition de l'image entiere $\lambda^+(I)$ et le temps de calcul avec un processeur Pentium III cadencé à $1,2$ GHz. Ce temps est évidemment lié à notre implémentation (l'algorithme est optimisé avec $P_{max} = 32$, voir l'annexe C). La dernière colonne donne le temps de calcul «normalisé» par la taille de l'image et rapporté à une image 256×256.

	Nom	Dimensions	Pixels (N)	$\lambda^+(I)$	t (s.)	$2^{16}\frac{t}{N}$ (s.)
(a)	IRM	256×256	65 536	$2,51$	159 s.	159 s.
(b)	clown	302×226	68 252	$3,29$	45 s.	43 s.
(c)	immeuble	387×351	135 837	$10,10$	114 s.	55 s.
(d)	rond point	230×253	58 190	$1,45$	38 s.	42 s.
(e)	champs	285×276	78 660	$1,38$	50 s.	41 s.
(f)	bâtiment	422×358	151 076	$2,97$	123 s.	53 s.

TAB. 9.1 – Algorithme LQR_1 : caractéristiques des images considérées, échelle d'apparition de l'image et temps de calcul.

La figure 9.2 représente les ensembles-échelle correspondant à chaque image. Nous avons ici aussi utilisé une échelle logarithmique avec une précision de 256 (même transformation que celle donnée par les équations 9.2 et 9.3 mais en remplaçant $\lambda^-(f)$ par l'échelle d'apparition $\lambda^+(x)$ d'une région x). Notons

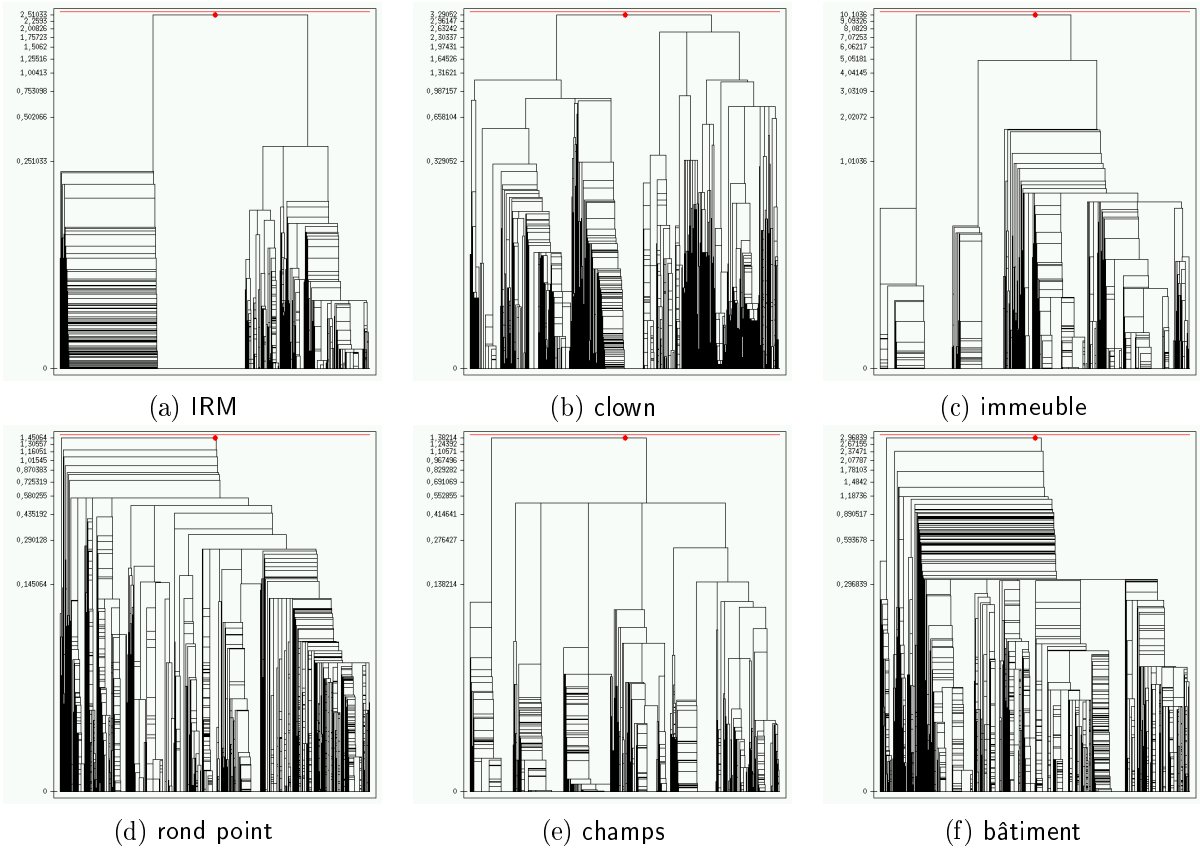
[1] Pour visualiser la carte interpixellaire G nous la transformons en une image pixellaire en affectant à un pixel x la valeur $G(x) = \min\{G(f_s), G(f_g)\}$, où f_s est sa frontière avec son voisin supérieur et f_g avec son voisin gauche.

FIG. 9.1 – Algorithme LQR_1 : dichotomies des ensembles-échelle.

que le 'zéro' indiqué à l'origine de l'axe des échelles correspond en réalité à λ_0 : le bas des hiérarchies est tronqué.

La figure 9.3 donne finalement la carte de contours de précision $\gamma_0 = 256$ correspondant à chaque scale-sets (notons donc que la coupe à λ_8 correspond à la limite de résolution de ces cartes de contours).

La figure 9.1 illustre bien la simplification (ou complexification) progressive de description qui intervient quand on navigue dans des scale-sets. Les fortes valeurs d'échelle mettent en évidence les structures les plus saillantes de l'image, la saillance dépendant à la fois de la taille, du contraste et de la simplicité géométrique de la forme. Les objets saillants peuvent ainsi être de grande taille (e.g. la tête dans l'IRM (a), l'immeuble dans l'image (c)) ou de petite taille mais fortement contrastés (le visage du clown (b), les marquages au sol dans (d), les véhicules dans (f)).



(a) IRM (b) clown (c) immeuble

(d) rond point (e) champs (f) bâtiment

FIG. 9.2 – Algorithme LQR_1 : ensembles-échelle. Échelle logarithmique de précision 256 (voir texte). Les intervalles d'échelle traduisent ici la log-persistance des objets : $\log \lambda^-(x) - \log \lambda^+(x) = \log \frac{\lambda^-}{\lambda^+}(x) = \log \pi(x)$.

L'objet central de l'image (a) IRM, i.e. la tête du patient, est très persistant : il est présent dans 2 coupes successives ($k = 1$ et $k = 2$), i.e. sur deux ordres de grandeur d'échelle (en base 2). Il vit en réalité pratiquement sur 3 ordres de grandeur : sa persistance exacte est $\pi = 7,84 \simeq 2^3$. Son complémentaire - le fond - est encore plus stable - on le trouve pour $k = 1, 2$ et 3 ($\pi = 11,64$) - car il est plus homogène. Cette structure objet/fond se retrouve au niveau du dendrogramme 9.2(a) qui est nettement «bimodal», indiquant qu'à grande échelle - au niveau global - la scène est composée de deux objets clairement distincts (notons la différence d'aspect entre la hiérarchie partielle correspondant au fond - à gauche - et celle correspondant à la tête - à droite. Nous verrons ci-dessous que la première structure est caractéristique d'une image désorganisée, d'un bruit). De même, les trois sous-ensembles de la tête (cerveau, couronne

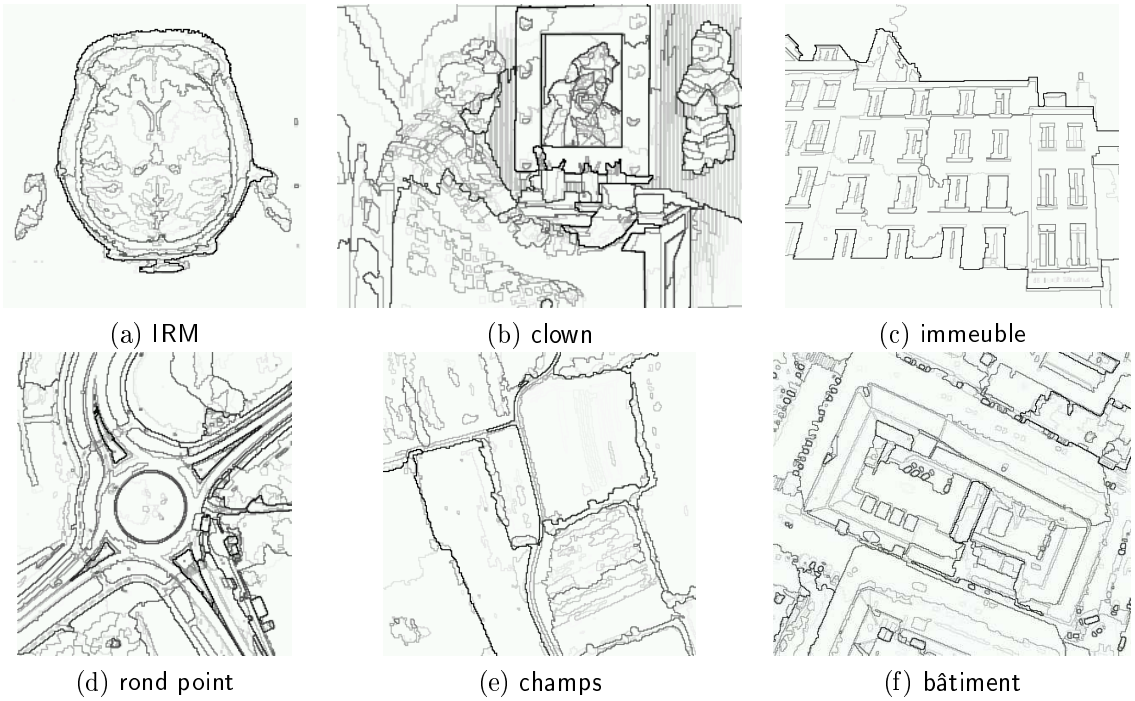


FIG. 9.3 – Algorithme LQR_1 : ultramétriques / cartes de contours. L'axe des échelles est logarithmique avec une «précision» de 256 (voir texte).

grise et os blanc) vivent pour $k = 3$ et $k = 4$ (resp. $\pi = 2,36$; $2,73$ et $2,60$). L'étape suivante distingue le lobe frontal du reste du cerveau, certains détails internes, etc.

L'image (b) clown apparaît dans le scale-sets comme étant constituée de 4 structures principales, que l'on voit sur la coupe à $k = 1$ ($\pi = 2,78$; $6,94$; $2,71$ et $2,12$). Le pallier correspondant à la première division du dendrogramme (en haut à droite) n'est que peu pertinent ($\pi = 1,30$). Se détache ensuite le visage du clown, son reflet dans le miroir, la poupée accrochée au mur et des zones globalement plus sombres que leur environnement. Notons que le reflet et la poupée sont groupés comme des touts par complémentarité : c'est leur fond cohérent (mais non homogène dans le cas de la poupée) qui les isole en se constituant. Notons à ce propos qu'un modèle constant par morceaux et surtout s'appuyant sur une optimisation locale peut difficilement parvenir à grouper correctement la texture complexe de l'habit du clown (il faudrait pour cela qu'il soit sur un fond très distinct). Les étapes suivantes font apparaître la table, les objets qui y sont posés, dont la main du clown, les bandes de la poupée, les parties du reflet, etc.

L'image (c) immeuble comporte deux structures globales nettes : le ciel ($\pi = 20,68$) et l'immeuble ($\pi = 2,04$). L'immeuble se divise en deux éléments forts, ombre ($\pi = 13,38$) et facade ($\pi = 2,92$). Les fenêtres basses s'ajoutent ensuite une à une à la facade, chacune d'elle étant très persistante ($2 \leq \pi \leq 10$) mais leurs complémentaires - les facades ayant partiellement absorbé des fenêtres - non ($\pi \leq 1,4$). L'absorption de ces fenêtres correspond à la chaîne de fusions rapprochées au sommet de la troisième hiérarchie partielle, à droite du dendrogramme 9.2(c)[2].

L'image (f) champs comporte deux objets globaux : un champ vert qui se comporte comme un fond ($\pi = 13,13$) et son complémentaire ($\pi = 2,76$). Ce dernier se décompose à son tour en 5 objets saillants ($\pi = 8,75$; $12,49$; $9,69$; $5,35$ et $2,02$), puis viennent les objets fins : haies, routes, puis inhomogénéïtés des champs, sillons...

[2]Notons que la largeur d'une hiérarchie partielle correspond à la taille de la région sommet. Ainsi des branches très voisines les unes des autres correspondent à de petits objets.

Les images (d) rond point et (f) bâtiment ne présentent pas la même structure globale que les images précédentes. Aucun grand objet persistant n'y est visible, ce qui donne une allure déséquilibrée caractéristique aux dendrogrammes. Les objets les plus saillants sont de petits objets très contrastés. On trouve cependant à plus petite échelle des formes intéressant le cartographe : la structure globale du rond point est donnée à $k = 3$, isolant le rond central, l'ensemble de la route bitumée en un bloc et les séparateurs des pattes d'oie. Les terre-plein apparaissent immédiatement en dessous, etc. Sur le batiment (f), les différents pans de toît sont tous trouvé à une certaine échelle et sont généralement persistants (voir leurs limites sur la figure 9.3(f)).

Nous venons de décrire les scale-sets en sens inverse de leur création. C'est en réalité le jeu des priorités de fusion qui permet d'aboutir aux structures globales, indiquant donc que l'itération de critères locaux peut conduire à des objets globaux, y compris avec un modèle très simple. La clef est l'ordre de fusion, et de ne pas s'arrêter aux objets «homogènes». La saillance visuelle est essentiellement affaire de démarquation par rapport à l'environnement, ce qui est l'idée de base du modèle de cocon et se retrouve ici avec la mesure de persistance.

Au sujet des temps de calcul, notons que rapporté au nombre de pixels de l'image, le temps de calcul sur l'IRM est environ 3 fois plus important que pour les autres images. C'est dû à la fusion du fond noir de l'image, qui se présente comme une zone homogène noyée dans un bruit impulsif. Nous verrons ci-dessous qu'il s'agit du pire des cas du point de vue de la complexité combinatoire de l'algorithme d'escalade. Les images (a) et (f) comportent environ deux fois plus de pixels que les autres. Le temps de calcul normalisé sur ces images est seulement légèrement supérieur aux autres, ce qui indique que le temps typique d'escalade sur l'énergie LQ est pratiquement linéaire en fonction de la taille de l'image.

Pour conclure sur ces premiers exemples, disons deux mots à propos des cartes de contours. Par nature, l'intensité des contours induits par un scale-sets repose sur des caractéristiques globales. Pour ne donner qu'un exemple, la limite supérieure de la tête dans l'IRM est très faiblement contrastée par rapport au fond, à l'inverse de la limite inférieure qui est très marquée. Le contour qui délimite le périmètre cranien a néanmoins dans le scale-sets partout la même intensité, qui correspond à la saillance de la forme globale par rapport au fond. Le contour supérieur est en quelque sorte rehaussé parce qu'il participe à un agencement globalement cohérent. Par contre, le modèle constant par morceaux place des limites - éventuellement franches - au milieu de zones lisses : voir par exemple la segmentation du dégradé de lumière sur le mur derrière le clown.

9.1.3 Étude de l'algorithme

Étudions maintenant d'un peu plus près le comportement de l'algorithme d'escalade pour le modèle constant par morceaux de Mumford-Shah.

Influence de la sur-segmentation

Les figures 9.4 et 9.5 illustrent l'influence de la sur-segmentation initiale sur l'image Lena (256×256).

Nous avons indiqué en regard de chaque coupe le nombre R de régions dans la partition. Ici les contours des régions ont été superposés aux modèles par plages. La figure 9.5 montre les cartes de contours et les hiérarchies correspondantes. Elle indique également le nombre de régions de la sur-segmentation, l'échelle d'apparition de l'image et le temps d'exécution de l'algorithme complet (comprenant donc le temps de sur-segmentation par LPE pour la colonne (c)).

Les résultats sont naturellement assez différents à petite échelle. On voit cependant que les coupes à grande échelle sont relativement similaires. L'algorithme est donc remarquablement stable à terme, quand l'effet de la division initiale ne se fait plus sentir. L'aspect des hiérarchies confirme ce comportement (elles sont représentées avec une échelle linéaire en λ qui met l'accent sur les plus grandes structures).

Notons que l'initialisation selon R_4 correspond à peu près à sous-échantillonner l'image d'un facteur 4. L'étude précise du comportement de l'algorithme en fonction de la résolution sera faite ci-dessous.

FIG. 9.4 – Influence de la sur-segmentation : comparaison entre LQR_1, LQR_4 et LQW_2 (1/2). (a) Sur-segmentation pixellaire (algorithme LQR_1). (b) Sur-segmentation en dalles régulières 4×4 (algorithme LQR_4). (c) Sur-segmentation par LPE, $\alpha = 2$, (algorithme LQW_2).

Notons également que Lena n'est absolument pas constante par morceaux. Il est naturel que le modèle considéré divise les portions lisses en plages (e.g. le chapeau, l'épaule).

Si l'on ne s'intéresse pas aux micro-structures de l'image on a donc tout intérêt à partir d'une LPE, le temps d'exécution est approximativement divisé par la taille moyenne des régions initiales (pour une énergie LQ), soit environ par 10 avec la LPE que nous avons utilisée.

Toutefois, le comportement à petite échelle des algorithmes LQR_1 et LQW_2 est sensiblement différent, comme l'illustre l'expérience de la figure 9.6. sur l'image chromosomes. La table 9.2 associée à cette figure indique différents chiffres concernant les scale-sets correspondants. Notons que les persistances indiquées en table 9.2 font référence aux régions A, B, C et D identifiées sur la première image de la figure 9.6. Les dendrogrammes sont également anotés pour identifier ces objets.

La figure 9.6 montre qu'en partant des pixels individuels, l'algorithme d'escalade forme progressivement des ensembles qui suivent à peu près les courbes de niveau de la fonction (rappelons que Kervrann et al. (1999) ont montré que pour certains régulariseurs simples, les frontières des minima d'une fonctionnelle proche de celle considérée ici correspondent exactement aux isophotes). La LPE rompt ces courbes

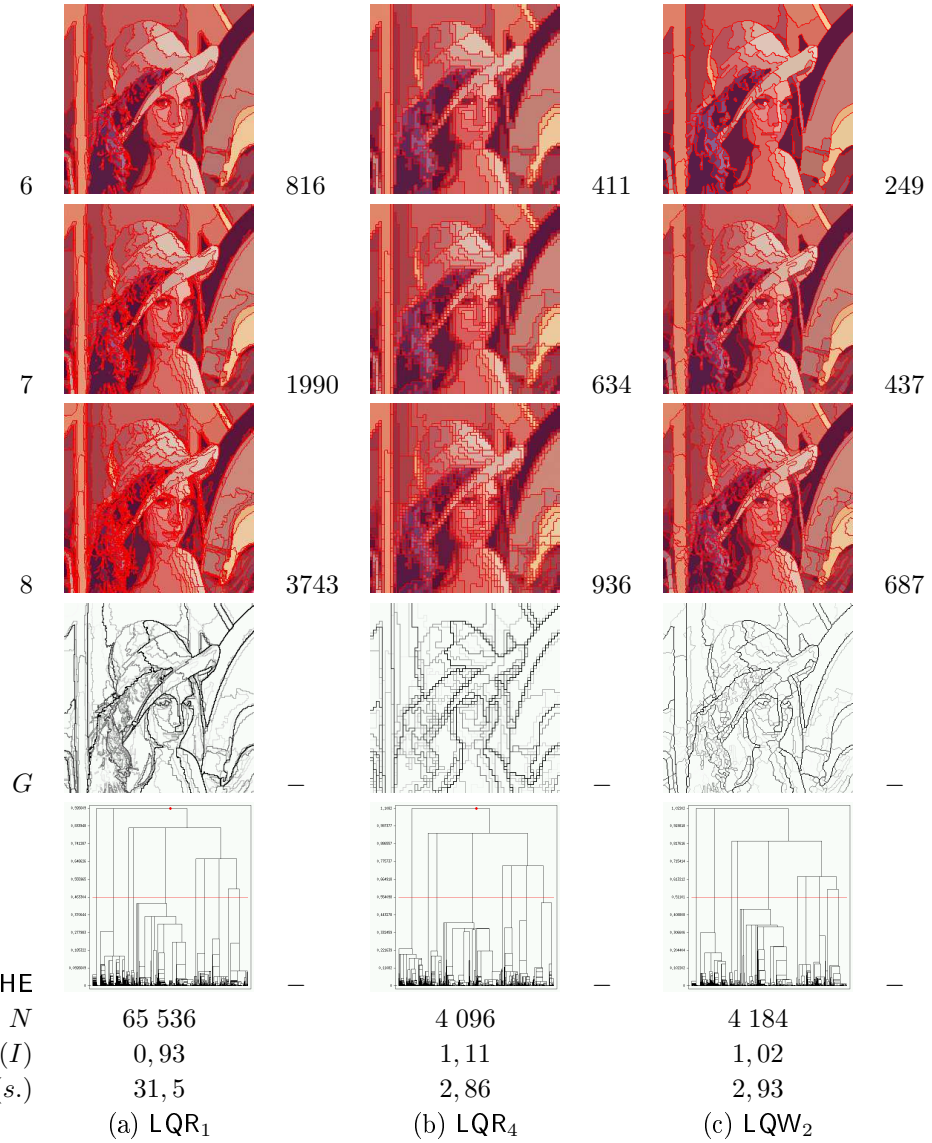


FIG. 9.5 – Influence de la sur-segmentation : comparaison entre LQR_1, LQR_4 et LQW_2 (2/2).

par construction même.

Les trois chromosomes sont bien entendus isolés par les deux méthodes. La table 9.2 indique que les échelles d'apparition de l'image sont très proches pour les deux initialisations. On voit également que les persistances des trois chromosomes sont extrêmement fortes, le critère de persistance identifie sans ambiguïté la présence de trois objets principaux (le petit chromosome A est absorbé par le fond avant les deux autres qui le sont en bloc, mais l'étape est relativement peu persistante, cf. $\pi(A \cup D)$).

Algo.	Base	t (s.)	$\lambda^+(I)$	$\pi(A)$	$\pi(B)$	$\pi(C)$	$\pi(D)$	$\pi(A \cup D)$
LQR_1	17 685	8, 54 s.	0, 294	11, 13	36, 17	9, 53	4, 82	1, 97
LQW_2	709	0, 77 s.	0, 298	36, 13	65, 73	71, 817	7, 04	1, 96

TAB. 9.2 – Comparaison entre LQR_1 et LQW_2 sur l'image chromosomes. Les ensembles A, B, C, D sont indiqués en figure 9.6.

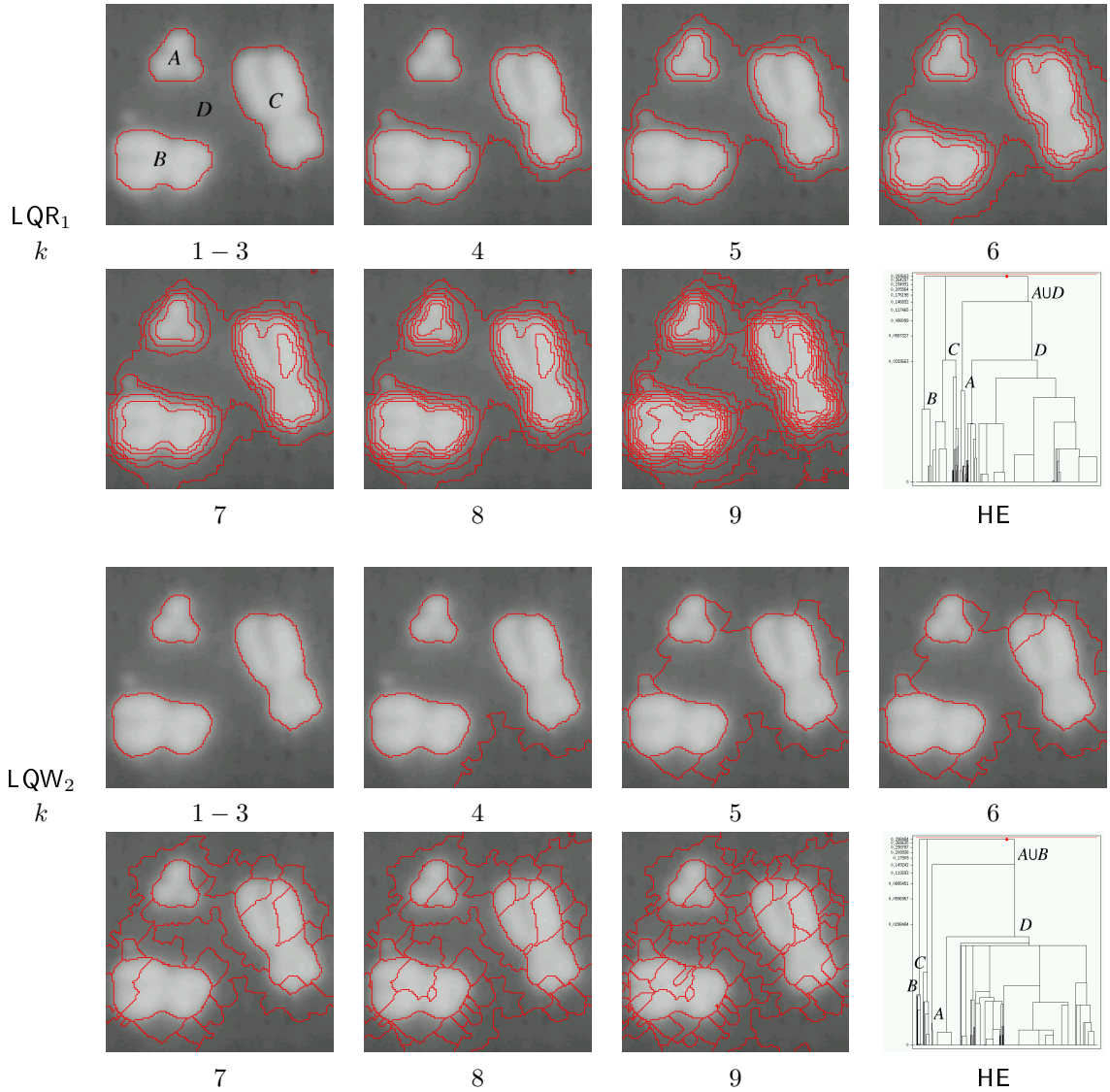


FIG. 9.6 – Comparaison entre LQR$_1$ et LQW$_2$ sur l'image chromosomes.

Escalade et bruit

Voyons maintenant le comportement de l'algorithme d'escalade face à des images bruitées.

Cette étude éclaircira de nombreux aspects de la méthode. Notons que plusieurs points qui sont discutés renvoient aux problèmes d'optimisation de l'algorithme qui sont décrits en annexe C.

Nous avons considéré deux types de bruitage d'une image : un bruitage par «désorganisation» des pixels et un bruitage par ajout d'un bruit gaussien i.i.d.

Images désorganisées À partir d'une image initiale I, nous créons une image «désorganisée» $A_p(I)$ de degré de désorganisation p en réalisant des permutations aléatoires des pixels de I. $A_p(I)$ a ainsi le même histogramme global que I mais une plus faible organisation spatiale. L'indice p dans A_p désigne la probabilité qu'un pixel ait sa valeur échangée avec celle d'un autre, tiré au hasard dans l'image[3].

[3]Concrètement, nous balayons l'image pixel par pixel et si le résultat d'un tirage équiprobable sur $[0, 1]$ est inférieur à p, nous permutons la valeur du pixel avec celle d'un autre pixel de l'image, tiré de manière équiprobable dans le domaine.

Nous étudions alors le comportement de l'algorithme LQR_1 (optimisé avec $P_{max} = 32$, cf. annexe C) sur des images de dégré de désorganisation croissant.

La figure 9.7 montre un extrait 64×64 de l'image girl (initialement 256×256), les images désorganisées A_p générées sur cet extrait et différentes coupes de leurs segmentations hiérarchiques. Les coupes sont ici réalisées selon une séquence fixe d'échelles et non relativement à l'apparition de l'image pour pouvoir comparer les résultats sur différentes images. Les hiérarchies associées sont également présentées.

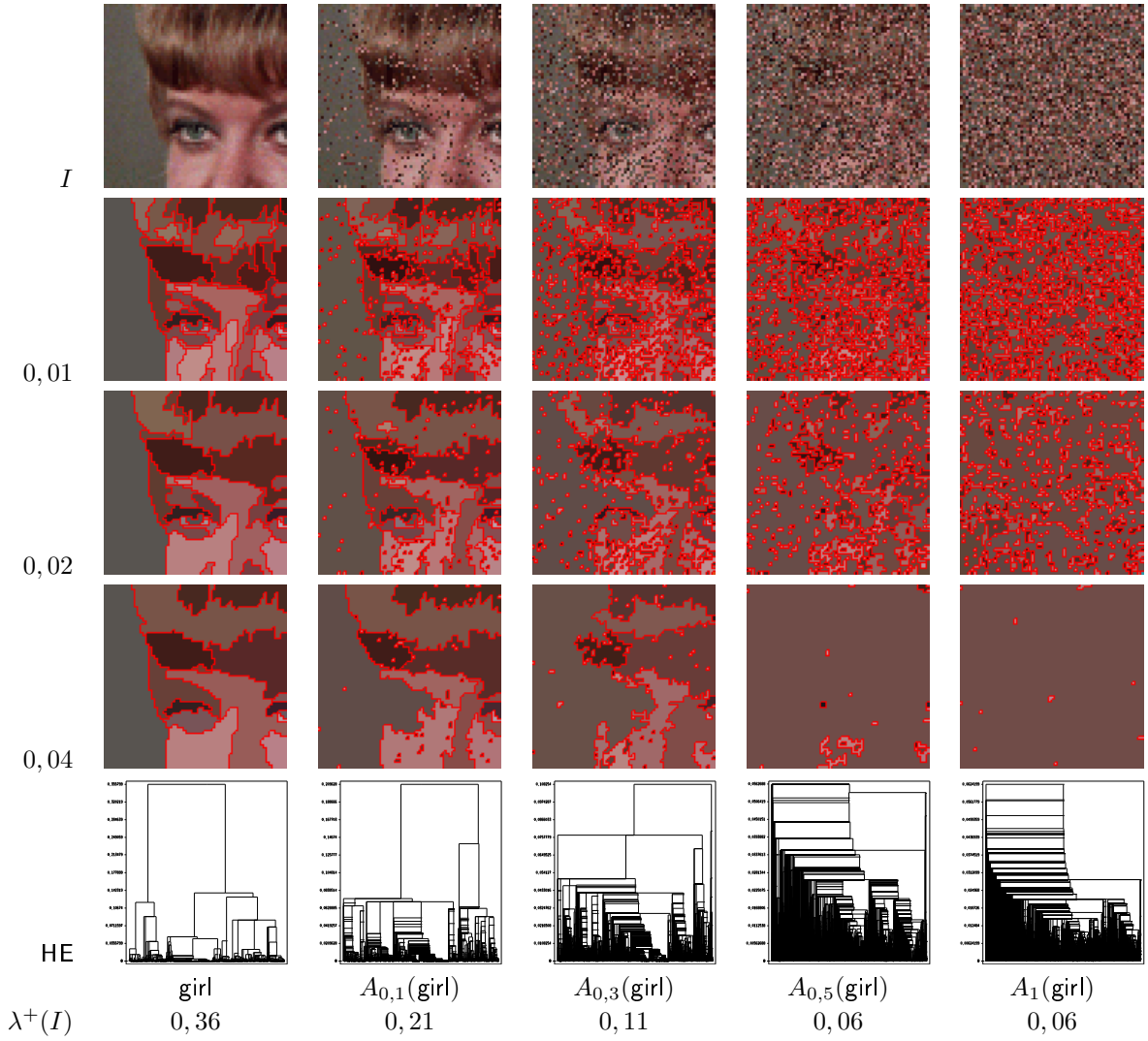


FIG. 9.7 – Comportement de l'algorithme LQR_1 sur une image bruitée par permutations aléatoires. Voir texte.

De son côté, le tableau 9.3 donne quatre quantités en fonction de p et de la taille de l'extrait considéré : le temps de calcul et l'échelle d'apparition du sommet de la hiérarchie ainsi que deux autres chiffres qui font référence à l'algorithme tel qu'il est décrit en annexe C et qui illustrent les problèmes d'optimisation qui y sont discutés. Ces chiffres sont :

- $\frac{F}{N}$: où F est le nombre d'appels à la fonction Calculer_apparition effectués au cours de l'algorithme, divisé par le nombre N de nœuds du graphe de base. F représente le nombre total d'hypothèses de fusion évaluées au cours de l'algorithme. F se décompose en deux termes $F = A + S$ où :

A est le nombre d'arêtes du graphe de base, qui sont toutes balayées au cours de l'initialisation. Sachant

qu'un graphe de N pixels en 4-connexité a environ $2N$ arêtes (aux effets de bord près), $A/N \simeq 2$.

S est la somme des degrés des nœuds résultants d'une fusion (notée $\sum d(z_i)$ dans l'annexe C).

F/N représente donc à peu près le nombre moyen de voisins d'une région venant d'être créée +2. La complexité combinatoire d'une escalade dépend fortement de ce terme.

• PM : Profondeur maximale au cours de l'algorithme de l'intersection d'un morceau affine avec une affine par morceaux f : 1 si l'intersection se situe sur le dernier morceau de f, 2 si elle se situe sur l'avant dernier, etc. Ce facteur intervient dans l'optimisation de l'algorithme.

		32×32	64×64	128×128	256×256
	$N \rightarrow$	1024	4096	16384	65536
Temps ($s.$)	girl	$0,7$	$2,3$	$8,9$	$42,1$
	$A_{0,1}$(girl)	$0,8$	$3,1$	$13,8$	$121,3$
	$A_{0,3}$(girl)	$1,2$	$4,4$	$19,2$	$283,9$
	$A_{0,5}$(girl)	$1,5$	$5,3$	$30,3$	$343,3$
	A_1(girl)	$1,4$	$9,2$	$93,8$	$1866,4$
Degré moyen +2 ($\frac{F}{N}$)	girl	$10,1$	$10,0$	$10,2$	$11,1$
	$A_{0,1}$(girl)	$12,6$	$14,5$	$16,2$	$29,4$
	$A_{0,3}$(girl)	$19,7$	$22,5$	$22,8$	$68,8$
	$A_{0,5}$(girl)	$26,7$	$27,7$	$38,1$	$84,7$
	A_1(girl)	$25,1$	$48,2$	$112,4$	$371,4$
Profondeur max. (PM)	girl	5	5	6	9
	$A_{0,1}$(girl)	5	7	12	25
	$A_{0,3}$(girl)	9	18	15	31
	$A_{0,5}$(girl)	10	12	22	28
	A_1(girl)	8	13	21	30
$\lambda^+(I)$	girl	$0,21$	$0,36$	$0,82$	$1,83$
	$A_{0,1}$(girl)	$0,14$	$0,21$	$0,51$	$1,17$
	$A_{0,3}$(girl)	$0,07$	$0,11$	$0,21$	$0,37$
	$A_{0,5}$(girl)	$0,04$	$0,06$	$0,09$	$0,16$
	A_1(girl)	$0,03$	$0,06$	$0,07$	$0,13$

TAB. 9.3 – Comportement de l'algorithme LQR_1 sur une image bruitée par permutations aléatoires. Voir texte.

Comportement en fonction du degré de désorganisation

À taille d'image fixée (colonnes du tableau 9.3), une augmentation de p se traduit globalement par

i) une augmentation du degré moyen F/N des nœuds,

ii) une augmentation du temps de calcul,

iii) une augmentation de PM et

iv) une diminution de $\lambda^+(I)$.

Les coupes des hiérarchies calculées sur les images désorganisées présentent des régions constellées de trous. Les graphes d'adjacence correspondants ont des degrés très déséquilibrés, les grandes régions trouées ayant des degrés très forts. L'augmentation du temps de calcul est directement liée à l'augmentation de ce degré moyen : l'annexe C montre que - toutes choses égales par ailleurs - la complexité d'une escalade est proportionnelle à F/N. On vérifie ici une assez bonne linéarité entre F/N et le temps.

Le temps de 42 secondes pour l'escalade complète sur l'image girl 256×256 non désorganisée donne l'ordre de grandeur typique d'exécution de l'algorithme à partir des 65 536 pixels d'une image «usuelle» (voir aussi la table 9.1). Par contre, si l'image est un bruit pur (A_1(girl)), il faut environ 30 minutes pour construire la hiérarchie sur la même base !

L'augmentation de PM correspond à une augmentation de la longueur des chaînes d'étapes non

persistantes *successives*. Concrètement, les grandes régions absorbent un à un leurs «bruits» internes mais de nombreuses étapes ne sont pas persistantes (optimales pour l'énergie). Les détails sont finalement absorbés par paquets.

Ceci se retrouve au niveau des hiérarchies. L'évolution de leur aspect en fonction de p est frappant. Leur déséquilibre augmente clairement avec p. La hiérarchie $A_1(\mathsf{girl})$ est caractéristique d'une image de bruit pur : très fort déséquilibre et surtout *aucune grande structure persistante* : on n'observe aucun pallier significatif dans le dendrogramme pour les grandes régions ($\pi(x) \simeq 1$). Les seules éléments persistants sont les petites régions qui sont absorbées par les grandes (et elles sont très persistantes, i.e. $\pi(x) \gg 1$, voire $\pi(x) = +\infty$ s'il s'agit de pixels individuels, de la base de H). Cette structure indique que les deux seuls niveaux pertinents d'analyse de l'image sont le niveau du bruit et celui de l'image entière.

Les échelles d'apparition du sommet de H décroissent avec p. Les dendrogrammes sont ici représentés avec la même hauteur, mais il faut imaginer qu'ils diminuent de taille. Intuitivement, le bruit ne «monte» pas, il «envahit» progressivement le bas de la hiérarchie mais sans augmenter d'échelle. Ce bruit possède d'ailleurs une échelle caractéristique de disparition : l'apparition du sommet de la hiérarchie de $A_1(\mathsf{girl})$. Pour l'extrait 64×64, cette échelle vaut environ 1/6 de l'échelle d'apparition de l'image non bruitée (cf $\lambda^+(I)$ sur la figure 9.7). Le bruit ne «monte» donc pas, ce sont les grandes structures qui «descendent» progressivement, pour finir par disparaître quand elle ont atteint le niveau de bruit, ou quand celui-ci a «envahi» tous le bas de la hiérarchie[4].

Comportement en fonction de la taille de la zone. Examinons maintenant l'effet d'une augmentation de la taille de la zone d'étude (lignes du tableau 9.3). Notons qu'ici, nous n'avons pas sous-échantillonné girl pour obtenir les images de petite taille, mais nous en avons pris des extraits. Les statistiques des sous-images ne sont donc pas les mêmes (le comportement de l'algorithme par rapport à des changements de résolution est étudié ci-dessous en section 9.1.3).

Examinons d'abord l'évolution du temps de calcul. On voit que pour l'image non permutée, ce temps est approximativement proportionnel à la surface de l'image, ce qui n'est absolument pas le cas pour l'image totalement désorganisée $A_1(\mathsf{girl})$. Ce point est directement lié à l'augmentation du degré moyen. Sur l'image non permutée, ce degré est à peu près constant quand la taille de l'image augmente. Il evolue par contre très rapidement sur l'image de bruit pur (quadratiquement en théorie dès que l'image est constituée d'une seule plage - le fond - constellée de «bruits» absorbés un à un).

Enfin, même si l'augmentation de taille ne correspond pas ici à une homothétie, on voit que $\lambda^+(I)$ croît grosso-modo linéairement avec la largeur de l'image.

Bruit gaussien Voyons maintenant le comportement de l'algorithme LQR_1 en fonction d'un bruit gaussien i.i.d. ajouté à l'image.

Notons $\mathcal{N}_\sigma$ une loi normale de moyenne nulle et d'écart-type σ. Considérons à nouveau l'extrait 64×64 de l'image girl de la figure 9.7 et ajoutons-lui un bruit gaussien d'écart-type croissant. La figure 9.8 présente, outre girl elle-même, quatre réalisations d'images bruitées $\mathsf{girl}+\mathcal{N}_\sigma$ pour $\sigma = 10, 20, 30, 40$ (les radiométries ont été tronquées si elles sortent de $\langle 0, 256\langle \ldots)$. Différentes coupes et les hiérarchies calculées sur ces images suivent.

La table 9.4 indique pour sa part les mêmes quantités que précédemment pour ces nouvelles images.

On retrouve différents points qui avaient été observés pour un bruitage par permutation. L'algorithme est toutefois plus résistant à un bruit gaussien - même fort - qu'à une désorganisation[5], ce qui était

[4]Notons que pour les bruitages par «désorganisation», il existe un seuil critique de p qui correspond à la valeur pour laquelle, statistiquement, les pixels non permutés se trouvent divisés en plusieurs composantes connexes. Ces pixels se trouvent subitement déconnectés en une multitude de composantes. Il s'agit d'un problème de seuil de porosité bien connu des physisiens. Ici le problème est légèrement plus complexe car un pixel permuté peut l'être avec un pixel de la même région «cohérente» que lui (e.g. dans girl un pixel gris de fond avec un autre). La permutation n'a alors pas d'effet visuel fort.

[5]Notons que visuellement, $\mathsf{girl}+\mathcal{N}_{40}$ correspond à peu près - s'il y a lieu de comparer - à un niveau de bruit de désorganisation situé entre $A_{0,1}(\mathsf{girl})$ et $A_{0,3}(\mathsf{girl})$. Il faut donc comparer les segmentations de $\mathsf{girl}+\mathcal{N}_{40}$ à celles de $A_{0,1}(\mathsf{girl})$ ou $A_{0,3}(\mathsf{girl})$.

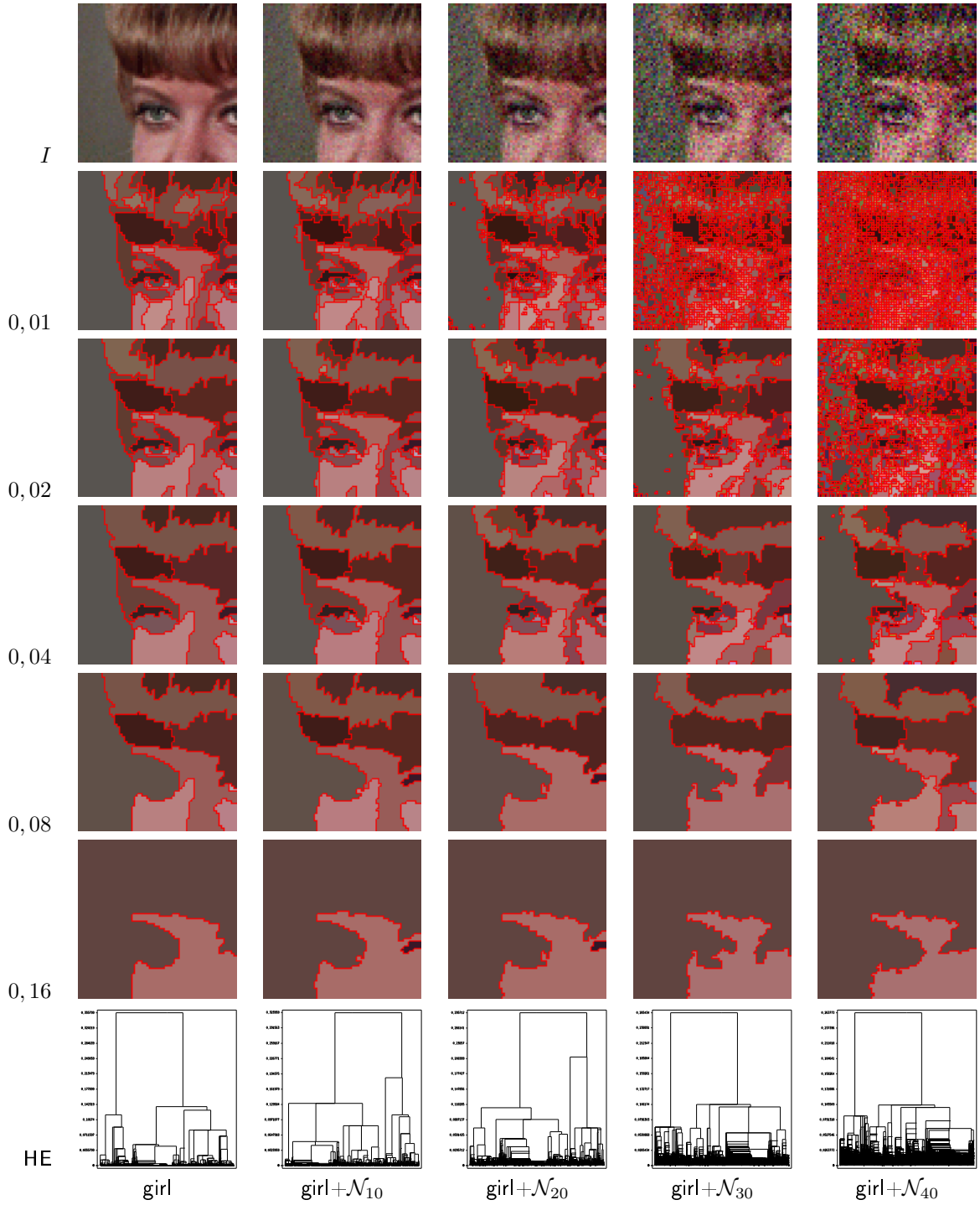


FIG. 9.8 – Comportement de l'algorithme LQR_1 sur une image bruitée par un bruit gaussien i.i.d.. Voir texte.

prévisible : un bruit gaussien est additif, il noie le signal, alors qu'une désorganisation l'occulte en introduisant des points totalement abbérants ; en outre la norme L_2 (énergie Q) est directement reliée à la loi gaussienne.

L'algorithme parvient donc à retrouver les structures principales - à grande échelle - même noyées

	girl	girl+$\mathcal{N}_{10}$	girl+$\mathcal{N}_{20}$	girl+$\mathcal{N}_{30}$	girl+$\mathcal{N}_{40}$
Temps ($s.$)	$2,37$	$2,85$	$2,80$	$3,36$	$3,86$
$\frac{F}{N}$	$10,0$	$12,3$	$13,5$	$15,9$	$17,8$
PM	5	7	7	8	8
$\lambda^+(I)$	$0,36$	$0,32$	$0,30$	$0,27$	$0,26$

TAB. 9.4 – Comportement de l'algorithme LQR_1 sur une image bruitée par un bruit gaussien i.i.d.. Voir texte.

dans un bruit important. Les coupes à λ fort sont très semblables pour les différents σ, dès que l'on passe au dessus de l'échelle propre du bruit, soit environ à partir de $\lambda = 0,04$ pour l'image la plus bruitée (on doit certainement pouvoir dériver théoriquement l'espérance de $\lambda^+(I)$ en fonction de σ et de la taille de l'image pour un image uniquement constituée d'un bruit $\mathcal{N}_\sigma$ i.i.d.).

En conclusion, retenons que l'algorithme LQR_1 est remarquablement robuste au bruit. En outre, l'analyse des scale-sets peut certainement permettre de répondre à une question aussi simple que : *y-a-t-il quelque chose à voir dans l'image ou n'est-elle qu'un bruit pur ?*

Résolution et échelle

Intéressons-nous maintenant au comportement de l'algorithme en fonction de la résolution de l'image.

L'idée est simple : l'algorithme d'escalade s'appuyant sur une énergie LQ est invariant par homothétie géométrique idéale de l'image, où «idéale» signifie que l'homothétie doit conserver les statistiques des régions. Par extension, nous avons cherché à évaluer le comportement de l'algorithme dans le cas d'une augmentation de résolution.

Nous avons tout d'abord vérifié l'invariance par homothétie «idéale». Soit I_0 est une image de taille $n \times m$. Définissons alors l'image I_k comme l'image de taille $k.n \times k.m$ dans laquelle on a répliqué les pixels de I_0 par $I_k(x, y) = I_0(\lfloor x/k \rfloor, \lfloor y/k \rfloor)$. On trouve alors en pratique des scale-sets HE_k associées aux I_k qui ont des structures identiques (à l'œil : nous ne les avons pas appariées...) et dont les échelles d'apparition des sommets sont reliées par $\lambda^+(\vee\mathsf{HE}_k) = k \times \lambda^+(\vee\mathsf{HE}_1)$ (nous avons testé $k = 1, 2, 4$ sur l'image Lena). Notons que la base de la hiérarchie persistante pour l'image I_k est formée de dalles $k \times k$, qui apparaissent toutes à $\lambda = 0$: la méthode «retrouve» donc en quelque sorte la résolution d'origine.

Dans l'expérience de sur-échantillonnage que nous venons de décrire, l'augmentation de définition ne correspond pas à une augmentation de résolution : les nouveaux pixels sont la réplique exacte d'anciens. Nous avons alors procédé à l'expérience inverse : sous-échantillonner progressivement une grande image et comparer les scale-sets obtenus. Dans ce cas, l'augmentation de définition est une augmentation de résolution, qui ajoute de l'information.

Nous avons considéré un extrait 1024×1024 d'une ortho-photographie aérienne à résolution au sol de 50cm (issue de la BDOrtho® IGN). Cette image est représentée en entier en figure 9.9(d). La figure 9.9(c) est un zoom sur sa portion 512×512 centrale, la figure 9.9(b) sur sa portion 256×256 centrale et la figure 9.9(a) sur sa portion 128×128 centrale. La figure 9.9(a) donne donc une bonne idée de la précision des détails de la photographie d'origine à 50cm (un pixel fait $0,23$ mm sur un A4).

Notons I_8 l'image 1024×1024 d'origine. Soit alors I_4 l'image 512×512 obtenue par sous-échantillonnage de I_8 d'un facteur 2, I_2 l'image 256×256 obtenue en sous-échantillonnant I_8 d'un facteur 4, et I_1 l'image 128×128 obtenue en réduisant 8 fois I_8. Les sous-échantillonnages ont été effectué par moyenne sur chaque bloc $k \times k$. Les images I_k représentent donc la même portion de terrain à 4 résolutions différentes : 4m, 2m, 1m et $0,5$m.

Les scale-sets HE_k sur les images I_k ont été calculés par l'algorithme LQW_2.

Comme nous l'avons vu, si I_k était une k-réplication des pixels de I_1 et si l'escalade partait des pixels, alors la coupe de HE_k à l'échelle $k \times \lambda$ serait la k-réplication de celle de HE_1 à l'échelle λ.

Les figures 9.9 et 9.10 présentent donc les coupes obtenues pour une suite géométrique d'échelles

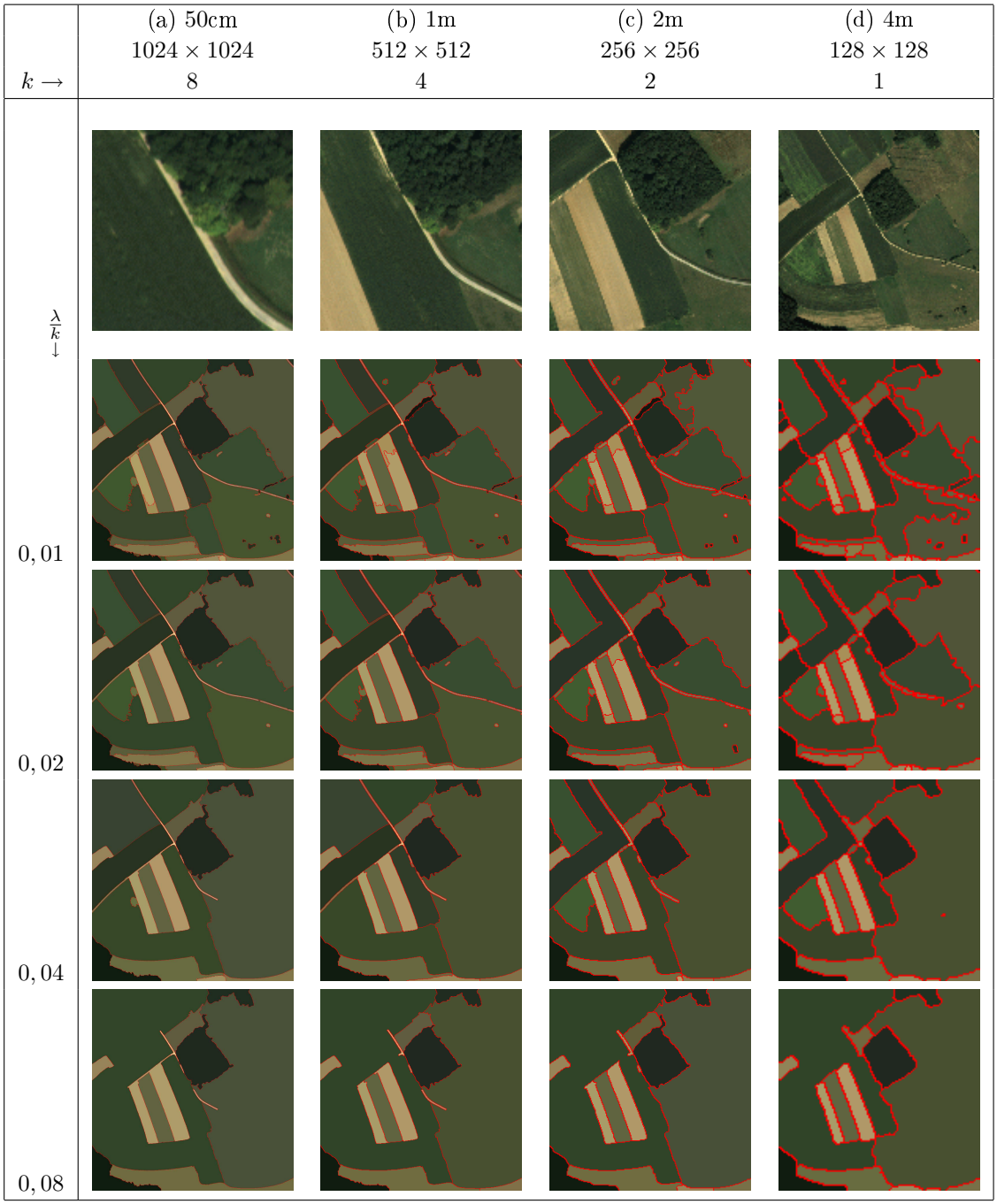


FIG. 9.9 – Changement d'échelle et de résolution (1/2). Voir texte.

$\lambda_i \propto 2^i$, chaque HE_k étant coupé à l'échelle $k \times \lambda_i$. Notons que chaque segmentation est représentée en entier, aux dimensions de I_1. Leur taille croît cependant avec k (on remarque à l'œil que la précision des limites décroît de la séquence (a) à la séquence (d)). La figure 9.10 présente également les scale-sets et indique les valeurs normalisées $\lambda^+(I)/k$, N/k^2 et t/k^2, où N est le nombre de régions de la sur-segmentation par LPE et t est le temps de calcul. Les tailles et temps normalisés illustrent à nouveau la quasi-linéarité du temps de calcul en fonction de la taille pour des images de contenu équivalent.

On constate que les segmentations obtenues à échelles normalisées égales sont très proches, à part

	(a) 50cm	(b) 1m	(c) 2m	(d) 4m
	1024×1024	512×512	256×256	128×128
$k \rightarrow$	8	4	2	1
$\frac{\lambda}{k} \downarrow$				
$0,16$				
$0,32$				
HE_k				
$\lambda^+(I)/k$	$0,567$	$0,627$	$0,616$	$0,644$
N/k^2	1150	1253	1240	1222
t/k^2	$0,97$ s.	$1,02$ s.	$0,96$ s.	$0,91$ s.

FIG. 9.10 – Changement d'échelle et de résolution (2/2). Voir texte.

évidemment pour les objets fins (e.g. la route) qui sont perdus à basse résolution. L'invariance par homothétie des scale-sets induit donc un comportement extrêmement robuste aux changements de résolution. En outre, la relation entre résolution et échelle est linéaire, ce qui signifie que si l'on sait comment régler l'outil pour une résolution donnée, on obtient directement le réglage pour d'autres.

Une idée importante en télédétection est que le succès d'un problème de photo-interprétation automatique dépend de manière critique de la résolution d'observation choisie. On admet alors généralement qu'il existe une résolution «optimale», ou un intervalle étroit de bonnes résolutions, dont la borne inférieure dépend essentiellement de la précision géométrique requise et la borne supérieure de la capacité à identifier et délimiter les objets recherchés comme des touts (Woodcock et Strahler 1987). L'escalade sur

l'énergie LQW_2 fournit à grande échelle des analyses globales qui sont localisées à la précision de l'image traitée. Comme les régions déterminées à un niveau global ne dépendent que faiblement de la résolution d'origine, la méthode permet donc de réduire la question du choix d'une résolution d'observation à un choix de précision géométrique. La démarche à suivre est immédiate : il faut choisir la résolution qui correspond à la précision géométrique requise par l'application et régler l'échelle d'analyse relativement à ce choix. Il n'y a nul besoin de sous-échantillonner l'image pour rendre les objets d'intérêt homogènes, l'escalade réalise en quelque sorte ce sous-échantillonnage virtuellement, tout en conservant la précision de l'image d'origine.

Dans le cadre du scale-space gaussien, transposant l'idée initiale de Witkin (1983) aux régions, plusieurs auteurs tel que Lindeberg (1994) et Olsen (1996, 1997) ont proposé de détecter les objets à grande échelle - par exemple lorqu'ils ne forment qu'un unique bassin versant - puis de remonter le scale-space pour les localiser précisément à faible échelle. L'analyse scale-sets réalise naturellement ce type d'opération en proposant une analyse en objets de plus en plus globaux mais qui restent localisés à la précision initiale.

La robustesse de la méthode à des changements de résolution la pose également comme une solution de choix pour des questions de fusion de données issues de capteurs de résolutions différentes (Laporterie 2002), ou d'appariement d'images de différentes résolutions (Dufournaud 2001).

9.2 Énergies géométriques «orientées régions»

Envisageons maintenant d'autres formes d'énergie que celle proposée par Mumford et Shah, en commençant par les énergies géométriques.

9.2.1 Importance de la régularisation géométrique

Montrons tout d'abord sur un exemple l'importance de la régularisation géométrique. Considérons pour cela l'énergie géométrique la plus simple possible, qui évalue la complexité d'une partition à son cardinal. Notons K l'énergie associée, qui attribue un coût fixe à chaque région.

(a) Image girl 256×256 (b) LQW_2 (c) KQW_2

FIG. 9.11 – Importance de la régularisation géométrique. Les segmentations correspondent aux deux coupes à 15 régions des scale-sets LQW_2 et KQW_2 sur l'image girl.

On constate sur la figure 9.11(b) que les frontières obtenues par l'algorithme KQW_2 sont très sinueuses, contrairement à celle obtenues avec une régularisation par l'énergie de longueur L.

A posteriori, les moins bons résultats expérimentaux obtenus par la méthode des cocons nous semblent principalement dûs au fait que ce modèle n'intègre pas de facteur géométrique dans la qualification des régions (tout comme les approches par prédicats d'homogénéité).

9.2.2 Énergies géométriques globales

En section 7.1.2 p.154, nous avions fait la distinction entre les énergies géométriques fortement séparables - ou locales - qui sont distribuables sur les frontières d'une partition, et les énergies globales, qui ne le sont pas. Nous nous intéressons ici aux énergies globales qui s'appuient sur un modèle d'approximation polygonale des frontières de chaque région. Dans cette section, nous considérons une région x d'une partition et la décomposition de sa frontière inter-pixellaire δx en composantes connexes $f_1 \cup \cdots \cup f_p$ correspondant chacune à un bord de la région (au sens topologique). Les courbes f_i sont ouvertes si elles se connectent au bord du domaine de l'image, fermées sinon.

Nous considérons alors des énergies qui s'écrivent globalement pour une partition P :

$$\mathsf{C}(P) = \sum_{x \in P} \sum_{f \in \delta x} \mathsf{C}(f)$$

ce qui signifie que les énergies des différentes composantes de la frontière d'une région sont traitées de façon indépendante. Notons que comme l'algorithme d'escalade est invariant par homothétie, nous omettons ici le facteur 1/2 de redondance globale de la description (cf. section 7.1.2 p.154).

Si f désigne une portion connexe de frontière inter-pixellaire nous notons alors $\check{f}_\sigma$ une approximation polygonale de f de précision σ (l'algorithme est décrit en annexe D).

Dans le formalisme du codage minimal, la longueur de description $\mathcal{L}_{1,2}(f\langle\check{f}_\sigma\rangle)$ de f utilisant le modèle géométrique $\check{f}_\sigma$, s'écrit alors

$$\mathcal{L}_{1,2}(f\langle\check{f}_\sigma\rangle) \quad = \quad \mathcal{L}_1(\check{f}_\sigma) + \mathcal{L}_2(f|\check{f}_\sigma)$$

où $\mathcal{L}_1$ et $\mathcal{L}_2$ sont les fonctions longueur-de-description associées à deux langages $\mathfrak{L}_1$ et $\mathfrak{L}_2$.

Si $\check{f}_\sigma$ est une polyligne composée de segments qui une fois discrétisés coïncident avec ceux de f, alors $\mathcal{L}_2(f|\check{f}_\sigma) = 0$: la donnée de $\check{f}_\sigma$ suffit à reconstruire exactement f. Si la polygonale ne correspond pas seulement à une version idéalisée des chemins discrets mais à une simplification, alors - dans l'esprit du codage minimal - il faut également transmettre les erreurs et optimiser le message total.

Nous adoptons toutefois un point de vue différent, plus proche de l'esprit débit/distorsion. Nous considérons en effet que la précision de description σ est un paramètre réglé par l'utilisateur qui correspond à la précision géométrique du modèle recherché, ou encore à son degré de généralisation en termes cartographiques. Dans ce cas, la complexité géométrique d'une région est donnée par le minimum de $\mathcal{L}_1(\check{f}_\sigma)$ sur toutes les σ-approximations de f (minimisation de complexité à distorsion bornée). Cette approche revient à considérer que le terme $\mathcal{L}_2(f|\check{f}_\sigma)$ est nul pour toute σ-approximation de f, i.e. que les erreurs inférieures à σ sont négligeables, et a contrario que les approximations de moindre précision que σ sont à distance infinie de f. La description de la partition s'effectue donc avec pertes (contrôlées) ce qui peut s'interpréter comme le fait que le modèle que l'on cherche à inférer est la partition polygonisée et non la partition pixellaire qui sert de support pour le calcul de l'énergie d'attache aux données.

Nous nous plaçons en outre dans une optique de caricature visuelle ou de segmentation et non dans une optique de compression, ce qui a plusieurs incidences importantes. Même lorsque nous adopterons une approche par codage pour dériver des énergies, l'objectif ne sera pas de mettre en place des schéma de codage qui permettraient une reconstruction effective de la géométrie - même polygonisée - de la partition, comme le fait par exemple Pateux (1998). Dans l'optique adoptée, s'appuyer sur l'imagination d'un langage de description $\mathfrak{L}_1$ pour évaluer la complexité d'une forme a essentiellement un rôle *qualitatif*. Les langages envisagés doivent alors refléter au mieux nos préférences pour certaines formes particulières. Insistons donc à nouveau sur l'importance de critères d'invariance dans le choix d'un langage. En particulier, si le codage de certains éléments conduirait à rompre certaines invariances alors nous choisirons de les omettre. Nous ne coderons par exemple pas la position des régions dans l'image : typiquement, spécifier une origine de tracé avant chaque description polygonale introduirait un terme constant qui, combiné avec d'autres termes, pourrait briser certaines invariances.

Dans les expériences qui sont reproduites ici, nous avons approximé f par un algorithme donné et nous avons attribué a posteriori une complexité à l'approximation trouvée. Le choix de la mesure de

complexité n'est donc pas nécessairement lié au terme de régularisation utilisé dans l'algorithme d'approximation polygonale. Nous voyons aujourd'hui qu'il serait plus cohérent de considérer un algorithme d'approximation qui tendrait à minimiser la complexité finalement attribuée au modèle polygonal (ce point n'est toutefois certainement pas critique pour les degrés d'approximation considérés).

Nous commençons par décrire toutes les énergies géométriques que nous avons envisagées. La comparaison des résultats interviendra à la fin de la section.

9.2.3 Nombre de traits d'une description graphique

Le premier coût de description géométrique d'une région que nous avons considéré est le *nombre de traits* nécessaire pour décrire sa frontière à précision σ, soit tout simplement le nombre de segments composant son approximation polygonale à σ. Nous noterons l'énergie correspondante $\mathsf{T}\sigma$.

Ce critère de complexité d'un tracé est relativement intuitif. Dans le domaine de la restitution photogrammétrique, on estime généralement que le temps nécessaire à un restituteur professionnel pour délimiter un objet géographique est globalement proportionnel à la complexité polygonale de la forme restituée. Dans un esprit «langage de description», T_σ représente le nombre d'instructions qu'il faut donner à un traceur pour qu'il dessine la ligne une fois qu'il est positionné sur son extrêmité. D'un point de vue codage, un tel terme de complexité revient à coder la position de chaque extrêmité d'un segment de manière équiprobable dans le domaine de l'image. Le nombre de bits nécessaire augmente donc avec la taille du domaine, ce qui signifie que le code n'est *a priori* pas invariant par focalisation. Toutefois, les hiérarchies d'escalade étant invariantes par homothétie, nous pouvons nous contenter de spécifier l'énergie à un facteur près.

Notons que pour évaluer le nombre de traits, la qualité géométrique de l'approximation n'a pas d'importance : elle peut par exemple délocaliser les angles sans que cela n'ait d'incidence. Par contre, la précision σ a évidemment une incidence forte : décomposer une ligne «presque» droite en deux ou, en augmentant à peine σ, la décrire par deux traits change du simple au double l'énergie (une idée pour réduire l'influence du choix a priori de σ serait de s'arrêter à une approximation stable, c'est-à-dire - en termes de scale-sets - sur une solution relativement persistante).

Forme non séparable

Disons un mot de la forme non séparable du critère du *nombre de traits*, qui est assez intéressante.

Soit P une partition décrite par son graphe de frontières $G = (J, B)$. Soit D_σ un prédicat «être droite» sur les courbes connexes, qui est vrai ssi la courbe est une droite à la précision σ, au sens d'une mesure d'erreur donnée (typiquement max, moyenne ou écart-type des écarts à une droite). Définissons alors une segmentation de G en droites comme une partition des arêtes de G en chaînes connexes[6] $C_1 \dots C_p$ toutes droites, i.e. telles que $\forall i \in \langle 1, p \rangle : D_\sigma(C_i)$. Le *nombre de traits*, $\mathsf{T}_\sigma(G)$, de G est alors le cardinal de la plus petite segmentation de G. $\mathsf{T}_\sigma(G)$ représente donc le nombre minimal de segments de droite qui permettent de recouvrir G à la précision σ.

Cette énergie n'est absolument pas monotone sur le treillis des partitions ; elle présente même de nombreux minima locaux au sens de la 2-normalité. La figure 9.12 illustre ce point pour un cas de synthèse utilisant le prédicat strict, qui n'est vrai que si tous les points d'une chaîne sont rigoureusement alignés ($\sigma = 0$).

Forme séparable

La forme séparable est, elle, décroissante pour des prédicats réalistes. La figure 9.13 l'illustre pour le prédicat strict D_0. Cette figure donne toutes les configurations géométriques possibles de deux régions

[6] Une telle partition s'appelle une *dissection* de G en théorie des graphes.

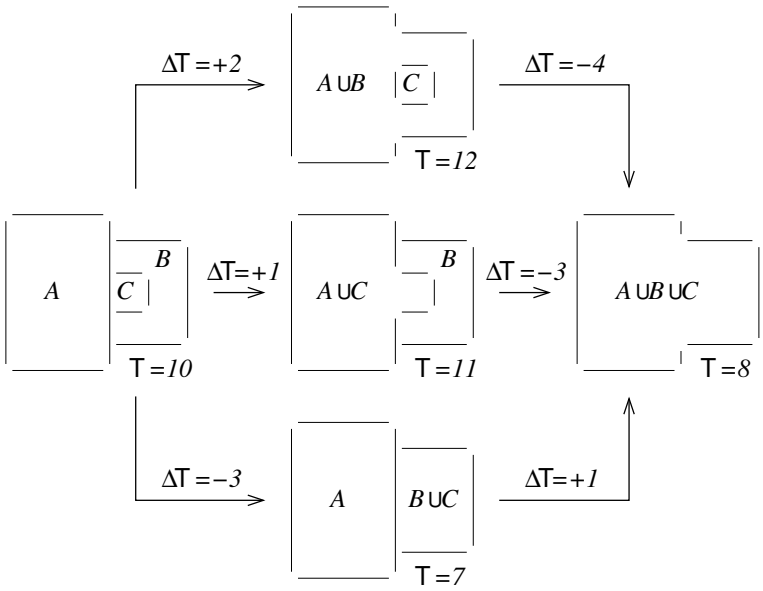


FIG. 9.12 – L'énergie géométrique T_0 non séparable n'est pas monotone. Les segments représentent les chaînes droites maximales du graphe. Dans cet exemple, il n'existe aucun chemin d'énergie décroissante entre $\{A, B, C\}$ et $\{A \cup B \cup C\}$, bien que l'énergie de l'union soit plus faible.

voisines séparées par une frontière composée d'un seul segment et les variations d'énergie associées à leur fusion.

La variation du nombre de traits dépend de la nature géométrique des jonctions au niveau de la frontière supprimée. Si la jonction est une jonction en 'T', supprimer une demi-barre du 'T' pour aboutir à un angle 'Γ' ne diminue pas l'énergie. Par contre supprimer le pied du 'T' pour aboutir à une frontière droite diminue l'énergie de 2. Si la jonction est une jonction en 'Y', supprimer la frontière entre les deux régions diminue l'énergie de 1. Évidemment si la frontière entre les régions n'est pas droite mais est une polyligne brisée, l'énergie diminue d'autant plus que la polyligne est brisée.

Cette énergie se comporte donc totalement différemment de l'énergie de longueur utilisée dans le modèle constant par morceaux de Mumford et Shah (1989). Ces auteurs ont en effet montré qu'en continu, les jonctions triples du minimum étaient toutes en 'Y', délimitant trois secteurs égaux de 120°. A contrario, l'énergie T favorise la formation de jonctions en 'T'.

La figure 9.14 illustre le comportement du critère de complexité T_1 sur un cas réel.

9.2.4 Codage polaire de lignes polygonales

Codage avec changements de direction équiprobables

Dans (Guigues et al. 2001b) nous avons proposé d'utiliser un mode de codage d'une polyligne qui généralise le codage par déplacement pixellaire de Freeman. Au lieu de n'autoriser que des déplacements d'un pixel, on autorise des sauts arbitraires. On pourrait alors coder chaque déplacement en coordonnées cartésiennes (dx, dy) (distance «city-bloc»), cependant le code obtenu n'est pas invariant par rotation. Nous choisissons donc de coder chaque saut en coordonnées polaires (d, a) (distance, angle).

Nous commençons par transmettre d par un code «universel» de Rissanen pour les entiers ce qui emploie $\mathcal{L}^0(d)$ bits (cf. équ. 6.12 p.123). L'emploi de ce code est parfaitement naturel ici : les distances sont positives et il est logique de requérir que les plus faibles distances obtiennent les plus faibles coûts de codage. En outre, les distances sont naturellement quantifiées à la résolution de l'image.

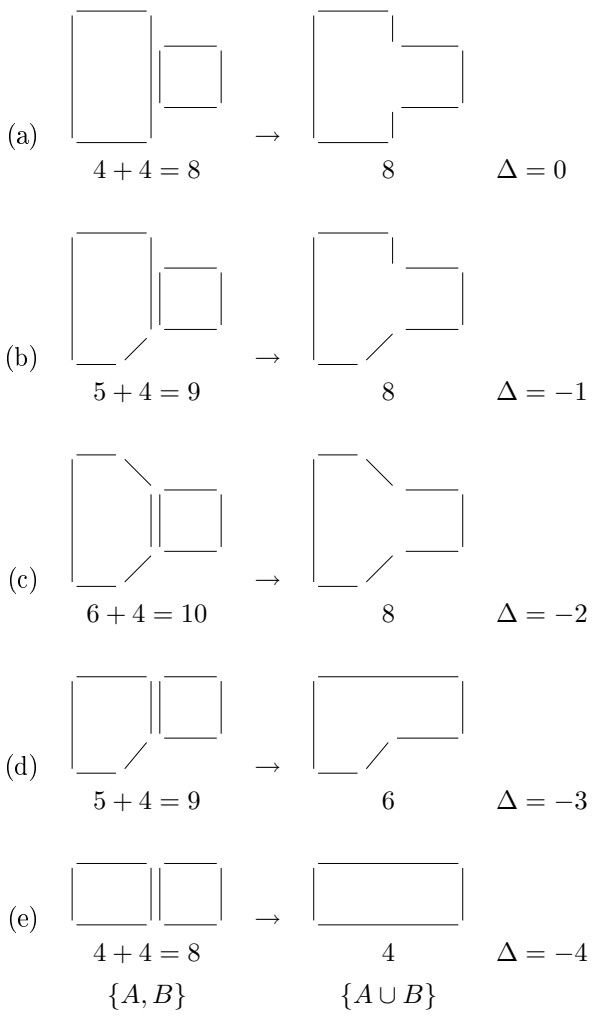


FIG. 9.13 – Variation de l'énergie géométrique T_0 par fusion de couples de régions (cas de synthèse).

Nous spécifions ensuite la position du point d'arrivée parmi les $2\pi d$ positions possibles sur le cercle de rayon d échantillonné à pas de 1, ce qui nécessite approximativement $\log(2\pi d)$ bits si les positions sur le cercle sont équiprobables. On obtient alors une longueur de code pour un saut à distance d :

$$\begin{aligned}
\mathsf{P}(d) &\triangleq \mathcal{L}^0(d) + \log(2\pi d) \\
&= \log 2.865 + \log d + \log\log d + \cdots + \log 2\pi + \log d \\
&= 4,17 + 2\log d + \log\log d + \ldots .
\end{aligned}$$

Si $\check{f}_\sigma$ est une polyligne qui comporte p segments de longueurs $d_1 \ldots d_p$, nous approximons finalement sa complexité pour le langage qui vient d'être décrit par :

$$\mathsf{P}(\check{f}_\sigma) \simeq 5p + \sum_{i=1}^{p} \log d_i^2. \tag{9.4}$$

Du fait de la sous-additivité du logarithme, à longueurs égales, ce code attribue les plus faibles énergies aux lignes les moins brisées. Il est invariant par rotation si l'algorithme d'approximation polygonale l'est, mais n'est absolument pas invariant par homothétie. Remarquons que le critère du nombre de traits est proportionnel au premier terme de 9.4. Ce terme est cependant négligeable quand $d_i \gg 2^{5/2}$.

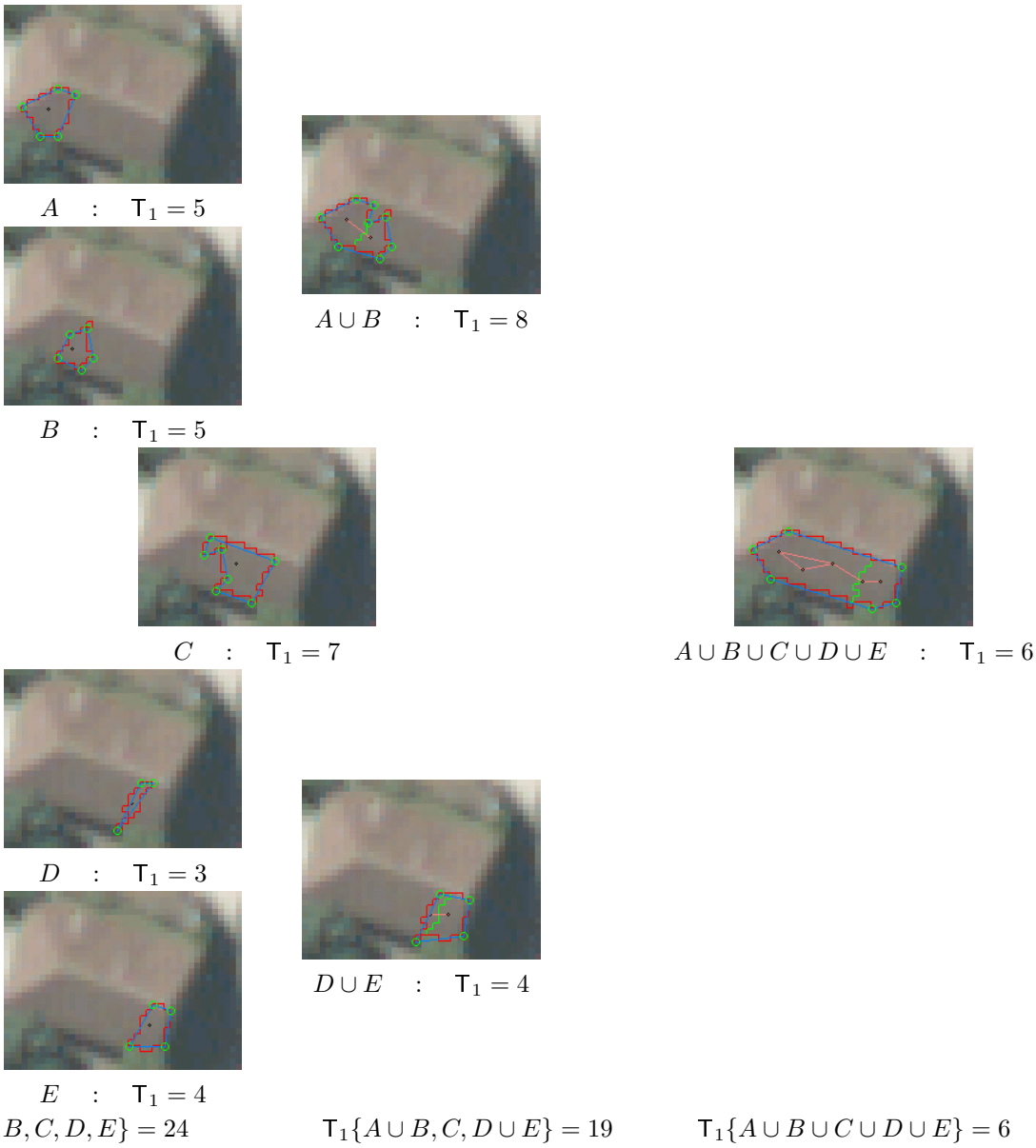


FIG. 9.14 – Variation de l'énergie géométrique T_1 par fusion de couples de régions (exemple réel). Rouge : frontières interpixellaires des régions. Bleu : frontière polygonisée ($\sigma = 1$). Rose : sous-graphes unions. Cet exemple correspond à une fusion par cocons durs, qui ne prend donc pas en compte la géométrie de l'union dans le calcul des priorités de fusion (de toute évidence $B \cup C$ aurait produit une bien plus forte diminution d'énergie géométrique que $A \cup B$).

Nous notons finalement P_σ l'énergie géométrique correspondant à une σ-approximation des frontières des régions suivie d'un codage polaire des polylignes par 9.4.

Codage avec changements de direction gaussiens

Le codage polaire que nous venons de voir considère tous les angles comme équiprobables (une simulation du modèle a priori donnerait quelque chose de proche d'un mouvement Brownien). On peut bien entendu envisager une autre loi a priori, typiquement qui favorise les changements de direction faibles.

L'angle sera donc spécifié relativement à la direction courante de tracé : si la polyligne est constituée de p vecteurs $v_1 \dots v_p$, nous notons $\theta_i \triangleq (\widehat{v_{i-1}, v_i})$, $i \in \langle 2, p \rangle$, le $i^{\text{ème}}$ changement de direction le long de la polyligne (voir la figure 9.15). Si la ligne est fermée, nous définissons alors θ_1 par $\theta_1 \triangleq (\widehat{v_p, v_1})$. Si la ligne se raccorde à un bord de l'image, dans le doute, nous posons ici $\theta_1 \triangleq (\widehat{n, v_1})$, où n est le vecteur normal à la frontière de l'image.

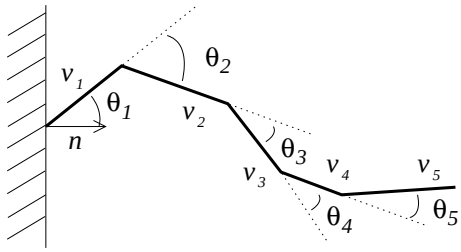


FIG. 9.15 – Changements de direction le long d'une ligne polygonale.

Si P est une loi a priori sur les changements de direction, alors en reprenant le schéma de codage polaire précédent nous obtenons une énergie qui prend la forme (avec les mêmes approximations) :

$$2p + \sum_{i=1}^{p} \log ||v_i|| - \sum_{i=1}^{p} \log P(\theta_i)$$

Nous avons alors considéré une loi gaussienne de moyenne nulle comme loi a priori.

Si la variance de la gaussienne est traîtée comme un paramètre inconnu à estimer alors en développant un schéma de codage analogue à celui décrit ci-dessous pour les modèles radiométriques (section 9.3, voir essentiellement l'équation 9.14 p.242), on aboutit à une énergie qui s'écrit

$$\mathsf{P}^{\mathcal{N}}(\check{f}) \quad \triangleq \quad 2p + \sum_{i=1}^{p} \log ||v_i|| + \frac{p}{2} \log \left(2\pi e \frac{\Sigma_\theta^2}{q^2} \right) \tag{9.5}$$

où $\Sigma_\theta^2 \triangleq \frac{1}{p} \sum_{i=1}^{p} \theta_i^2$ désigne la variance des changements de direction et q est le pas de quantification des angles utilisé pour leur codage.

Nous noterons $\mathsf{P}_\sigma^{\mathcal{N}}$ l'énergie qui correspond à des approximations polygonales de précision σ et leur codage selon 9.5 où q est arbitrairement fixé à 256 (le choix n'est cependant pas critique).

9.2.5 Énergie de courbure/concavité

Définition

L'énergie de codage que nous venons de voir n'est pas invariante par homothétie en raison des deux premiers termes dans l'expression 9.5. Nous écartant d'une approche par codage, nous avons alors considéré les énergies qui s'écrivent

$$\sum_{i=1}^{p} g(\theta_i)$$

où g est une fonction paire croissante sur $\mathbb{R}^+$. De telles énergies sont liées à la courbure (discrète) des frontières.

Après plusieurs essais, le choix de g qui a donné les meilleurs résultats est tout simplement $g(x) = |x|$.

Cette énergie a une interprétation géométrique simple : elle mesure le degré de non convexité des formes. En effet, la somme des angles d'un polygone convexe vaut systématiquement 2π. Si le polygone

présente des concavités, alors la somme des *valeurs absolues* de ses angles augmente avec l'importance des concavités, aussi bien en nombre qu'en «profondeur».

Nous appelons donc l'énergie

$$\mathsf{C}(\check{f}) \quad \triangleq \quad \frac{1}{2\pi}\sum_{i=1}^{p}|\theta_i| \tag{9.6}$$

l'énergie de concavité (normalisée) de la polyligne[7].

La figure 9.16 illustre le comportement de cette énergie sur un certain nombre de formes (les polygones ont été dessinés à la main, les sommets cliqués sont représentés).

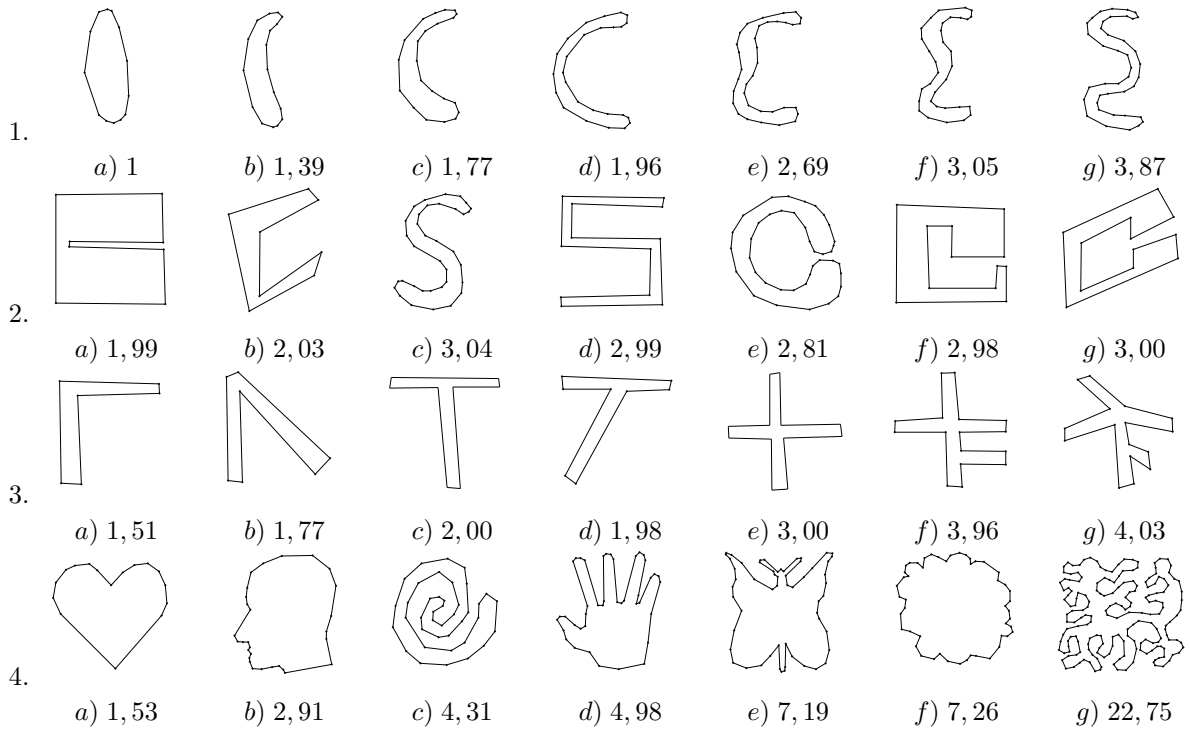


FIG. 9.16 – Exemples de valeurs d'énergie de concavité normalisée. Voir texte.

Si la forme est convexe son énergie vaut 1 strictement (exemple de la forme 1.a de la figure 9.16). Si la forme présente une concavité, son énergie augmente avec la courbure totale de cette concavité (1.b à 1.d). Le $'C'$ de la figure 1.d a ainsi une énergie d'environ 4π (soit 2 après normalisation) : il est approximativement formé de 4 demi-cercles, trois convexes et un concave. Les formes 2.a, 2.b, 3.c et 3.d on également une énergie d'à peu près 2. Elles montrent que pour l'énergie C, seule la somme des changements de direction a de l'importance, indépendamment de leur position ou de leur «concentration» : deux angles brusques de $\pi/2$ - comme en 2.a - ou une somme d'angles faibles induisant un changement de direction total de π - comme en 1.d - sont équivalents. De même les deux concavités à $\pi/2$ du 'T' de la figure 3.c ont beau être séparées par deux convexités (le haut et le pied du 'T') elles totalisent la même énergie.

La fin de la séquence 1. (formes 1.e à 1.g) simule l'introduction progressive d'une nouvelle concavité à partir du 'C' vu en 1.d. La forme 1.g a une énergie proche de 4 : elle comporte approximativement 3 creux en demi-cercle. De leur côté, les formes 2.c à 2.g ont pour point commun de comporter deux creux de courbure totale équivalente à un celle d'un demi-cercle. Elle ont donc une énergie de concavité d'environ 3. La forme 3.a possède une seule concavité de courbure équivalente à un quart de cercle, celle 3.b une concavité de courbure équivalente à un huitième de cercle.

[7] Notons qu'ici nous posons $\theta_1 = 0$ si la ligne est ouverte.

L'énergie de concavité d'une forme convexe est inchangée par toute transformation qui laisse la forme convexe, soit en particulier par transformations affines et projectives (par exemple tous les triangles, quadrilatères convexes ou ellipses ont la même énergie de concavité). La figure 9.16 montre que l'énergie est également robuste à un certain nombre de déformations sur des objets concaves (comparer 2.a et 2.b, 2.f et 2.g, 3.c et 3.d, 3.f et 3.g), mais n'est en général pas invariante par transformation affine (comparer 3.a et 3.b).

Elle est bien entendu invariante par homothétie, ainsi que par une sorte d'homothétie «locale», i.e. par dilatation de seulement une excroissance de la forme (étirer par exemple les branches du 'T' ne change pas son énergie du moment qu'on ne les courbe pas).

Cette qualification des formes a certainement déja été utilisée en reconnaissance de forme, nous n'avons cependant pas fait de recherches dans la littérature consacrée. Elle n'a par contre pas été utilisé - à notre connaissance - en segmentation.

Variation de concavité par fusion

Voyons maintenant la manière dont varie l'énergie C lors de la fusion de deux régions. Considérons pour cela deux régions A et B voisines, séparées par une frontière f d'un seul tenant qui se connecte au niveau de deux jonctions j_1 et j_2 (voir figure 9.17).

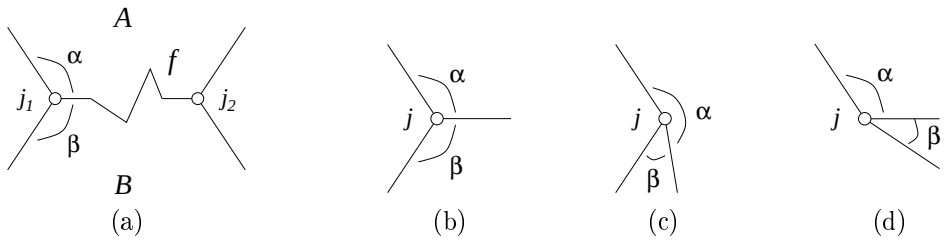


FIG. 9.17 – Variation d'énergie de concavité par fusion. Voir texte.

Lors de la fusion de A et B la variation d'énergie de concavité est de

$$\Delta\mathsf{C} \quad = \quad \Delta\mathsf{C}(j_1) + \Delta\mathsf{C}(j_2) - 2\mathsf{C}(f)$$

où $\Delta\mathsf{C}(j)$ désigne la variation d'énergie survenant au niveau de la jonction j et $\mathsf{C}(f)$ représente l'énergie de courbure propre de la polyligne f (la somme de ses changements de direction interne), indépendemment de ses connexions terminales. Voyons alors ce qui se produit au niveau d'une jonction j.

Notons α et β les angles respectifs que font les frontières des régions A et B au point j (cf. fig.9.17(a)). On voit alors immédiatement que

$$\Delta\mathsf{C}(j) \quad = \quad |\pi - \alpha - \beta| - |\pi - \alpha| - |\pi - \beta|.$$

Plusieurs cas se présentent donc :

1. Si $\alpha + \beta > \pi$ (i.e. si la région union est localement concave)

i. Si $\alpha < \pi$ et $\beta < \pi$ (i.e. si les deux régions A et B sont localement convexes, cf. fig.9.17(b)) on trouve alors $\Delta\mathsf{C}(j) = 2(\alpha + \beta) - 3\pi$. Ainsi $\Delta\mathsf{C}(j) < 0$ si et seulement si $\alpha + \beta < \frac{3\pi}{2}$. La fusion fait donc décroître l'énergie au niveau de la jonction j quand le changement de direction de la frontière de l'union en j est inférieur à $\frac{\pi}{2}$.

ii. Si $\alpha > \pi$ (i.e. si A est localement concave, cf. fig 9.17(c) - le cas $\beta > \pi$ est symétrique) on trouve $\Delta\mathsf{C}(j) = 2\beta - \pi$. La fusion fait donc décroître l'énergie si et seulement si $\beta < \frac{\pi}{2}$, soit quand le changement de direction de la frontière de B est supérieur à $\frac{\pi}{2}$ en j.

2. Si $\alpha + \beta < \pi$ (i.e. si la région union est localement convexe), alors $\alpha < \pi$ et $\beta < \pi$ (A et B sont localement convexes). On trouve dans ce cas $\Delta\mathsf{C}(j) = -\pi$. L'énergie décroît donc toujours.

L'énergie de concavité peut donc être sur-additive si la fusion introduit de nouvelles concavités.

Toutefois, les calculs qui viennent d'être fait concernent seulement la variation d'énergie au niveau d'une jonction. La variation totale implique l'autre jonction et la courbure de la frontière séparant A et B. Quand la frontière entre les deux régions est brisée, l'énergie totale est généralement décroissante. Si cela n'est pas le cas, nous forçons la sous-additivité en posant $\Delta\mathsf{C} = 0$, ce qui revient à interdire la fusion (échelle d'apparition de l'union infinie).

Par la suite, nous noterons C_σ l'énergie géométrique correspondant à une σ-approximation des frontières des régions et l'attribution d'une énergie de concavité aux approximations (la normalisation n'a aucune importance pour l'escalade).

9.2.6 Une variante séparable de l'énergie géométrique de Ballester et al.

Ballester, Caselles, et González (1994) ont proposé de remplacer l'énergie de longueur dans la fonctionnelle de Mumford-Shah par une énergie qu'ils appellent *variation affine totale* (Affine Total Variation, ATV) qui est invariante par transformation affine de la géométrie de la partition. Les auteurs soulignent l'importance de cette invariance en tant qu'invariance projective approchée.

Soit f et g deux courbes connexes rectifiables paramétrées par leur abcisses curvilignes s et t. Notons $\tau(x)$ le vecteur tangent à une courbe à l'abcisse x. L'interaction de f et g est alors définie par

$$\mathsf{Inter}(f,g) = \int_0^1 \int_0^1 |\tau(s) \wedge \tau(t)| \, ds \, dt$$

Si P est une partition constituée d'un ensemble B de frontières alors Ballester et al. (1994) définissent sa variation affine totale par

$$\mathsf{ATV}(P) = \sum_{f \in B} \sum_{g \in B} \mathsf{Inter}(f,g).$$

Cette énergie n'est cependant pas séparable. Elle est toutefois décroissante sur le treillis des partitions : supprimer des frontières d'une partition diminue de toute évidence son ATV.

Nous avons envisagé une variante séparable de cette énergie qui consiste simplement à restreindre les interactions entre courbes d'une partition aux seules interactions entre frontières d'une même région. Si R est une région de P, décomposons comme précédemment sa frontière δR en composantes connexes. Définissons alors la *variation affine régionale* (ARV) de R par

$$\mathsf{ARV}(R) = \sum_{f \in \delta R} \sum_{g \in \delta R} \mathsf{Inter}(f,g). \tag{9.7}$$

L'énergie de partition s'obtient alors par sommation des énergies régionales.

Si les frontières de la région R sont polygonisées, notons comme précédemment $v_1 \dots v_p$ les vecteurs composant les approximations (ici ces vecteurs proviennent de toutes les composantes connexes). L'ARV de R s'écrit alors

$$\mathsf{ARV}(R) \quad = \quad \sum_{i=1}^{p} \sum_{j=1}^{p} |v_i \wedge v_j| \quad = \quad \sum_{i=1}^{p} \sum_{j=1}^{p} ||v_i|| \, ||v_j|| \, |sin(\widehat{v_i, v_j})|. \tag{9.8}$$

Malheureusement, ARV est globalement sur-additive plutôt que sous-additive. Cela provient d'une non normalisation de la mesure, qui est apparente dans l'expression de droite dans 9.8 : globalement chaque *point* possède une énergie qui dépend de la longueur totale des courbes[8]. Cela signifie que globalement,

[8] Les interactions sont homogènes à des surfaces. L'invariance affine provient précisément du fait que la surface d'un parallélogramme est indépendante de son applatissement.

l'énergie ne décroît par fusion que si la longueur de la frontière supprimée $f = \delta A \cap \delta B$ est supérieure aux deux longueurs des frontières restantes $\delta A \backslash f$ et $\delta B \backslash f$ (la variation dépend aussi des directions mutuelles). En pratique, si l'on force la sous-additivité, c'est-à-dire si l'on interdit les fusions en cas de sur-additivité de l'énergie géométrique, on ne parvient alors à fusionner que de petites régions. Nous avons donc considéré une énergie «normalisée», qui s'écrit

$$\mathsf{B}(R) \quad = \quad \sum_{i=1}^{p} \frac{1}{p-1} \sum_{j=1}^{p} \frac{|v_i \wedge v_j|}{||v_i||\ ||v_j||} \quad = \quad \frac{1}{p-1} \sum_{i=1}^{p} \sum_{j=1}^{p} |sin(\widehat{v_i, v_j})|. \tag{9.9}$$

On peut interpréter cette énergie comme une mesure de la cohérence globale des directions de la frontière de R : chaque terme mesure la coïncidence moyenne entre la direction du vecteur courant et celle des autres vecteurs. Cette énergie «normalisée» n'est cependant plus affine-invariante. Elle est par contre invariante par homothétie.

Remarquons que l'on pourrait ne normaliser que par la longueur d'un des deux vecteurs et non par les deux, ce qui reviendrait à pondérer l'énergie de «cohérence» de chaque vecteur par sa longueur. Ce choix a toutefois donné de moins bons résultats. Les exemples présentés ci-dessous étayeront ce point.

Nous noterons B_σ l'énergie 9.9 associée à une σ-approximation des frontières d'une partition.

9.2.7 Résultats comparatifs

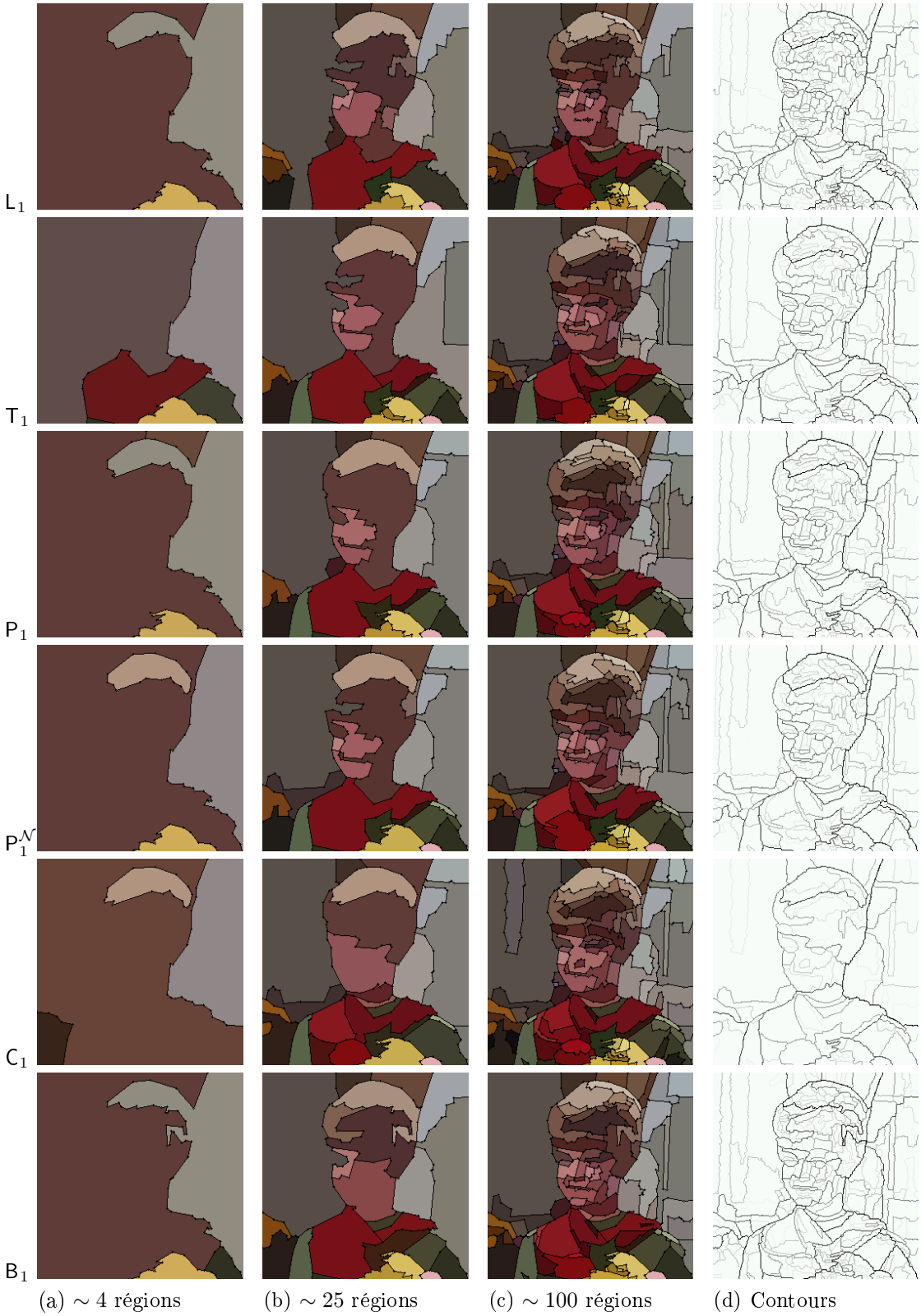
La table 9.5 résume les différentes énergies géométriques considérées.

Énergie		Forme
Constante	K	1
Longueur	L_σ	$\sum_{i=1}^{p} ||v_i||$
Nombre de traits	T_σ	p
Codage polaire	P_σ	$5p + \sum_{i=1}^{p} \log ||v_i||^2$
Codage polaire gaussien	$\mathsf{P}_\sigma^{\mathcal{N}}$	$2p + \sum_{i=1}^{p} \log ||v_i|| + \frac{p}{2} \sum_{i=1}^{p} \log \left(2\pi e \frac{\Sigma_\theta^2}{q^2}\right)$
Concavité	C_σ	$\sum_{i=1}^{p} |(\widehat{v_i, v_{i+1}})|$
Cohérence des directions	B_σ	$\frac{1}{p-1} \sum_{i=1}^{p} \sum_{j=1}^{p} |sin(\widehat{v_i, v_j})|$

TAB. 9.5 – Résumé des énergies géométriques considérées. La table donne l'énergie attribuée à une composante connexe de la frontière d'une région qui a été polygonisée à précision σ, aboutissant à un polyligne composée de p vecteurs v_i (dans le cas de l'énergie de cohérence, toutes les composantes de la frontière sont considérées en bloc). $\Sigma_\theta^2 = \frac{1}{p} \sum_{i=1}^{p} \left(\widehat{v_i, v_{i+1}}\right)^2$.

L'énergie constante K est donnée pour mémoire. Par ailleurs, nous considérons ici une énergie de longueur L_σ qui est définie comme la longueur d'une σ-approximation polygonale de chaque région afin que la comparaison avec les autres critères porte véritablement sur la mesure et ne soit pas biaisée par la discrétisation des frontières.

La figure 9.18 présente les résultats obtenus sur l'image girl pour ces différentes énergies. La précision des approximations polygonales est systématiquement fixée à $\sigma = 1$ pixel (les partitions présentées ont leurs frontières polygonisées à cette précision). L'énergie radiométrique utilisée est à nouveau l'énergie Q de déviations quadratiques et la sur-segmentation est obtenue par LPE, $\alpha = 2$. Nous comparons ici les coupes à nombre équivalent de régions car les dynamiques des scale-sets ne sont pas les mêmes pour les différentes énergies (le nombre de régions est approximatif car il se peut qu'il n'existe pas de partition optimale pour un cardinal donné en raison de la non persistance de certaines régions).



(a) $\sim$ 4 régions (b) $\sim$ 25 régions (c) $\sim$ 100 régions (d) Contours

FIG. 9.18 – Comparaison des énergies géométriques sur l'image girl.

La structure globale de l'image - un objet sombre sur un fond plus clair - est la même pour les différents critères, avec toutefois des différences géométriques. Les critères L_1, P_1 et B_1 fusionnent la partie claire supérieure de la chevelure au fond. Par ailleurs, B_1 incorpore également une excroissance assez irrégulière dans cette partie de la chevelure. D'importantes différences sont visibles sur les coupes à 25 régions. Les énergies L_1, T_1 et $\mathsf{P}_1^{\mathcal{N}}$ attribuent la partie frontale de la chevelure au fond. Elles fusionnent également la partie sombre de la chevelure à la partie ombragée du visage, jusqu'au cou. P_1 réagit également de cette manière. L'œil droit du personnage se présente comme une pièce distincte du visage, rattachée soit au fond (L_1, T_1 et B_1), soit à la chevelure (P_1 et $\mathsf{P}_1^{\mathcal{N}}$). L'énergie C_1 a un comportement relativement différent des autres. Elle délimite le visage en une pièce, incorporant l'œil et la nuque en une pièce globalement convexe (le seul défaut majeur de la coupe à 25 régions est le rattachement de la chevelure sombre au fond par un goulot étroit).

La figure 9.19 compare les coupes à 25 régions obtenues sur l'image champs pour les différentes énergies géométriques. Cet exemple met en évidence une importante différence de comportement entre les énergies C et B - liées à la courbure - et les autres - globalement liées à la longueur. L'énergie L, dépendante de la longueur des frontières, ne segmente pas correctement les objets filiformes (la route, les bas côtés). En effet, pour cette énergie, fusionner un élément fin et long à un autre élément fin dans son prolongement est nettement moins intéressant que fusionner l'objet à une zone adjacente : grouper deux objets fins alignés ne supprime qu'une très courte frontière, alors que grouper un objet fin et long avec une zone latérale supprime une frontière longue. Les énergies T, P et $\mathsf{P}^{\mathcal{N}}$ réagissent également de cette manière, mais dans une moindre mesure. T ne restitue pas les éléments fins et courbes (le virage en haut de l'image) car ils sont constitués de nombreux traits. P et $\mathsf{P}^{\mathcal{N}}$, qui dépendent des logarithmes des longueurs des portions droites de frontières, ne parviennent pas non plus à segmenter les objets filiformes. On notera également la différence de segmentation du champs texturé en bas à droite de l'image.

Après de nombreux autres essais, l'énergie de concavité C nous paraît globalement fournir de meilleurs résultats que les autres énergies. En outre, nous avons vu qu'elle était invariante par homothétie et robuste à d'importantes déformations des objets. C'est l'énergie géométrique que nous préconisons aujourd'hui (la variante B de l'énergie de Ballester et al. n'est pas mauvaise non plus mais le critère de cohérence globale des directions a parfois des comportements étranges).

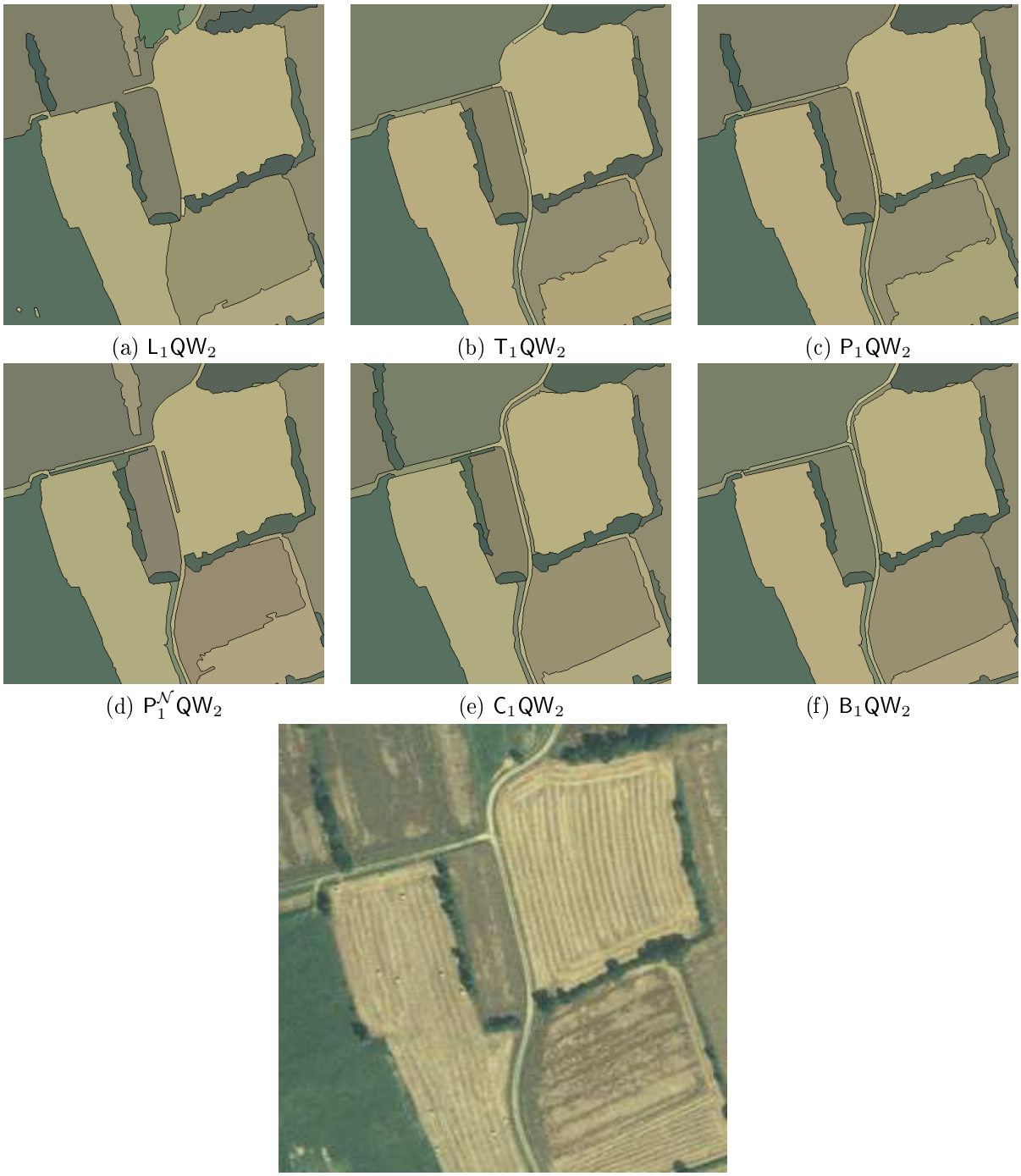
Notons que dans notre implémentation actuelle de l'algorithme, l'exécution sur l'image girl (256×256, 3905 bassins versants) selon l'énergie L (sans polygonisation) emploie environ 2 secondes de CPU à $1,2$ GHz alors que pour les autres énergies, le temps de calcul est d'environ 15 secondes (17 secondes pour l'énergie B qui est quadratique en fonction de p, cf. équ. 9.9). Pour l'image champs (512×512, 17 254 bassins versants) le temps passe d'une dixaine de secondes à approximativement 3 minutes.

Le code actuel n'est cependant absolument pas optimisé : nous re-parcourrons systématiquement le graphe de base pour trouver la frontière inter-pixellaire d'une région que nous re-polygonisons complètement pour en évaluer la complexité. Le temps de calcul de l'énergie géométrique croît donc avec la taille des régions (voir l'annexe C pour une étude théorique de l'impact du coût de calcul des énergies sur la complexité combinatoire d'une escalade). On pourrait toutefois réduire le temps de calcul de polygonisation à une constante (faible) en réalisant des approximations incrémentales : comme nous l'avons vu, les seuls points où peuvent intervenir des simplifications de forme par fusion sont les jonctions entre régions voisines. Nous avons tout lieu de penser que l'on pourrait implémenter des escalades s'appuyant sur des énergies utilisant des polygonisations qui soient pratiquement aussi efficaces que des escalades bâties sur une énergie de longueur.

9.3 Modèles radiométriques

Intéressons-nous maintenant à l'énergie d'attache aux données D.

Leclerc (1989a) a proposé de modéliser les valeurs d'une image numérique sur une région comme



(a) $\mathsf{L}_1\mathsf{QW}_2$ (b) $\mathsf{T}_1\mathsf{QW}_2$ (c) $\mathsf{P}_1\mathsf{QW}_2$

(d) $\mathsf{P}_1^{\mathcal{N}}\mathsf{QW}_2$ (e) $\mathsf{C}_1\mathsf{QW}_2$ (f) $\mathsf{B}_1\mathsf{QW}_2$

(g) champs. 512×512

FIG. 9.19 – Comparaison des énergies géométriques sur l'image champs.

résultant de la quantification d'une séquence de réalisations indépendantes d'une loi gaussienne. Les paramètres de la loi sont inconnus et doivent donc être estimés. Dans une approche par codage, Leclerc propose alors d'attribuer une énergie à chaque région qui correspond à un codage optimal des observations selon ce modèle. Leclerc dérive l'expression de cette énergie pour les images à un canal. Nous étendons ici son approche au cas multivarié. Nous voyons alors que l'énergie prend une forme relativement simple.

Une comparaison de l'expression trouvée à l'énergie de déviations quadratiques Q nous conduit ensuite à proposer plusieurs variantes. La fin de la section compare les résultats expérimentaux obtenus pour les différentes énergies envisagées.

9.3.1 Codage des radiométries selon un modèle gaussien

Notons $\mathcal{N}_{\mu,\Sigma}$ la distribution normale à d dimensions de moyenne μ et de matrice de variance/covariance Σ. Supposons que les observations X_i correspondent à la quantification d'observations réelles ayant été émises selon une loi $\mathcal{N}_{\mu,\Sigma}$. Supposons que le pas de quantification des observations est identique dans toutes les directions de l'espace. Notons q ce pas (en pratique $q = 1$ sur une image couleur à valeurs dans $\langle 0, 256\langle^3)$. La probabilité pour qu'un tirage selon $\mathcal{N}_{\mu,\Sigma}$ tombe dans la cellule de centre $X = (X^1 \dots X^d)$ est alors donnée par

$$P(X|\mathcal{N}_{\mu,\Sigma}) = \int_{X^1-q/2}^{X^1+q/2} \dots \int_{X^d-q/2}^{X^d+q/2} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(Y-\mu)^t\Sigma^{-1}(Y-\mu)} dY^1 \dots dY^d. \tag{9.10}$$

Cette intégrale d-dimensionnelle porte sur la cellule de volume q^d centrée en X.

Supposons que le pas de quantification est suffisant, i.e. que q est inférieur aux écarts types de la gaussienne dans chacune de ses directions principales. En suivant Leclerc, il est alors légitime d'approximer l'intégrale 9.10 par

$$P(X|\mathcal{N}_{\mu,\Sigma}) \simeq \frac{q^d}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu)^t\Sigma^{-1}(X-\mu)}.$$

Si l'on utilise cette loi discrète pour définir un code préfixé optimal au sens de Shannon, la longueur du mot-code d'une observation X est alors (exprimée en nats[9]) :

$$\begin{aligned} \mathcal{L}\big(X|\mathcal{N}_{\mu,\Sigma}\big) &= -\ln P(X|\mathcal{N}_{\mu,\Sigma}) \quad \text{nats} \\ &\simeq \frac{1}{2}\left[\ln\left((2\pi)^d \frac{|\Sigma|}{q^{2d}}\right) + (X-\mu)^t\Sigma^{-1}(X-\mu)\right] \quad \text{nats.} \end{aligned}$$

Intéressons-nous maintenant au codage d'un échantillon $X = \{X_1 \dots X_N\}$ de N observations, supposées avoir été produites de manière indépendante. La probabilité de l'échantillon est donc

$$P(X_1, \dots, X_N|\mathcal{N}_{\mu,\Sigma}) = \prod_{i=1}^{N} P(X_i|\mathcal{N}_{\mu,\Sigma})$$

et sa longueur totale de description

$$\begin{aligned} \mathcal{L}\big(X|\mathcal{N}_{\mu,\Sigma}\big) &= -\ln P(X_1, \dots, X_N|\mathcal{N}_{\mu,\Sigma}) = \sum_{i=1}^{N} -\ln P(X_i|\mathcal{N}_{\mu,\Sigma}) = \sum_{i=1}^{N} \mathcal{L}\big(X_i|\mathcal{N}_{\mu,\Sigma}\big) \\ &\simeq \frac{N}{2}\ln\left((2\pi)^d \frac{|\Sigma|}{q^{2d}}\right) + \frac{1}{2}\sum_{i=1}^{N}(X_i-\mu)^t\Sigma^{-1}(X_i-\mu). \end{aligned} \tag{9.11}$$

Les paramètres du modèle gaussien qui minimisent 9.11 sont la moyenne empirique $\overline{X}$ et la matrice de variance/covariance Σ_X des X_i (L'équation 9.11 est l'opposé de la log-vraisemblance de X sachant un modèle gaussien). La longueur minimale de description s'obtient alors en remarquant qu'à l'optimum le dernier terme de 9.11 se simplifie en

$$\sum_{i=1}^{N}(X_i-\overline{X})^t\Sigma_X^{-1}(X_i-\overline{X}) = Nd.$$

[9] Natural digits : unité de codage associée au logarithme naturel, de base e. Diviser les nats par $\ln 2$ pour obtenir des bits (binary digits).

On aboutit donc à

$$\mathcal{L}(X|\mathcal{N}_{\overline{X},\Sigma_X}) \simeq \frac{N}{2}\left[\ln\left((2\pi)^d\frac{|\Sigma_X|}{q^{2d}}\right)+d\right] = \frac{N}{2}\ln\left((2\pi e)^d\frac{|\Sigma_X|}{q^{2d}}\right). \tag{9.12}$$

Notons λ_i les valeurs propres de Σ_X et $\sigma_i \triangleq \sqrt{\lambda_i}$ les écarts-types associés. Comme $|\Sigma_X| = \prod_{i=1}^d \lambda_i$, 9.12 s'écrit également

$$\mathcal{L}(X|\mathcal{N}_{\overline{X},\Sigma_X}) \simeq \frac{N}{2}\ln\left((2\pi e)^d\prod_{i=1}^d\frac{\lambda_i}{q^2}\right) \tag{9.13}$$

$$= \frac{N}{2}\sum_{i=1}^d\ln\left(2\pi e\frac{\lambda_i}{q^2}\right) = N\sum_{i=1}^d\ln\left(\sqrt{2\pi e}\frac{\sigma_i}{q}\right). \tag{9.14}$$

Cette expression peut finalement s'interpréter comme la somme de d expressions de coûts de codage 1D dans chaque direction propre de la distribution.

Remarquons que la longueur de description trouvée *ne dépend pas de la loi effectivement suivie par* X. En particulier, elle ne dépend absolument pas du fait que X ait ou non été produit par une loi normale, mais seulement de ses statistiques d'ordre deux. Ceci traduit le fait que la longueur de codage (ou de manière équivalente $P(X|M)$) ne contient absolument aucune information sur la validité de l'hypothèse gaussienne pour l'échantillon. Cette longueur de description (ou probabilité) ne peut être utilisée que de manière *relative*, en comparaison aux longueurs de description obtenues pour d'autres modèles (dans notre cas la comparaison s'effectue entre une région fusionnée, donc modélisée par une seule gaussienne, et la région divisée, modélisée par plusieurs gaussiennes).

À cause du logarithme, l'expression 9.14 n'est pas invariante par transformation linéaire des observations. Toutefois, une escalade utilisant ce type d'attache aux données l'est.

Rappelons en effet que l'échelle d'apparition d'un nœud x s'écrit globalement $-\frac{\Delta\mathsf{D}}{\Delta\mathsf{C}}$. Notons alors $N = |x|$ et supposons que la partition P de la hiérarchie partielle $H(x)\backslash x$ qui est optimale en $\lambda^+(x)$ est composée de k régions x_i de cardinaux $N_1 \dots N_k$ (évidemment $\sum_1^k N_j = N$). Notons finalement σ_i^j l'écart-type de la distribution de la région j dans la $i^{\text{ème}}$ direction propre. La variation d'énergie d'attache aux données D quand P est fusionnée en x s'écrit alors

$$\begin{aligned}
\Delta\mathsf{D} &= \mathsf{D}(x) - \sum_{j=1}^k \mathsf{D}(x_i) \\
&= N\ln\left[(2\pi e)^{d/2}\frac{\prod_{i=1}^d\sigma_i}{q^d}\right] - \sum_{j=1}^k N_j\ln\left[(2\pi e)^{d/2}\frac{\prod_{i=1}^d\sigma_i^j}{q^d}\right] \\
&= \ln\left[(2\pi e)^{Nd/2}\frac{\left(\prod_{i=1}^d\sigma_i\right)^N}{q^{Nd}}\right] - \sum_{j=1}^k\ln\left[(2\pi e)^{N_j d/2}\frac{\left(\prod_{i=1}^d\sigma_i^j\right)^{N_j}}{q^{N_j d}}\right] \\
&= \ln\left[(2\pi e)^{Nd/2}\frac{\left(\prod_{i=1}^d\sigma_i\right)^N}{q^{Nd}}\right] - \ln\left[(2\pi e)^{\frac{d}{2}\sum_{j=1}^k N_j}\times\frac{\prod_{j=1}^k\left(\prod_{i=1}^d\sigma_i^j\right)^{N_j}}{q^{d\sum_{j=1}^k N_j}}\right] \\
&= \ln\frac{\left(\prod_{i=1}^d\sigma_i\right)^N}{\prod_{j=1}^k\left(\prod_{i=1}^d\sigma_i^j\right)^{N_j}}.
\end{aligned}$$

$\Delta\mathsf{D}$ ne dépend donc pas du pas de quantification q et est inchangée par multiplication des observations par une constante. Si C est également invariante par une telle transformation alors l'échelle d'apparition des nœuds est inchangée. L'invariance est donc plus forte que pour l'énergie de déviations quadratiques :

l'échelle d'apparition d'une région est indépendante de son contraste absolu. Nous verrons ci-dessous que des régions de variance très faibles mais présentant deux pôles gaussiens bien distincts peuvent ainsi rester divisées très haut dans les hiérarchies.

9.3.2 Entre trace et déterminant...

Vu la propriété qui vient d'être développée pour les escalades, on peut finalement réduire l'énergie de codage gaussien à la plus simple expression :

$$\mathsf{G}(X) \triangleq |X| \ln |\Sigma_X| = |X| \sum_{i=1}^{d} \ln \lambda_i. \quad [10] \tag{9.15}$$

La parenté avec l'expression

$$\mathsf{Q}(X) \triangleq |X| \mathrm{Tr}(\Sigma_X) = |X| \sum_{i=1}^{d} \lambda_i$$

de l'énergie de déviations quadratiques est alors frappante.

Les deux énergies dépendent d'un invariant par rotation de la matrice de variance/covariance, sa trace ou son déterminant. En norme L_2, le coût moyen par échantillon est la somme des valeurs propres de Σ_X. Pour une formulation par codage gaussien optimal, ce coût moyen dépend globalement de la somme des *logarithmes* des valeurs propres de Σ_X[11].

Nous avons alors considéré trois variantes autour des énergies G et Q, que nous notons $\mathsf{Q}^{\frac{1}{2}}$, $\mathsf{D}^{\frac{1}{2}}$ et D. Le tableau 9.6 indique la forme de ces variantes et la manière dont elles se positionnent par rapport à G et Q.

Énergie	G	$\mathsf{Q}^{\frac{1}{2}}$	Q	$\mathsf{D}^{\frac{1}{2}}$	D
Forme matricielle	$\lvert X\rvert \ln \lvert\Sigma_X\rvert$	$\lvert X\rvert \mathrm{Tr}(\sqrt{\Sigma_X})$	$\lvert X\rvert \mathrm{Tr}(\Sigma_X)$	$\lvert X\rvert \lvert\sqrt{\Sigma_X}\rvert$	$\lvert X\rvert \lvert\Sigma_X\rvert$
Forme «propre»	$\lvert X\rvert \sum \ln \lambda_i$	$\lvert X\rvert \sum \sqrt{\lambda_i}$	$\lvert X\rvert \sum \lambda_i$	$\lvert X\rvert \prod \sqrt{\lambda_i}$	$\lvert X\rvert \prod \lambda_i$
Sur-additivité	NON	OUI	OUI	NON	NON

TAB. 9.6 – Résumé des énergies radiométriques considérées.

La variante $\mathsf{Q}^{\frac{1}{2}}$ correspond à un choix intermédiaire entre attribuer $\ln \lambda_i$ et λ_i comme coût moyen par échantillon et par dimension. Les variantes $\mathsf{D}^{\frac{1}{2}}$ et D changent le logarithme du déterminant qui intervient dans l'expression de l'énergie G en une racine ou en l'identité.

On vérifie aisément que les expressions qui impliquent la trace de Σ_X fournissent des énergies sur-additives, contrairement à celles qui impliquent son déterminant. Q et $\mathsf{Q}^{\frac{1}{2}}$ conduisent donc à ce que nous

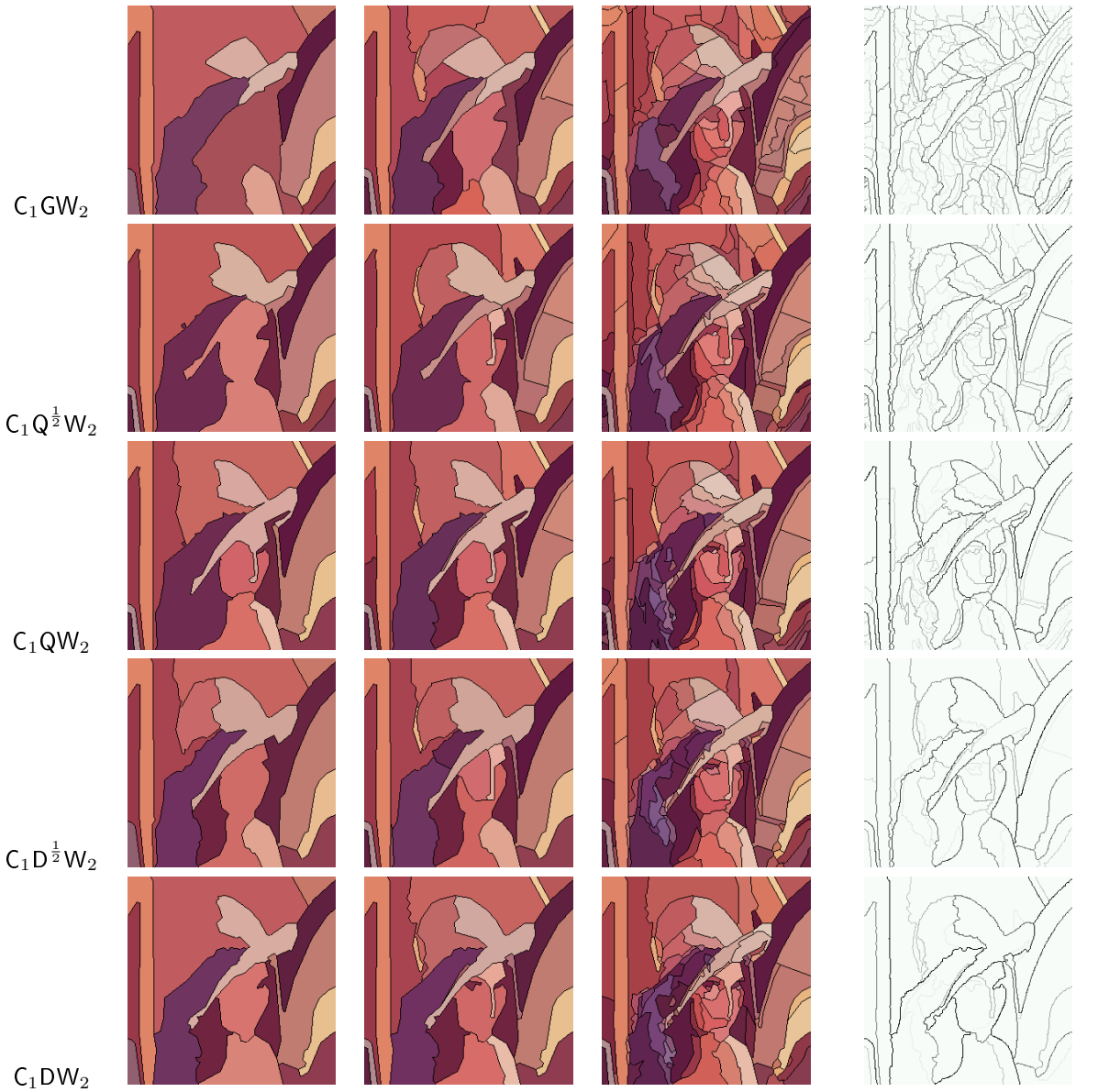
[10] Rappelons que pour que l'approximation de l'intégrale 9.10 soit valable, il faut s'assurer que les λ_i sont supérieures à q^2. q correspond en théorie au pas de quantification effectif des observations. Ainsi pour les images codées en valeurs entières - typiquement sur $\langle 0, 256\langle$ - en théorie $q = 1$. Nous venons cependant de voir que le choix de q n'avait pas d'importance pour l'escalade. En pratique, le seul problème est d'éviter les énergies de $-\infty$ que l'on obtient quand $|\Sigma_X| = 0$ i.e. quand il existe une valeur propre nulle. Typiquement pour les images en couleurs - à 3 canaux - le déterminant est nul pour tous les échantillons de moins de 4 points. Et même pire, si l'image est en réalité une image en niveaux de gris, i.e. telle que les trois composantes rouge, vert et bleu des pixels sont systématiquement égales, alors tous les échantillons sont en réalité mono-dimensionnels et conduisent à des longueurs de $-\infty$. Il est donc nécessaire de passer par le calcul des valeurs propres de Σ_X, de les normaliser par q^2 et de ramener à 1 les valeurs qui sont inférieures avant d'utiliser l'expression en $\sum \ln \lambda_i$. En pratique, nous utilisons un q très faible et non de 1 comme la théorie du codage le voudrait : il peut en effet être intéressant d'ordonner deux hypothèses de fusion de variances inférieures à 1 et non de leur attribuer une énergie nulle.

[11] Notons que l'énergie Q correspond au coût de codage selon une loi normale de variance fixe plutôt qu'optimisée pour l'échantillon.

avons appelé aux chapitres précédents des formulations multi-échelles «pures» (dont les solutions vivent sur le demi-axe positif des échelles et qui rendent équivalentes les escalades pour les deux problèmes duaux de débit/distorsion).

9.3.3 Résultats expérimentaux

La figure 9.20 montre des résultats obtenus avec ces différentes énergies sur l'image Lena. Côté géométrique, nous avons utilisé l'énergie de concavité C_1. L'initialisation est faite par l'algorithme W_2.



(a) $\sim$ 15 régions (b) $\sim$ 25 régions (c) $\sim$ 100 régions (d) Contours ($\gamma_0 = 256$)

FIG. 9.20 – Comparaison des énergies radiométriques G, $\mathsf{Q}^{\frac{1}{2}}$, Q, $\mathsf{D}^{\frac{1}{2}}$ et D sur Lena. Voir texte.

Les résultats sont assez difficiles à comparer en dehors de toute application aval. Quelques comportements généraux ressortent toutefois.

On observe tout d'abord une progression nette dans la dynamique des cartes de contours. Les ultramétriques sont de plus en plus contrastées de l'énergie G à l'énergie D.

Par ailleurs, les coupes à 100 régions mettent en évidence des comportements vis à vis des textures relativement différents. La segmentation selon G fusionne les parties texturées de la plume violette du chapeau de Lena avant les parties lisses de la grande plage grise à droite de l'image ou avant les parties lisses du fond. Inversement, l'énergie D fusionne en priorité la plage grise ou le fond. Les autres énergies ont un comportement intermédiaire. Le comportement de G s'explique par l'indépendance de l'échelle d'apparition d'une région à la valeur absolue de sa variance : seul compte le rapport entre sa variance propre et celle de ses sous-parties. Des régions peu contrastées mais comportant des sous-régions gaussiennes de paramètres très différents peuvent donc apparaître très tardivement. Inversement des régions visuellement texturées mais relativement stationnaires peuvent vivre à des échelles faibles.

Notons toutefois que, même si l'ordre dans lequel sont produites les régions est différent, de nombreuses régions se retrouvent dans toutes les analyses, à une échelle ou à une autre, avec seulement quelques différences géométriques.

À grande échelle (15 ou 25 régions) Les différences essentielles sont localisées au niveau des zones réputées difficiles de Lena : les limites ténues entre l'épaule et le visage et entre la partie supérieure du chapeau et le fond.

Finalement, il est important de rappeler que l'énergie géométrique de concavité est en grande partie responsable des unités globales qui sont délimitées.

9.3.4 Critère MDL - débit/distorsion

Voyons maintenant sur un exemple les résultats que fournissent des critères MDL.

Dans le cadre de l'inférence par codage minimal, outre les erreurs du modèle, il faut transmettre ses paramètres. Une gaussienne de $\mathbb{R}^d$ est déterminée par $\frac{d(d+3)}{2}$ paramètres : les d coordonnées de sa moyenne et $\frac{d(d+1)}{2}$ entrées indépendantes de la matrice de variance/covariance (qui est symétrique). Si l'on utilise l'expression BIC/MDL asymptotique 6.13 (p.124) du coût de codage optimal de ces paramètres pour un échantillon de taille N, on obtient alors

$$\mathcal{L}(\mathcal{N}_{\overline{X},\Sigma_X}) \simeq \frac{d(d+3)}{4} \log N \quad \text{bits.} \tag{9.16}$$

Côté géométrique, nous avons considéré deux types de descriptions de partitions : une description orientée frontières par un code de Freeman, qui attribue à une région R une complexité géométrique d'environ $2\mathsf{L}(\delta R)$ bits, et un codage polaire d'une approximation polygonale des régions, qui attribue à R une complexité de $\mathsf{P}_\sigma(\delta R)$ bits. Nous prenons ici $\sigma = \sqrt{2}/2$ pour pouvoir négliger les écarts entre les polygonisations et les frontières inter-pixellaires. La longueur est également évaluée après polygonisation à précision $\sqrt{2}/2$ afin d'éviter les anisotropies liées à la discrétisation.

La figure 9.21 montre différentes coupes des ensembles-échelle obtenus par escalade sur les quatre énergies suivantes :

(a)	$\lambda(2\mathsf{L}_{\sqrt{2}} + \mathcal{L})$	$+$	G
(b)	$\lambda(\mathsf{P}_{\sqrt{2}} + \mathcal{L})$	$+$	G
(c)	$\lambda\mathsf{P}_{\sqrt{2}}$	$+$	$(\mathcal{L} + \mathsf{G})$
(d)	$\lambda\mathsf{P}_{\sqrt{2}}$	$+$	G

où $\mathcal{L}$ représente la longueur de description 9.16 des paramètres d'une gaussienne.

Les coupes à $\lambda = 1$ pour les énergies (a) et (b) représentent donc des approximations de deux critères MDL utilisant deux langages de description géométrique différents et un schéma de description radiométrique identique. Les énergies (b) et (c) coïncident à $\lambda = 1$, toutefois les stratégies d'escalade sont différentes : (b) oppose la complexité du modèle à ses erreurs alors que (c) oppose la part géométrique de la description à sa part radiométrique. Dans le cas (c), regrouper le terme $\mathcal{L}$ avec le terme G introduit une composante sous-additive dans l'énergie d'«attache aux données», rendant moins «pure» l'énergie

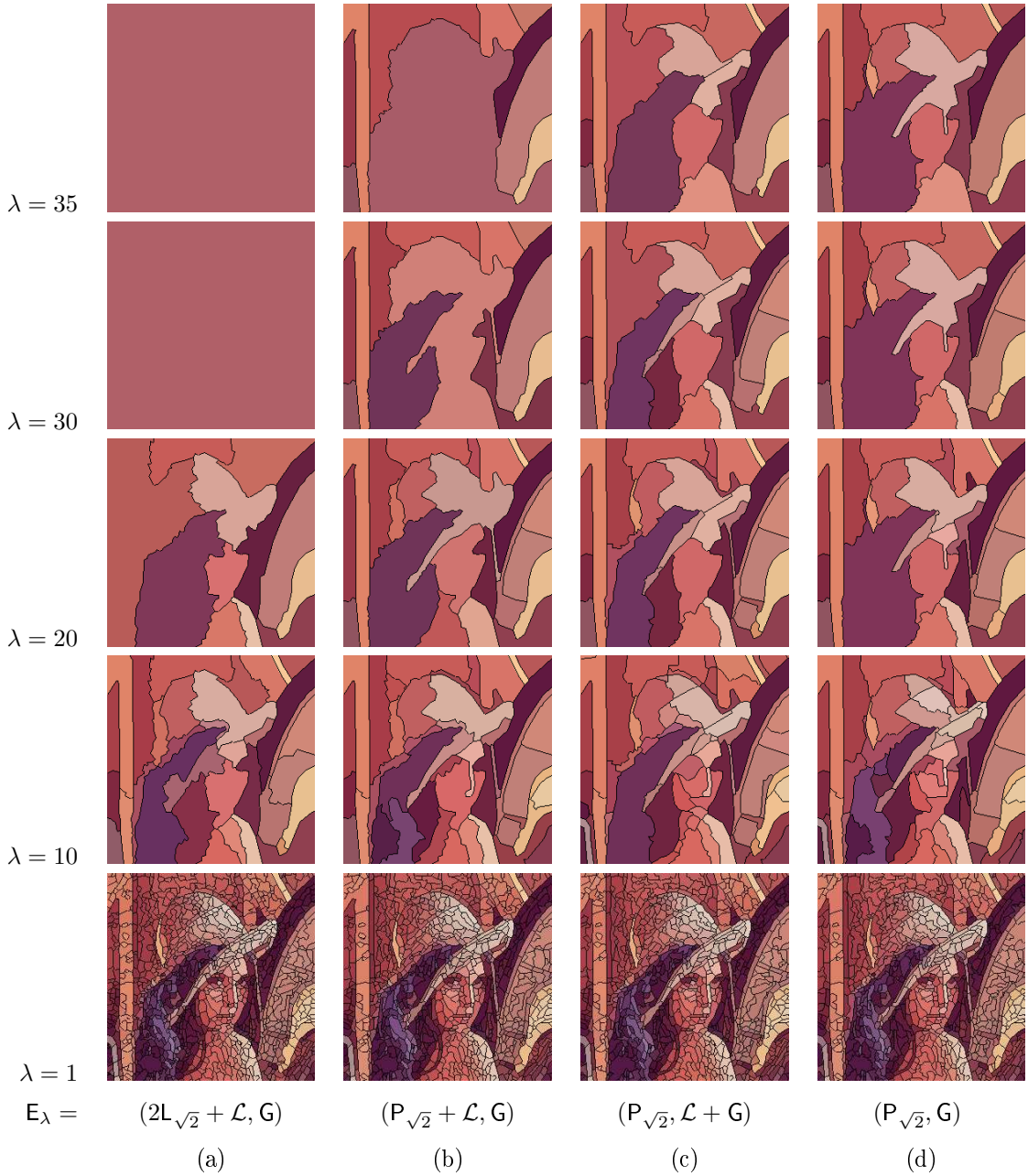


FIG. 9.21 – Escalades sur des critères de codage minimal.

multi-échelles. Pour sa part, l'énergie (d) ne correspond pas à un critère MDL : le codage des paramètres est omis.

On constate que les segmentations à $\lambda = 1$ sont très morcellées, ce qui est bien naturel : seule de petites plages de Lena sont correctement modélisées par une distribution gaussienne.

Notons à nouveau l'influence décisive à grande échelle du choix du langage de description géométrique. Une description selon P est beaucoup plus succinte en cas d'alignements qu'une description selon L, c'est pourquoi le sommet de la hiérarchie (a) apparaît avant celui de (b) (resp. $\lambda^+(I) = 22$ et 58). Cependant, à l'échelle MDL ($\lambda = 1$), les résultats sont peu différents. Autrement dit, dans un cadre MDL, introduire des langages géométriques complexes n'a de sens que si l'on introduit parallèlement des

langages radiométriques complexes, qui peuvent permettre des descriptions économiques de régions de taille significative.

Globalement, les niveaux de description utiles pour une tâche d'interprétation se situent à des échelles bien plus fortes que 1. On se trouve alors dans un cadre débit/distorsion de compression avec perte. Mais dans ce cas, le terme $\mathcal{L}$ - qui est nécessaire dans l'optique de description sans perte MDL - n'a a priori plus de sens. On observe pourtant qu'il produit des résultats étonnants aux fortes échelles (Lena groupée en une seule région pour l'énergie (b)).

9.3.5 Introduction d'un terme de gradient

Voyons finalement une façon d'introduire un terme de gradient dans l'énergie géométrique d'une partition (C) de manière à favoriser les partitions dont les frontières suivent les contours de l'image.

Considérons l'énergie de longueur L. On peut l'écrire pour une région R sous la forme $\mathsf{L}(R) = \sum_{x \in \delta R} 1$. Au lieu d'attribuer un coût constant à chaque élément de frontière on peut alors considérer un coût qui dépend du module du gradient le long de la frontière :

$$\mathsf{L}^f(R) \triangleq \sum_{x \in \delta R} f\left(||\nabla I(x)||\right) \tag{9.17}$$

où f est une fonction positive décroissante (on pourrait aussi estimer le gradient dans la direction normale à la courbe). L'énergie obtenue est bien sous-additive.

Notons $G = ||\nabla I||$. $\mathsf{L}^f(R) = \sum_{x \in \delta R} f(G(x))$ peut alors aussi s'écrire $\mathsf{L}^f(R) = \frac{\sum_{x \in \delta R} f(G(x))}{|\delta R|} |\delta R|$. L^f peut donc être vue comme une modulation de l'énergie de longueur L par un terme de gradient moyen le long de la frontière. Cette expression se généralise alors à n'importe quelle énergie géométrique E en définissant

$$\mathsf{E}^f \triangleq \frac{\sum_{x \in \delta R} f(G(x))}{|\delta R|} \times \mathsf{E}. \tag{9.18}$$

Remarquons qu'E^f n'est plus nécessairement sous-additive. On forcera alors la sous-additivité comme nous l'avons déja fait ci-dessus.

En pratique, nous avons considéré des fonctions f de la forme $f(g) = 1/g^\gamma$ (notons que si le gradient est mesuré sur les frontières d'une LPE, il n'est jamais nul).

La figure 9.22 montre qu'introduire un terme favorisant les partitions s'appuyant sur des gradients importants permet de nettement améliorer les segmentations de Lena. Ici $\gamma = 1$. L'énergie est donc invariante par transformation linéaire des valeurs de l'image.

Nous avons toutefois constaté que pour d'autres images, un terme de gradient avait tendance à dégrader les résultats en introduisant un critère de décision trop local là où certaines limites ténues doivent être rehaussées sur un critère purement régional (repenser par exemple au cas de la limite frontale de l'IRM). De même, un terme de gradient a tendance à empêcher la délimitation correcte de textures présentant d'importantes variations (e.g. en imagerie aérienne, forêts présentant des ombres marquées).

9.4 Modélisation d'un MNE par facettes planes

Franck Taillandier a exploité la méthode d'escalade pour réaliser des segmentations de Modèles Numériques d'Élévation (MNE) en facettes planes. Disons quelques mots de cette application pour conclure ce chapitre (pour plus de détails lire Taillandier, Guigues and Deriche (2003) ainsi que le futur mémoire de thèse de Franck).

Franck Taillandier s'est initialement inspiré de la méthode que nous avions précédemment développée avec Didier Boldo (2002) (cf. p.166). Les données d'entrées sont un MNE ainsi qu'une ortho-image de gradient (Boldo 2002). L'objectif est de modéliser le MNE par une surface plane par morceaux, la surface au sein de chaque région étant de la forme $z = ax + by + c$. On souhaite également que les frontières des

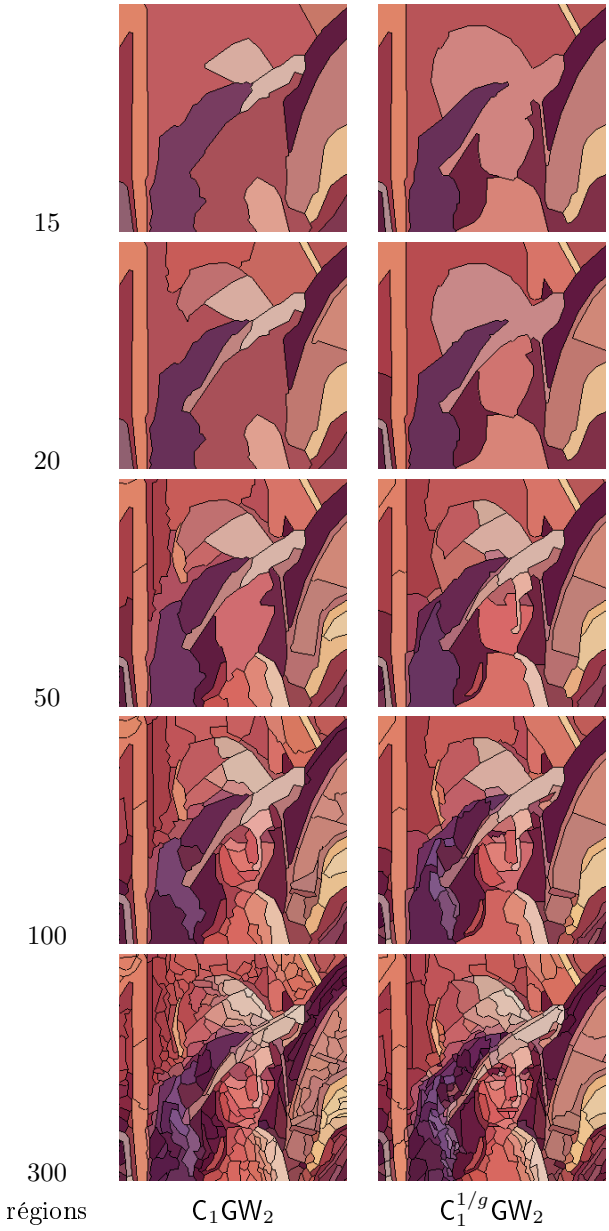
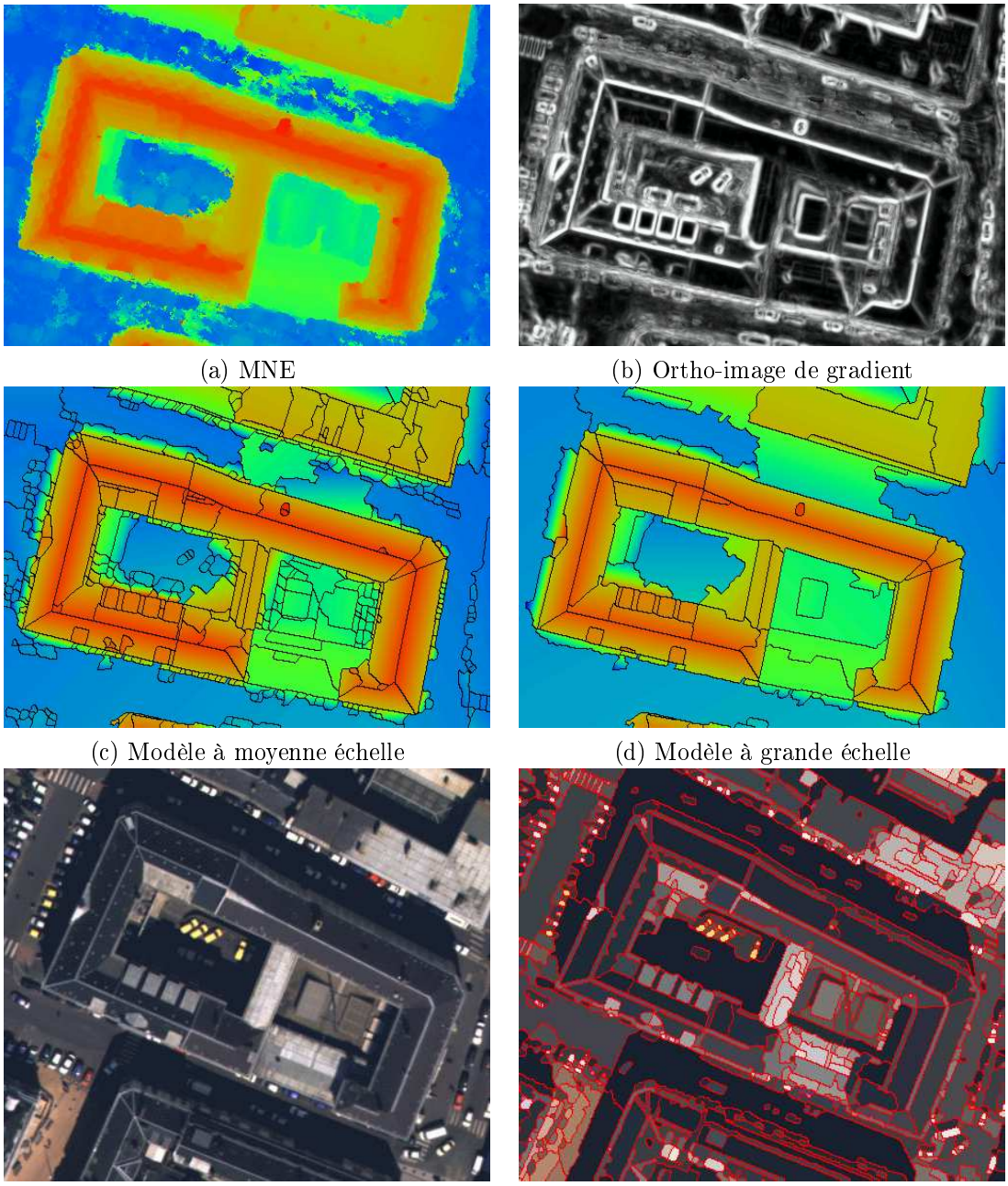


FIG. 9.22 – Introduction d'un terme de gradient dans l'énergie de la partition.

régions soient localisées sur des lignes de fort gradient afin de corriger les erreurs classiques des MNE au niveau des lignes de discontinuité (voir la figure 9.23(a)). Pour renforcer la qualité géométrique des limites, Frank Taillandier s'appuie également sur la détection de segments 3D. Les segments détectés sont réinjectés dans l'ortho-image de gradient avec une saillance importante.

Les plans de régression sont calculés par un estimateur robuste (m-estimateur en norme $L_{1,2}$). L'énergie «radiométrique» est alors une énergie de codage qui incorpore un terme en $\frac{3}{2}\log N$ pour le codage des paramètres du plan, un terme de codage des déviations en norme $L_{1,2}$ et un terme de codage des gradients internes à une région afin de favoriser les régions exemptes de discontinuités radiométriques internes. L'énergie géométrique est l'énergie de codage polaire que nous avons décrite ci-dessus. L'escalade s'appuie sur un calcul de LPE sur l'ortho-image de gradient.

La figure 9.23 montre un MNE, l'ortho-image de gradients associée ainsi que deux coupes de la



(a) MNE (b) Ortho-image de gradient

(c) Modèle à moyenne échelle (d) Modèle à grande échelle

(e) Une vue du bâtiment (perspective) (f) $\mathsf{C}_1^{1/g}\mathsf{Q}^{\frac{1}{2}}\mathsf{W}_2$ sur (e) et coupe à $\lambda = 10$

FIG. 9.23 – Segmentation d'un MNE en facettes planes.

description ensembles-échelle obtenue par cette méthode. Les couleurs représentent du bleu au rouge des altitudes croissantes. Les altitudes figurées sur les coupes correspondent aux altitudes données par le modèle calculé (on voit clairement que les images sont bien plus lisses que le MNE d'origine).

Les figures 9.23(a) et (b) présentent une image aérienne du même bâtiment ainsi qu'une coupe de l'escalade sur cette image par l'algorithme $\mathsf{C}_1^{1/g}\mathsf{Q}^{\frac{1}{2}}\mathsf{W}_2$.

On sent bien qu'en réalisant simultanément les segmentations radiométrique et altimétrique par combinaison des deux données (nous pensons ici à l'extension multi-dimensionnelle de l'escalade expliquée p.180) on doit pouvoir converger vers un résultat robuste. En particulier les régions persistantes à la fois

radiométriquement et altimétriquement ont de fortes chances de correspondre à des pans de toît.

9.5 Conclusion

Nous venons de passer en revue de nombreux choix possibles d'énergies.

Il est clairement apparu que l'énergie géométrique avait une influence décisive sur les résultats. Nous avons alors pu mettre en évidence certains choix préférentiels. Par contre, si l'énergie radiométrique a également une influence certaine, il est plus difficile de départager les différentes formes que nous avons envisagées.

Nous avons finalement une certaine préférence pour l'énergie $\mathsf{C}_\sigma \mathsf{Q}^{\frac{1}{2}}$, alliant une énergie de concavité à des déviations proportionnelles aux écarts types des distributions - intermédiaires entre un codage gaussien et des déviations quadratiques. Sa forme est remarquablement simple. En outre, il s'agit d'une énergie multi-échelles pure qui verifie le plus d'invariances parmi les énergies étudiées (déplacements, homothéties, projections pour les formes convexes et certaines déformations de formes concaves, ainsi que rotations du repère colorimétrique et transformations linéaires des radiométries).

Pour conclure, les planches 9.24 à 9.29 présentent des résultats obtenus avec cette énergie sur différentes scènes aériennes. Les coupes ont été sélectionnées manuellement (mais sans précaution particulière) pour correspondre à des niveaux de description significatifs de chaque scène.

Remarquons sur ces exemples la capacité de la méthode à correctement délimiter certaines textures complexes (le hangar en construction, le bourg de campagne, les faces de la Tour Eiffel...) et ce en dépit de l'extrême simplicité du modèle radiométrique, qui ne comporte aucune caractéristique liée à l'organisation spatiale des valeurs dans une région. Ce comportement a trois origines.

En premier lieu, si deux régions présentent des textures identiques, elles possèdent a fortiori les mêmes statistiques gaussiennes. L'ordre de fusion des régions sera donc le bon si les statistiques gaussiennes suffisent à discriminer deux textures adjacentes, ce qui est le cas dans de nombreuses situations réelles. Pour la segmentation de textures plus générales, nous pensons qu'une simple modélisation gaussienne dans l'espace de quelques caractéristiques texturales bien choisies devrait suffir à discriminer de très nombreuses textures.

En deuxième lieu, certaines régions peuvent apparaître en une pièce par effet de complémentarité. Globalement, le mécanisme d'escalade est un mécanisme de compétition entre différentes fusions possibles. Une régions d'aspect très disparate peut être groupée comme un tout si elle s'inscrit sur un fond très cohérent.

En troisième lieu, les fusions sont également dirigées par la géométrie de l'union, ce qui pointe à nouveau l'importance du facteur de forme en segmentation.

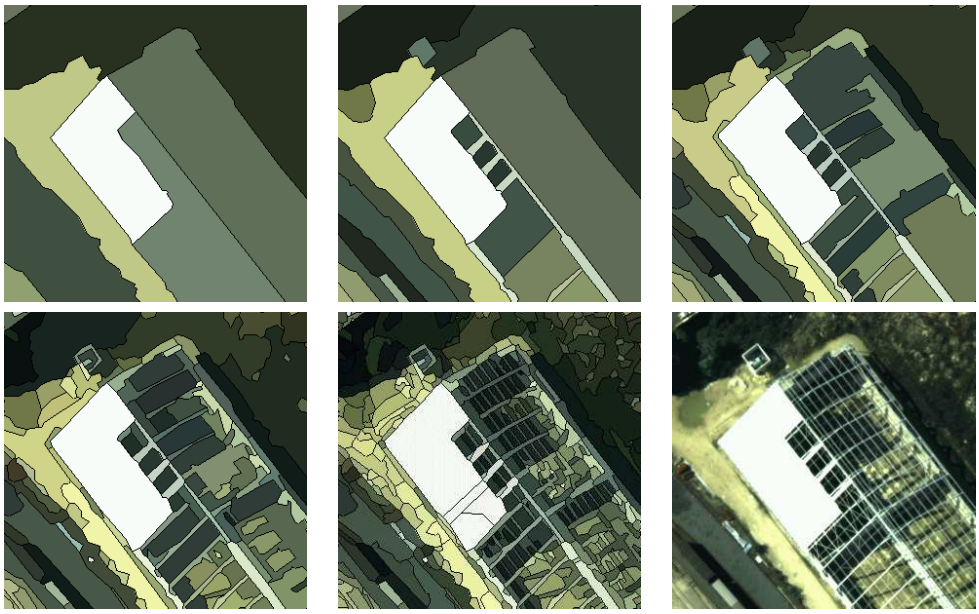


FIG. 9.24 – Hangar en construction.

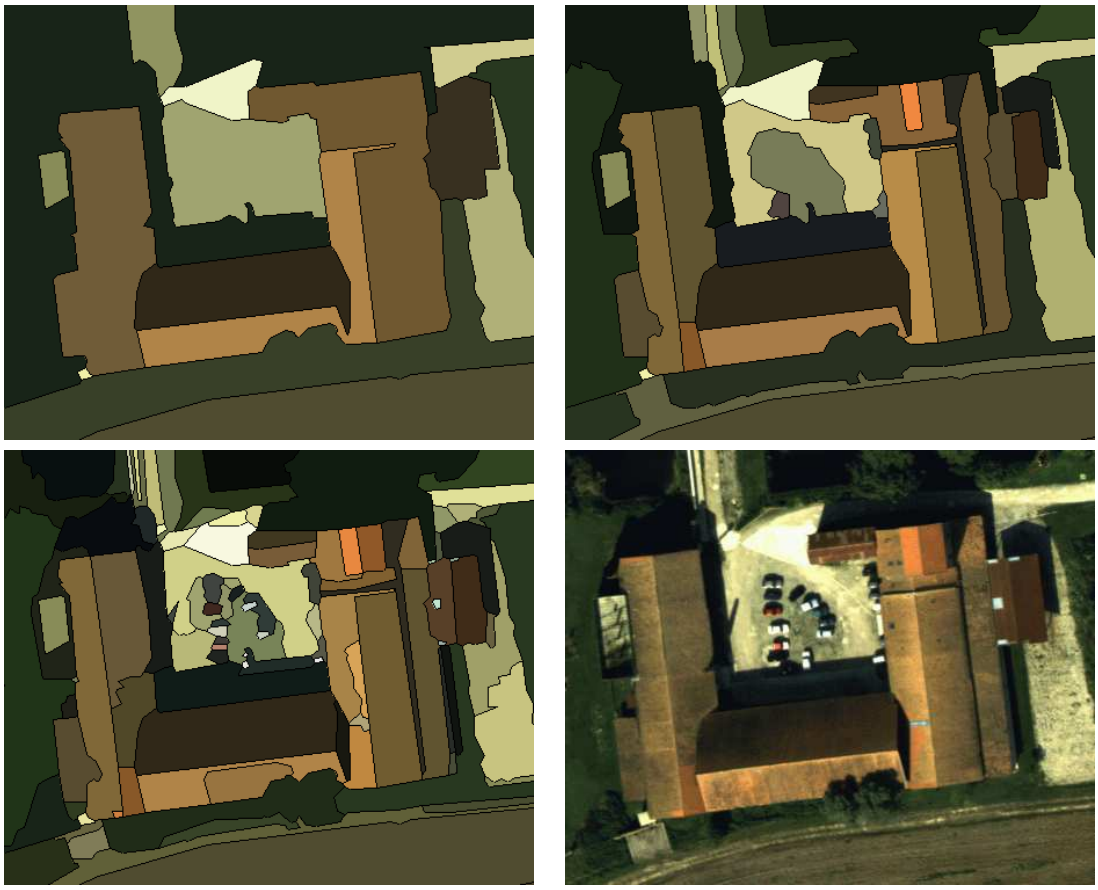


FIG. 9.25 – Une ferme.

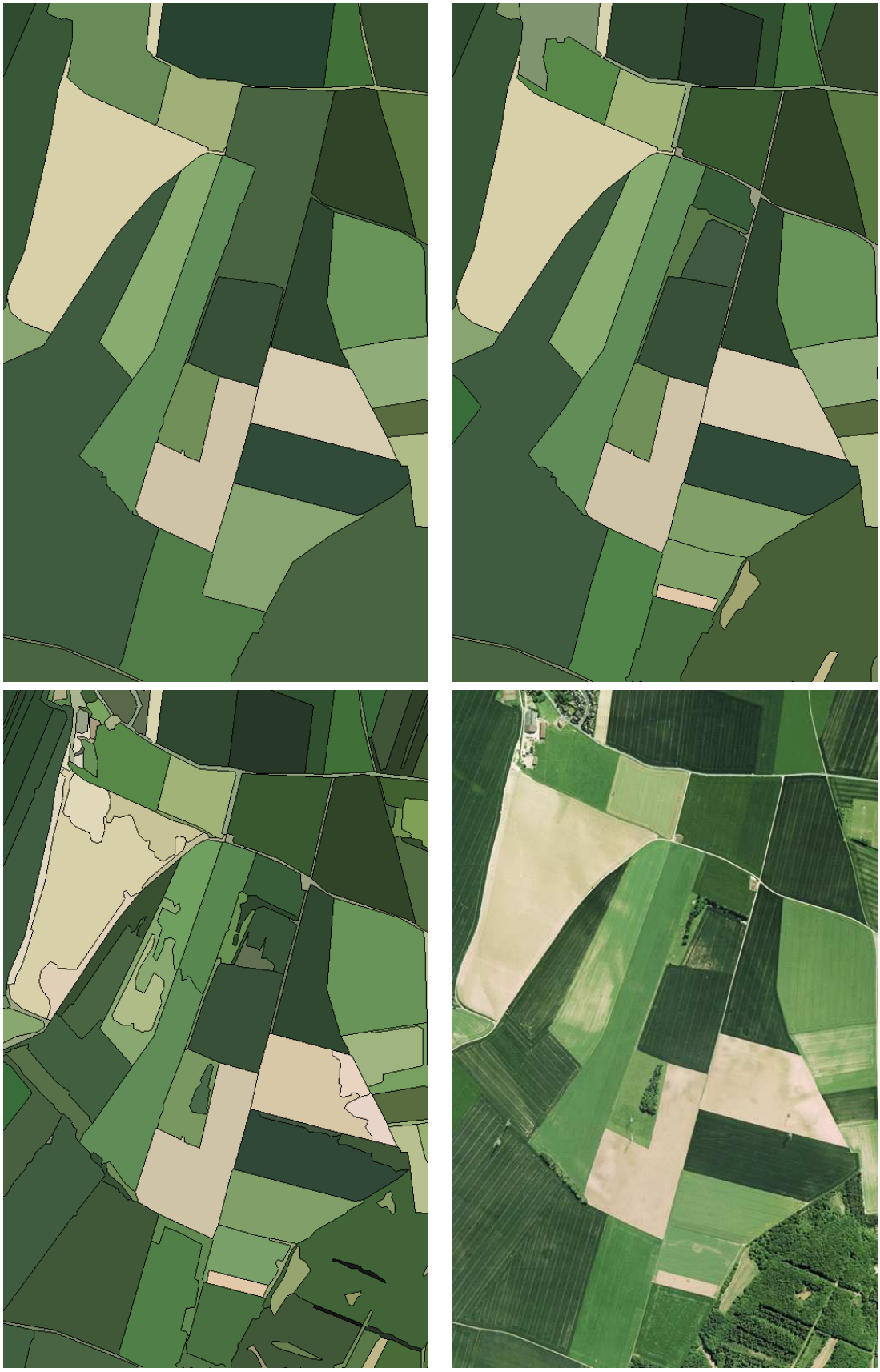


FIG. 9.26 – Paysage de Beauce.

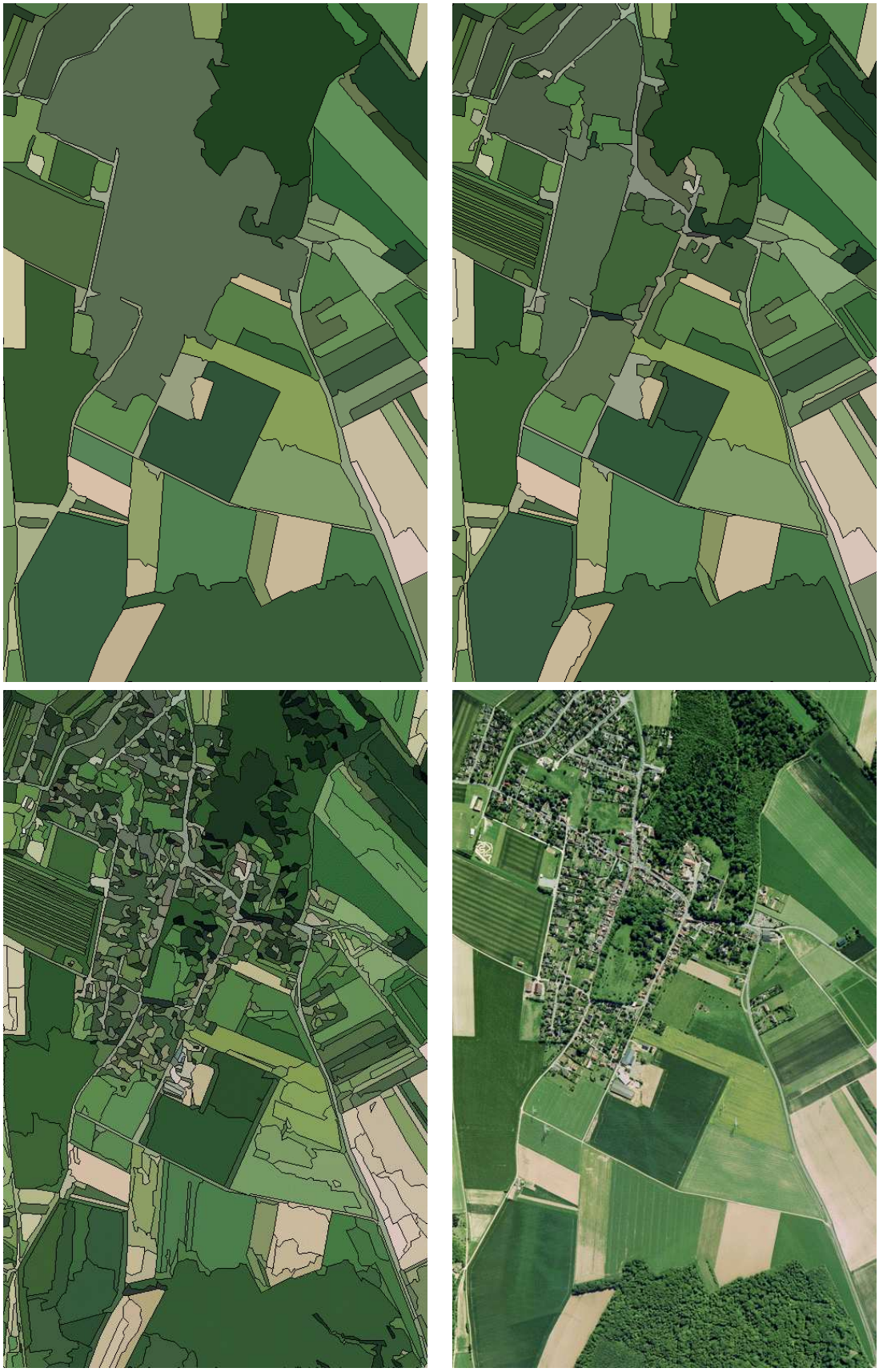


FIG. 9.27 – Un bourg de campagne.

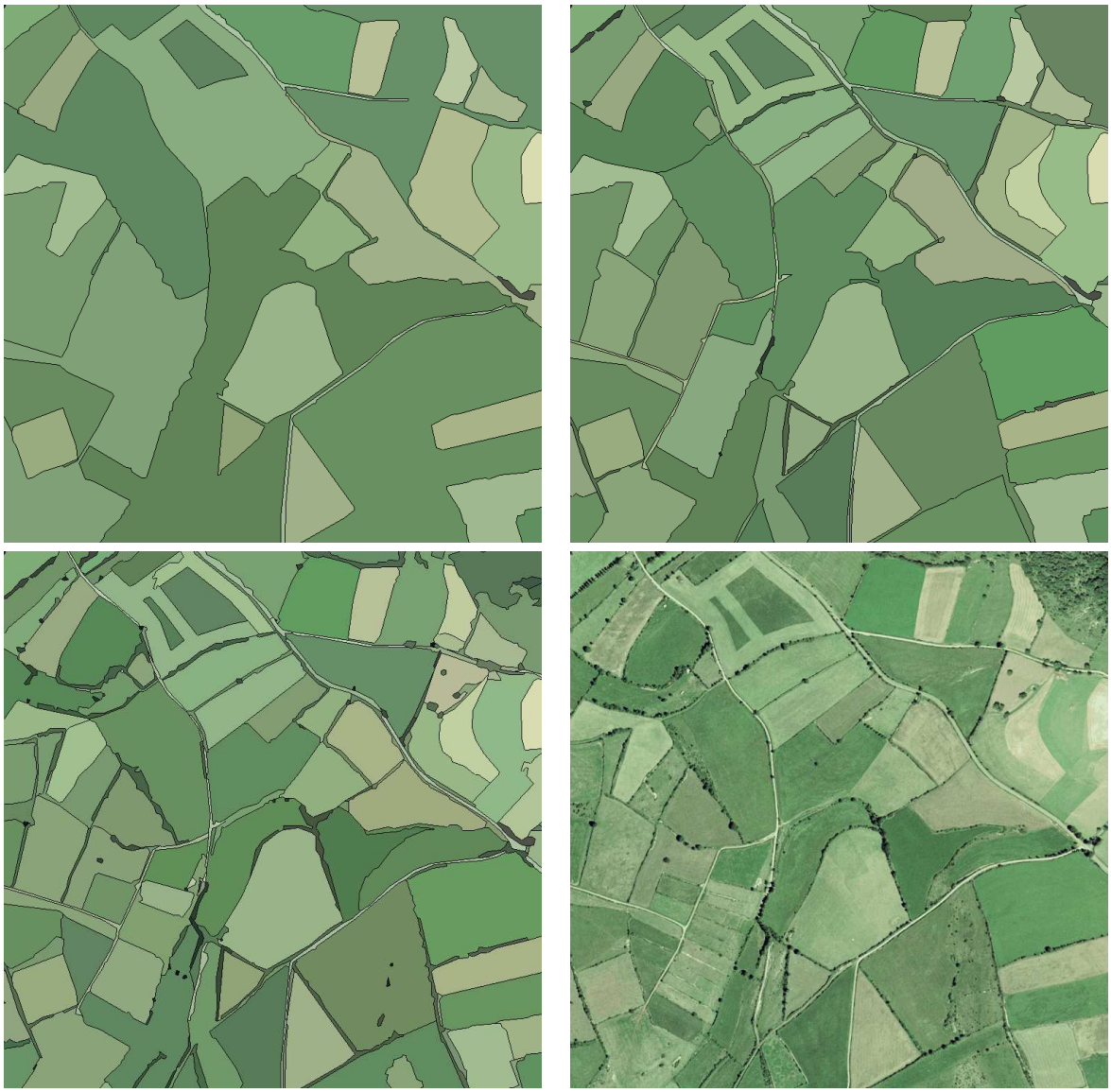


FIG. 9.28 – Paysage rural.

FIG. 9.29 – La tour Eiffel.

Chapitre 10

Conclusion et perspectives

10.1 Conclusion

10.1.1 Ensembles et échelles

L'échelle des valeurs, l'échelle des gris, l'échelle des tailles... Une échelle est un axe monodimensionnel qui permet d'*ordonner totalement* certains phénomènes.

Quand on s'intéresse à des phénomènes spatiaux, l'idée d'échelle est associée à une autre notion : celle de phénomènes imbriqués les uns dans les autres, *hiérarchisés*.

Une image possède deux «dimensions» : une dimension spatiale et une dimension de valeur. De même, un objet visuel possède deux caractéristiques principales : une forme, une silhouette, et un aspect (clarté, couleur, texture...). «La Terre est bleue comme une orange» écrivait ainsi Paul Éluard, jouant sur la confusion des deux caractéristiques. La silhouette est un élément du support de l'image, de l'espace. L'aspect, lui, est lié aux valeurs de l'image.

Une première idée qui a été développée dans ce mémoire est qu'il est possible, et souhaitable, de concilier les deux idées d'échelle qui sont associées aux deux dimensions d'une image : il est possible de développer des analyses des valeurs d'une image qui induisent des familles d'ensembles hiérarchisés. C'est le modèle de **cocon dur**.

L'analyse produite est alors un volume, une séquence continue de partitions emboîtées de l'espace, séquence qui peut être représentée exactement mais sous forme implicite, comme *la famille des coupes naturelles d'une hiérarchie indicée* (ou de manière équivalente d'un espace ultramétrique). Dans le cadre de l'analyse d'images, nous avons proposé d'appeler une telle structure une **représentation ensembles-échelle**, car elle constitue le pendant structurel d'une représentation scale-space. L'espace est structuré sous forme d'une hiérarchie d'ensembles et l'indice, l'échelle, confère un ordre total à ces ensembles.

Les formes, les silhouettes, émergent avant tout de l'existence de variations spatiales dans l'aspect d'une image : c'est en premier lieu la mesure qui induit un morcellement de l'espace. Mais inversement, les formes ont une influence sur notre vision de la valeur : nous régularisons visuellement les silhouettes, éliminons le bruit, complétons des contours manquants, jusqu'à imaginer des variations totalement inexistantes (formes illusoires). Le facteur géométrique a un rôle central dans notre perception des formes.

Le modèle de cocon est un modèle multi-échelles mono-critère, qui repose sur une unique mesure, une dissemblance, liée à l'aspect. L'échelle est alors directement un seuil sur ce critère. De son côté, le deuxième modèle proposé s'appuie sur la compétition de deux critères : un critère de simplicité de description et un critère de fidélité de la description aux données. L'approche s'appuie sur une idée extrêmement générale et intuitive : simplicité et fidélité sont deux qualités antagonistes pour un modèle. Faire une description suffisamment concise impose de caricaturer ou encore, plus on souhaite modéliser

fidèlement un phénomène, plus le modèle employé devra être complexe.

La formulation de la segmentation comme problème d'optimisation dérive alors d'un principe naturel de modélisation, **le principe de parcimonie**, selon lequel, à complexités égales, nous devons retenir le modèle le plus fidèle ou à précisions égales, retenir le modèle le plus simple. Il s'agit du principe central de la théorie de la compression avec pertes de Shannon aussi est-il est naturel d'aboutir à deux problèmes duaux d'optimisation sous contrainte qui sont formellement identiques à des problèmes de débit/distorsion opérationnels. La formulation lagrangienne est alors le pont entre les solutions de ces deux problèmes, et dans notre cadre le pont entre la complexité des formes, i.e. la part *spatiale* du modèle, et la fidélité du modèle aux données, aux *valeurs* de l'image. Avec l'expression Lagrangienne, nous retrouvons les formulations énergétiques classiques du problème de partitionnement : formulation variationnelles ou bayesiennes. Dans ces théories, le fait de devoir introduire des énergies à deux termes antagonistes est justifié par le caractère mathématiquement mal posé des formulations «naïves» du problème de modélisation, ne prenant en compte qu'une distance modèle-données. Il y a donc un rapport très étroit entre les problèmes de segmentation et les problèmes de compression orientée régions.

Nous avons alors montré que si l'on impose une structure hiérarchique aux régions présentes dans les solutions - ce qui est, insistons sur ce point, une contrainte plus faible que d'imposer une structure pyramidale sur les partitions - et si l'on utilise des énergies séparables dont le terme de complexité va en croissant avec la finesse des descriptions spatiales alors on aboutit effectivement à des analyses multi-échelles. L'hypothèse de croissance de l'énergie de régularisation est parfaitement naturelle. D'un point de vue descriptif, elle revient à stipuler qu'une description est d'autant plus complexe qu'elle est détaillée. D'un point de vue probabiliste, elle revient à stipuler qu'a priori, avant de voir l'image, l'hypothèse la plus probable est que l'image est *vide*, et que pour infirmer cette hypothèse, i.e. percevoir des formes, morceller l'espace, il faut trouver un motif valable dans l'image. A priori, une partition en deux classes est donc moins probable que la partition en une seule classe. Appliquée récursivement - indépendamment pour chaque nouveau détail considéré comme une image isolée - cette idée conduit à une probabilité a priori décroissante avec la finesse des partitions de l'espace.

Un seul paramètre, le paramètre de Lagrange, appelé paramètre d'échelle dans nos modèles, relie les solutions des deux problèmes duaux. C'est pourquoi, pour simultanément optimiser ces deux problèmes, il faut évoluer sur l'axe d'échelle. En termes techniques, il faut progresser le long de la courbe de débit/distorsion du problème. C'est le principe de l'algorithme d'**escalade**. Un unique paramètre mono-dimensionnel permet alors à la fois de «rendre clair l'obscur et simple le complexe»[1], c'est-à-dire de naviguer selon les deux dimensions fondamentales d'une image : l'espace et la mesure. Diminuer l'«échelle» a pour effet de simultanément rafiner les descriptions spatiales et radiométriques, de rentrer dans des détails plus fins et plus ténus. Autrement dit, dans un cadre hiérarchique, sous une hypothèse sur la loi a priori parfaitement naturelle, l'échelle des tailles, l'échelle de «complexité» de forme et celle de «complexité» d'aspect se confondent.

10.1.2 Développer la dimension d'échelle

Quel est l'intérêt de développer une dimension d'échelle ?

1) Obtenir des méthodes robustes de segmentation de bas niveau – Le deuxième chapitre du mémoire a argumenté que, d'un point de vue systémique, une approche multi-échelles de l'analyse de bas niveau en régions était *nécessaire* pour espérer pouvoir aboutir à une indépendance complète entre bas et haut niveau d'analyse visuelle. D'un point de vue pratique, nous avons vu qu'une approche multi-échelles

[1] Al-Kwharizmi - mathématicien Bagdadi dont le nom est à l'origine du mot algorithme - a inventé Al-jabr (littéralement le «raboutement» et qui a donné le mot algèbre) au début du $IX^{\text{ème}}$ siècle, à savoir la première méthode systématique - ou mécanique, algorithmique - de résolution d'équations. L'idée fondamentale d'Al-Kwharizmzi a été de *donner un nom* à l'inconnue et de la manipuler comme s'il s'agissait d'une quantité connue. Les manipulations formelles ont alors pour objectif d'isoler l'inconnue, dont la valeur est alors «révélée». L'obscur est rendu clair et le complexe simple. Nous aimons comparer la segmentation à un mécanisme de «raboutement» des pixels d'une image dont l'objectif est de révéler l'inconnu, mettre en évidence les formes de l'image. Selon l'«échelle», des formes différentes sont rendues explicites, claires, simples...

permettait d'aboutir à des méthodes **sans paramètre** et **robustes**, i.e. vérifiant certaines propriétés d'invariance fondamentales.

L'algorithme d'escalade - qui correspond au modèle le plus riche - est ainsi totalement exempt de paramètre une fois un couple d'énergies antagonistes spécifié. Par ailleurs, une propriété fondamentale de cet algorithme est qu'il produit des descriptions ensembles-échelle dont la *structure* est invariante par transformations linéaires des énergies. L'*échelle*, elle, est covariante. Ce dernier point signifie que c'est la description complète, prise comme un tout, qui est conservée, mais que ses coupes individuelles ne le sont pas. Nous retrouvons ici Koenderink (1984), qui mettait en avant l'idée que le scale-space doit être considéré comme un tout, une famille d'images reliées par une équation d'évolution et non comme une collection d'images.

Concrètement, cette covariance des ensembles-échelle obtenus par escalade signifie que l'algorithme permet véritablement de **supprimer le paramètre d'échelle du bas niveau d'analyse**. En effet, régler l'échelle a priori - lors de la spécification des énergies - ou a posteriori - par coupe dans la solution obtenue - est équivalent. La méthode fait donc de l'échelle un paramètre de réglage au sens fort, un potentiomètre mis à disposition de l'utilisateur (homme ou machine) pour naviguer dans la «structure» de l'image (telle qu'elle est perçue par le modèle énergétique).

Nous avons également vu que l'utilisation d'énergies apropriées permettait de conférer à la description finale des propriétés d'invariance à différentes transformations des données d'entrée et une robustesse au bruit ou à des changements de résolution.

Ces différentes qualités permettent d'envisager une analyse ensemble-échelle comme un *frontal visuel*, une interface permettant à des mécanismes ultérieurs d'appréhender l'image à plus haut niveau que le niveau des perceptions brutes.

2) Intérêt algorithmique du plongement en échelle – D'un point de vue mathématique et algorithmique, plonger le problème de partitionnement dans un cadre plus large, loin de le complexifier, permet d'en trouver des méthodes d'approximation efficaces.

Envisager un objet de façon implicite, comme une coupe d'un objet de plus grande dimension a déja montré sa fécondité en analyse d'images : pensons à la théorie des ensembles de niveau d'Osher et Setian (1988). Plusieurs méthodes d'optimisation globale (GNC ou continuation, recuit) s'appuient sur une idée apparentée, à savoir plonger le problème d'origine dans une famille de problèmes, par l'introduction d'une une variable supplémentaire (température...), et envisager la solution du problème d'origine comme une limite. Le principe de programmation dynamique de Richard Bellman s'appuie également sur une idée similaire, à savoir le plongement dans une famille de problèmes partiels qui, combinés, fournissent la solution du problème d'origine.

L'échelle induit un plongement naturel, qui est intimement lié à l'idée de causalité. D'un point de vue structurel, la décomposition d'une hiérarchie en hiérarchies partielles permet de minimiser exactement une énergie séparable, par composition ascendante de résolutions partielles sur des sous-parties de l'espace. Inversement, le plongement en échelle du problème de partitionnement optimal permet d'engendrer une structure remarquable.

Donc, envisager la famille de solutions comme un continuum permet non seulement d'obtenir une représentation riche mais conduit également à un principe d'optimisation qui fournit de bonnes solutions *mono-échelle*. En particulier, l'escalade permet d'atteindre «continuement» de bonnes solutions pour des énergies très fortement régularisées (qui sont évidemment les plus difficiles à optimiser directement). En pratique, la méthode parvient à délimiter des unités globales complexes, y compris avec des énergies extrêmement simples.

3) Maximalité locale et persistance en échelle

Décomposer un phénomène «cohérent» en parties est inutile et tout morcellement qui peut être proposé est nécessairement arbitraire. Cette idée soutend un principe central en segmentation, le **principe de**

maximalité : on doit rechercher les plus grandes parties cohérentes pour le critère qui nous intéresse. Nous avons vu avec le modèle de cocon que si l'on exprime la «cohérence» de manière *relative* - afin de supprimer un seuil arbitraire - on aboutit alors à un principe de *maximalité locale* plutôt que globale. Des zones cohérentes apparaissent alors à différentes échelles.

De même, nous avons montré que l'optimalité d'une coupe minimale dans une hiérarchie relevait d'un principe de maximalité : les régions globalement optimales sont les plus grandes régions partiellement optimales, i.e. optimales vis-à-vis de leurs sous-structures.

Dans les deux cas, le critère de sélection des unités pertinentes est alors un **ratio** : un ratio entre contraste externe et interne ou un ratio entre échelle de disparition et d'apparition (persistance). Un objet est pertinent pour l'analyse si ce ratio est supérieur à 1 et d'autant plus pertinent que ce ratio est grand. Le modèle de cocons part d'un critère relatif et aboutit à des hiérarchies. Le modèle énergétique, lui, part d'une contrainte hiérarchique et parvient à un critère d'optimalité qui prend la forme d'un ratio.

Il y a alors continuité entre ne jamais être solution et exister sur un court laps d'échelles. Et plus le temps de vie en échelle est important, plus la forme est stable : la persistance s'impose donc comme un critère de saillance de forme. Par ailleurs, la mesure est invariante d'échelle.

10.2 Perspectives

10.2.1 Perspectives théoriques

1) Maximisation de la co-énergie totale et couplage structurel

Nous sommes parvenus à la page 200 à une formulation du problème de segmentation multi-échelles comme un problème de recherche d'une hiérarchie de co-énergie totale maximale pour une énergie multi-échelles donnée.

Il s'agit d'un problème ouvert. Nous avons montré que sa solution n'était pas triviale, i.e. n'était pas la hiérarchie creuse, mais nous ne savons pas si elle est unique. Nous avons montré qu'une solution était invariante d'échelle et que l'escalade pure se présentait comme une stratégie d'approximation gloutonne localement optimale, consistant à suivre localement la plus grande pente dans l'espace débit/distorsion du problème.

Revenons brièvement sur le problème proposé afin d'expliquer en quoi nous pensons qu'une résolution plus globale pourrait s'avérer pertinente.

Le problème consiste à trouver une description multi-échelles ensembliste des données, sous la forme d'une description ensembles-échelle. La manière naturelle de relier cette description volumique à une image plane est alors de considérer ses coupes naturelles et de stipuler que chaque coupe doit correspondre à une description à échelle donnée (contrainte sur la «complexité» de la description) et être la plus «fidèle» possible à l'image à cette échelle (minimisation de la distance modèle-image). Une manière naturelle de requérir cela à toutes les échelles est alors de minimiser le lagrangien de cette famille de problèmes dans sa globalité, i.e. de minimiser sont intégrale en échelle (le fait que le problème ne soit bien posé qu'en terme de maximisation d'une quantité complémentaire est un détail technique).

Nous avons alors montré que si i) les énergies de complexité et de fidélité sont *séparables* et ii) au moins une des deux énergies est *monotone* dans le treillis des partitions, alors le problème se réduit à la recherche d'une hiérarchie de parties. En effet, les coupes minimales de la hiérarchie sont monotones et définissent sans ambiguïté l'axe des échelles de l'analyse, i.e. l'indice. En outre, l'algorithme de programmation dynamique fonctionnelle permet de calculer cet indice et l'intégrale de co-énergie de cette hiérarchie. L'ensemble du problème peut donc être posé comme une un problème de recherche d'une *hiérarchie optimale* (mais le résultat final est une description ensembles-échelles, i.e. une hiérarchie indicée).

La condition ii) de monotonie d'une des deux énergies est une condition génératrice, qui assure que l'on obtient bien une description multi-échelles. La seule limite forte du modèle est donc la condition i), à savoir la séparabilité des énergies.

Cette condition signifie que le modèle ne permet pas de prendre en compte des relations *entre régions d'une partition*, comme par exemple favoriser l'alignement de formes. Autrement dit, le modèle énergétique *mono-échelle* ne permet pas de modéliser des dépendances contextuelles ou encore de favoriser la formation de Gestalts supra-régionales.

Toutefois, le fait de rechercher une famille de partitions *monotones* qui soit *globalement* d'énergie minimale pourrait permettre réintroduire certaines formes de relations contextuelles, par un phénomène de **couplage structurel** entre les solutions aux différentes échelles. Tentons d'expliquer ce phénomène.

Notons en premier lieu que si l'on n'impose aucune contrainte sur la *structure* de la famille de partitions recherchée, alors la famille de partitions qui minimise l'intégrale du Lagrangien est tout simplement l'ensemble des minima ponctuels du Lagrangien pour chaque valeur de λ : si l'on minimise en chaque point, on minimise l'intégrale. Cependant, contraindre la famille complète de solutions à constituer une séquence de partitions *monotone* induit un couplage entre échelles. En effet, comme nous l'avons vu, en règle générale les minima d'une énergie affine ne sont pas emboîtés. Pour qu'ils le deviennent, il faut donc que certaines partitions s'écartent de la solution optimale à échelle donnée, prise indépendamment des autres.

Considérons alors l'image (a) de la figure 10.1. Les deux partitions (b) et (c) sont deux solutions indépendantes plausibles pour deux échelles $\lambda_2 > \lambda_1$. La solution à λ_2 peut être obtenue avec une énergie de complexité de forme qui privilégie les objets convexes : inclure la petite portion de fond dans la forme globale diminue peu l'attache aux données mais régularise beaucoup la forme. À λ_2 assez fort, cette solution est donc plausible. Par contre, quand les régions A et B sont divisées, exclure la portion de fond de l'ensemble A conserve la convexité et diminue peu l'attache aux donnée. Cette solution est donc plausible à une échelle λ_1. Ces deux solutions indépendantes ne sont cependant pas emboîtées.

Si l'on force les solutions à s'emboîter, il est alors possible que l'on obtienne la séquence de partitions (d)-(f). On remarque alors qu'à l'échelle λ_1, *tout se passe comme si B avait une influence* sur A, poussant le bord supérieur de A à s'aligner le sien. Diminuons alors encore l'échelle et demandons-nous comment a pu être obtenue A. La solution de moindre énergie est vraisemblablement celle décrite par la partition à λ_0 de la figure (f) : la portion de fond se détache de A. Il se crée donc une région *purement virtuelle*.

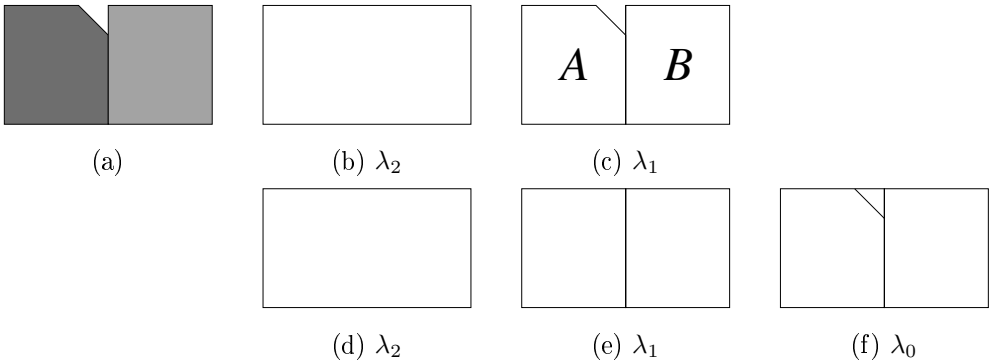


FIG. 10.1 – Le couplage d'échelle peut-il induire des influences contextuelles et des phénomènes de formes illusoires ?.

Le phénomène est encore plus frappant si l'on autorise les régions non connexes. Considérons par exemple l'image 10.2. À grande échelle, modéliser l'ensemble des taches noires comme un alignement de *disques identiques* est certainement beaucoup plus économique que de décrire chaque tache noire comme une forme différente. Si le modèle de complexité géométrique tient compte de cette régularité alors, à petite échelle, chaque tache noire individuelle peut être vue comme un disque parfait. En outre, à grande échelle, il peut être plus intéressant de voir l'ensemble comme une séquence ininterrompue, ce qui produit à faible échelle l'apparition d'une tache inexistante.

FIG. 10.2 – Un alignement de taches.

Il est donc possible que le couplage d'échelles permettre de modéliser l'idée centrale de la théorie de la Gestalt, à savoir l'idée que *la perception du tout a une influence sur celle de la partie*, à tel point qu'une organisation globale peut faire émerger des formes illusoires.

Tout ceci est relativement intuitif et demanderait à être démontré rigoureusement ou vérifié expérimentalement. Notons à ce sujet que l'approximation de la solution par escalade (même pure) ne peut pas faire apparaître ce type de phénomène - s'il existe - car l'optimisation est uniquement locale et gloutonne. Pour que le couplage puisse apparaître, il faut minimiser l'intégrale à plus longue échéance (sur plusieurs pas de complétion).

En conclusion, le fait que les minima du problème énergétique ne s'agencent pas spontanément de façon multi-échelles peut s'avérer être une aubaine : leur imposer cette contrainte supplémentaire peut permettre d'enrichir le modèle, ce qui nous ramène à l'idée qu'un mariage entre approches énergétiques et structurelles du problème de segmentation peut s'avérer prolifique.

2) Extension à des vecteurs d'échelle et à des structures plus riches

En dehors de la résolution globale du problème de recherche d'une hiérarchie optimale, mentionnons deux autres pistes de recherche qui nous semblent intéressantes :

• *Extension multi-dimensionnelle*

Nous avons vu que la méthode de programmation dynamique fonctionnelle s'étendait naturellement à des énergies à plusieurs paramètres du moment que chaque énergie de régularisation est monotone. Il reste à voir comment on peut étendre le reste de la démarche au cas multi-dimensionnel.

• *Extension à des structures plus riches que des hiérarchies*

Nous pensons ici à une extension à des structures pyramidales au sens du clustering, c'est-à-dire aux hypergraphes d'intervalles, qui autorisent des recouvrements limités entre régions. En particulier, si elle est possible, une extension de l'approche aux 2-3-hiérarchies de Bertrand (2002) pourrait s'avérer pertinente : la combinatoire est toujours linéaire et chaque région peut en recouvrir une autre. La méthode de programmation dynamique n'est cependant pas directement extensible.

10.2.2 Perspectives applicatives

Citons finalement quelques perspectives applicatives qui nous paraissent intéressantes.

1) Reconnaissance de formes, segmentation de haut niveau –

Une partie de notre argumentaire en faveur d'une approche multi-échelles de la segmentation repose sur l'idée de différer l'introduction de connaissances spécifiques à une étape aval. Nous pensons que les analyses ensembles-échelle proposées peuvent effectivement servir de socle à des problèmes variés d'interprétation d'image.

Le point important est que comme une hiérarchie possède un nombre raisonnable d'ensemble, des critères élaborés de qualification/reconnaissance de formes peuvent être envisagés. On peut alors fiablement sélectionner *a posteriori*, sur des critères de *haut niveau*, des parties ou des partitions pertinentes pour une application donnée. Il est possible de sélectionner des coupes horizontales, à échelle fixe, mais aussi sélectionner des coupes irrégulières, ou simplement des régions spécifiques, éventuellement emboîtées.

Plus spécifiquement, nos expériences en segmentation d'images aériennes nous ont montré que les unités d'occupation du sol qui intéressent un cartographe apparaissaient quasi-systématiquement dans les hiérarchies, mais à des échelles variables. Roger Trias Sanz travaille actuellement sur l'exploitation

des analyses ensembles-échelles dans le cadre de sa thèse au MATIS sur l'interprétation de l'occupation du sol guidée par le parcellaire cadastral.

2) Segmentation interactive –

Meyer (1999b) Zanoguera, Marcotegui, et Meyer (1999) et Zanoguera (2001) ont montré que l'on pouvait réaliser des systèmes de segmentation interactive efficaces en s'appuyant sur des descriptions multi-échelles. L'ensemble des outils qu'ils proposent peut reposer sur une analyse ensembles-échelle.

Un logiciel de saisie semi-automatique des zones de végétation à partir de segmentations par escalade de la BDOrtho® est en cours de développement à l'IGN. Il doit servir à constituer, d'ici à 2006, la couche végétation du futur Référentiel à Grande Échelle français.

3) Segmentation d'images variées –

La formulation énergétique proposée est relativement générale et permet donc d'élaborer des méthodes de segmentation adaptées à des types d'images différentes.

Outre la segmentation d'images classiques en plages constantes, nous avons donné un exemple d'exploitation pour la modélisation de MNE en facette planes (thèse de Franck Taillandier).

Dans le cadre d'une thèse EDF/IGN, Mathieu Deveau explore actuellement le problème de segmentation multi-échelles conjointe de données laser (distance et intensité) et d'images optiques pour la modélisation géométrique de sites industriels.

4) Stéréovision –

Différents auteurs ont proposé des méthodes de reconstruction stéréocopique s'appuyant sur l'appariement de régions. Les différences de niveau de description qui apparaissent dans deux segmentations «plates» de deux images prises de points de vue différents posent cependant d'important problèmes. Pour résoudre ces problèmes, des processus d'appariement hiérarchique «coarse to fine» ont été développés au début des années 90 dans le cadre du projet SYNTIM de l'INRIA (Vinet 1991; Randriamasy 1992). Les descriptions multi-échelles proposées peuvent certainement permettre d'envisager le problème sous un angle nouveau (e.g. comme une mise en correspondance globale de deux structures hiérarchiques par couplage de poids maximal respectant la contrainte d'inclusion).

Nesrine Shehata travaille sur ce sujet au laboratoire MATIS, dans le cadre d'une thèse CNES/IGN portant sur l'exploitation d'images satellitales à haute résolution.

5) Indexation d'image - élaboration de descripteurs –

Les descriptions multi-échelles proposées peuvent s'avérer intéressantes pour les problèmes d'indexation de base d'images, de recherche par le contenu, etc. En particulier, leurs propriétés d'invariance ou de robustesse à certaines transformations les rendent particulièrement pertinentes dans un tel cadre. Par exemple, une question intriguante est : quelle information sur la structure d'une image est véhiculée par la distribution des persistances en fonction de l'échelle (diagramme $(\frac{\lambda^+(x)}{\lambda^+(I)}, \pi(x))$) ? Dans le même ordre d'idée, certains auteurs ont proposé des index s'appuyant sur l'agencement spatial relatif des régions d'une segmentation (RAG attribué par des attributs de couleur, clarté ou forme des régions). De tels index sont bien évidemment très sensibles au niveau de segmentation considéré. Quelle information supplémentaire ou quelle robustesse apporte la structure en échelle de l'image ?

Une thèse doit commencer à l'UTC sous la direction de Jean-Pierre Cocquerez sur le problème de fouille de base de données d'images médicales. Il est prévu d'étudier la pertinence des analyses ensembles-échelle pour ce problème.

6) Compression orientée régions, caricature visuelle –

Nous avons montré que la méthode d'escalade conduisait des solutions remarquables d'un point de vue débit/distorsion. Il reste à evaluer les performances effectives de la méthode pour la compression d'images orientée régions. Un atout important est que comme toute une famille de solutions est calculée, l'utilisateur peut régler interactivement le compromis débit/distorsion qui lui convient, en inspectant *visuellement* le degré de dégradation obtenu pour un volume donné.

7) Croisement de données de différentes résolutions, appariement multi-échelles –

La robustesse de la méthode à des changement de résolution (liée à la propriété de covariance d'échelle

de la solution) permet d'envisager des utilisations pour le croisement ou l'appariement de données de différentes résolutions.

10.3 Conclusion finale

Concluons ce document par deux citations tirées du premier cycle de conférences de l'Université de tous les savoirs (an 2000). Ce cycle de conférences comportait dix-sept interventions sur le thème des mathématiques, parmi lesquelles deux conférences étaient consacrées à des théories utilisés en analyse d'images, voire développées pour l'analyse d'images. La première, d'Yves Meyer, portait sur les ondelettes et la seconde, de Pierre-Louis Lions, portait sur les modélisations par EDP.

Meyer disait : «La révolution numérique envahit notre vie quotidienne, elle modifie notre travail, notre relation aux autres, notre perception des sons et des images. En ouvrant un magazine, on a toutes les chances d'y trouver une apologie des caméras numériques ; ces publicités sont en général bien faites et nous fournissent une première définition de la finalité de la révolution numérique : la possiblité d'agir, d'intervenir sur les images et les sons, de les manipuler, de les améliorer, etc.» (Meyer 2000, p.69).

De son côté, Lions disait : «La notion d'échelle est une notion centrale : les modèles ne sont valables qu'à une certaine échelle. Et un enjeu scientifique (et applicatif) fondamental est la compréhension des relations entre les échelles et les modèles associés, ce qui conduit à la modélisation du couplage d'échelles» (Lions 2000, p.106).

Il nous semble que l'approche ensembles-échelle proposée dans ce document s'inscrit pleinement dans ce double programme.

Listes des communications

Nous avons inclus ici toutes les communications dont nous sommes auteur ou co-auteur.

Celles qui sont précédées du symbole ▲ concernent les méthodes de segmentation décrites dans ce mémoire.

Article de revue

▲**Guigues, L., Le Men, H. et Cocquerez, J-P. 2003a**. The hierarchy of the cocoons of a graph and its application to image segmentation. *Pattern Recognition Letters* 24(3), pp. 1024-1066, may 2003.

Communications dans des conférences avec actes

▲**Guigues, L., Le Men, H. et Cocquerez, J-P. 2003b.** Scale-sets image analysis. In *proc. of IEEE Int. Conf. on Image Processing (ICIP'03), Barcelona, Spain*, 14-17 sept. 2003.

▲**Guigues, L., Le Men, H. et Cocquerez, J-P. 2003c.** Représentation et analyse ensembles-échelle d'une image. In *actes du XIXème colloque GRETSI sur le Traitement des Images et du Signal, Paris, France*, sept. 2003.

▲**Taillandier, F., Guigues, L. et Deriche, R. 2003.** A framework for constrained multi-scale range image segmentation. In *proc. of IEEE Int. Conf. on Image Processing (ICIP'03), Barcelona, Spain*, 14-17 sept. 2003.

Viglino, J.-M. et Guigues, L. 2002. Géorérérencement automatique de feuilles cadastrales. In *actes du XIIIème congrès RFIA, Angers, France*, pp. 135-143, jan. 2002.

▲**Guigues, L., Le Men, H. et Cocquerez, J-P. 2001a.** Graphs, cocoons and image segmentation. In *proc. of 3rd IAPR TC-15 workshop on Graph-based Representations in Pattern Recognition (GbR), Ischia, Italy*, pp. 22-31, may 2001.

▲**Guigues, L., Le Men, H. et Cocquerez, J-P. 2001b**. Segmentation d'image par minimisation d'un critère MDL dans une pyramide de segmentations. In *actes du XVIIIème colloque GRETSI sur le Traitement des Images et du Signal, Toulouse, France*, sept. 2001.

Viglino, J.-M. et Guigues, L. 2001a. Cadastre map assembling : A puzzle game resolution. In *proc. of 6th International Conference on Document Analysis and Recognition (ICDAR), Seattle, USA*, pp. 1235-1239, sept. 2001.

Viglino, J.-M. et Guigues, L. 2001b. Géoréférencement automatique de feuilles cadastrales. In *Bulletin d'information de l'IGN* no. 72, Journées de la recherche de l'IGN 2001, Saint-Mandé, France, février 2001.

Paparoditis, N., Maillet, G., Taillandier, F., Jibrini, H., Jung, F., Guigues, L. et Boldo, D. 2001. Multi-image 3D feature and DSM extraction for change detection and building reconstruction. In *proc. of 3rd workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, Ascona, Switzerland*, pp. 217-230, june 2001.

Viglino, J.-M. et Guigues, L. 2000. Automatic road extraction through a light propagation simulation. In *proc. of. XXXIIIth ISPRS congress, Amsterdam, The Netherlands*, jul. 2000.

Guigues, L. 1999a. Comparison of images segmentations using a hierarchical model for n to m regions matching. In *proc. of 2nd IAPR TC-15 workshop on Graph-based Representations in Pattern Recognition (GbR), Haindorf, Austria*, pp. 41-50, may 1999.

Ruskoné, R., Guigues, L., Airault, S. et Jamet, O. 1996. Vehicle detection on aerial images : a structural approach. In *proc. ICPR'96, Vienna, Austria*, vol. III, pp. 900-904.

Rapports techniques IGN/MATIS

Guigues, L. 2001 Extracting vegetation patches from colour aerial images. English, 8 pages, 2001.

▲**Guigues, L. 2000** Graphes, cocons et segmentation d'image. Français, 27 pages, juin 2000.

Guigues, L. 2000b Détection et correction automatique des rayures sur les clichés aériens numérisés. Français, 5 pages, mars 2000.

Guigues, L. 1999b. A light propagation principle for parallel curves detection. Anglais, 6 pages, 1999. (Article refusé par ICCV'99 qui n'a pas été resoumis à d'autres conférences mais déposé comme rapport technique)

Guigues, L. 1998. Tronçonnement optimal d'un arbre en composantes de diamètre borné. Français, 7 pages, décembre 1998.

Guigues, L. et Pierrot Deseilligny, M. 1998. Non interleaved partitions : a framework for the comparison of image segmentations. Anglais, 10 pages, 1998. (Article refusé par ICPR'98 qui n'a pas été resoumis à d'autres conférences mais déposé comme rapport technique)

Guigues, L. 1997a. Relations de non entrelacement entre deux partitions d'un ensemble et applications au traitement d'image. Français, 21 pages, décembre 1997.

Guigues, L. 1997b. Recalage de modèles indéformables de bâtiments pour la restitution semi-automatique du bâti répétitif. Français, 10 pages, 1997.

Communications orales

▲**Guigues, L.**. Segmentation multi-échelles d'images. Communication orale, séminaire de vision par ordinateur de l'INRIA Rhône-Alpes, Montbonnot, 15 Mars 2002 (un résumé est disponible sur http ://www.inrialpes.fr/movi/seminar/).

▲**Guigues, L.**. Approches structurelles de la segmentation d'images. Communication orale en séance plénière du GT3 «systèmes de segmentation pour la représentation et l'analyse des images» du GdR ISIS (CNRS/MENRT), ENST, Paris, Mars 2001 (voir http ://www-isis.enst.fr).

Guigues, L.. Un algorithme pour la détection de zones de végétation sur images aériennes. Communication orale au GT6.2 «systèmes complexes pour l'analyse d'images» du GdR ISIS (CNRS/MENRT), ENST, Paris, Oct. 2000 (un résumé est disponible sur http ://www-isis.enst.fr/Kiosque/journees/Annonces/annonce-reunion-119.html).

Autres

Deux articles de vulgarisation portent sur les méthodes de géoréférencement automatique de plans cadastraux que nous avons développées avec Jean-Marc Viglino :

Un grand puzzle de 500 000 pièces. In *SIG La Lettre*, no. 25, mars 2001, pp.1-2.

Géoréférencement automatique des planches cadastrales. In *IGN magazine* no. 6, juillet-août 2001, numéro spécial sur le Référentiel à Grande Échelle (RGE), p.10.

Références

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B. Petrov et F. Caski (Eds.), *Second International Symposium on Information Theory, Budapest*, pp. 267–281.

Aloimonos, Y. et Badyopadhyay, A. 1987. Active vision. In *IEEE ICCV*, pp. 35–54.

Alvarez, L., Guichard, F., Lions, P.-L., et Morel, J.-M. 1993. Axioms and Fundamental Equations of Image Processing. *Arch. Rational Mech. Anal.* 123 :199–257.

Arnheim, R. 1969. *Visual Thinking.* Univ. of california press (La pensée visuelle - Flammarion 1976 pour l'édition française).

Arnoux, P. et Chemla, K. 1992. Systèmes dynamiques et théorie ergodique. in Chaos et détermisme, ouvrage collectif sous la dir. de A. Dahan Dalmedico, J.-L. Chabert et K. Chemla. Coll. Points sciences.

Asseraf, M. 1998. *Extension et optimisation pour la segmentation de la distance de Kolmogorov-Smirnov.* Ph. D. thesis, Univ. Paris Dauphine.

Atlan, H. 1982. *Entre le cristal et la fumée.* PUB.

Auger, P. 1992. *Hiérarchies et échelles en écologie.* Naturalia Publications.

Baillard, C. 1997. *Calcul de modèles numériques de surface, etc.* Ph. D. thesis, ENST.

Ballard, D. H. 1981. Generalized Hough Transform to Detect Arbitrary Patterns. *IEEE Trans. on Pattern Annalysis and Machine Intelligence 3*(2) :111–122.

Ballester, C., Caselles, V., et González, M. 1994. Affine Invariant Segmentation by Variational Method. In *Proc. of 9th RFIA, Paris*, pp. 379–390.

Batbedat, A. 1990. *Les approches pyramidales dans la classification arborée.* Masson.

Baxter, R. et Oliver, J. 1994. MDL and MML : similarities and differences. Technical Report TR 207, Dep. Comp. Science, Monash Univ., Clayton, Victoria 3168, Australia.

Bellmann, R. 1957. *Dynamic Programming.* Princeton University Press.

Bernardo, J. 1997. Noninformative Priors Do Not Exist : A Discussion. *J. Statististics Planning and Inference* 65 :159–189.

Bertolino, P. 1995. *Contribution des pyramides irrégulières en segmentation d'images multirésolution.* Ph. D. thesis, Institut National Polytechnique de Grenoble.

Bertrand, G., Everat, J.-C., et Couprie, M. 1997. Image Segmentation Through Operators based upon Topology. *Journal of Electronic Imaging 6*(4) :395–405.

Bertrand, P. 2002. Set systems for which each set properly intersects at most one other set - Application to pyramidal clustering. Technical report, Cahiers de Mathématiques du Ceremade - Univ. Paris 9.

Besag, J. 1986. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society* B 48 :259–302.

Beucher, S. 1990. *Segmentation d'images et morphologie mathématique.* Ph. D. thesis, Ecole des Mines de Paris - Centre de Morphologie Mathématique.

Beucher, S. et Lantuejoul, C. 1979. Use of watersheds in contour detection. In *Proceedings of Int. Workshop on Image Processing, Real-time Edge and Motion Detection/Estimation, Rennes (France).*

Beucher, S. et Meyer, F. 1993. The morphological approach to segmentation : the watershed transformation. In Dougherty (Ed.), *Mathematical Morphology in Image Processing*, Chapter 12 :pp. 433–481. Marcel Dekker.

Bister, M., Cornelis, J., et Rosenfeld, A. 1990. A critical view of pyramidal segmentation algorithms. *Pattern Recognition Letters* 11 :605–617.

Blake, A. et Zisserman, A. 1987. *Visual reconstruction.* MIT Press, Cambridge, MA.

Boldo, Fuchs, Jung, et Wehrstedt 2001. Automating the aerial triangulation of digital aerial frame camera images. In *Proc. of 3rd Int. Symp. on Mobile Mapping Technology, Le Caire.*

Boldo, D. 2002. *Fabrication conjointe de MNS et d'ortho-images pour la visualisation perspective de scènes urbaines.* Ph. D. thesis, Univ. René Descartes - Paris V.

Boucart, M. 1996. *La reconnaissance des objets.* La psychologie en plus, Presses Universitaires de Grenoble.

Bouman, C. A. et Shapiro, M. 1994. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing 3*(2) :162–177.

Boykov, Y., Veksler, O., et Zabih, R. 1999. Fast Approximate Energy Minimization via Graph Cuts. In *Proc. of 7th Int. Conf. on Comp. Vision (ICCV), sept. 20-27, Kerkyra, Greece.*

Braquelaire, J.-P. et Brun, L. 1998. Image Segmentation with Topological Maps and Inter-Pixel Representation. *Journal of Visual Comm. and Image Repres. 9*(1) :62–79.

Breiman, L., Friedman, J. H., Olshen, R. A., et Stone, C. J. 1984. *Classification and regression trees.* Wadsworth and Brooks.

Brun, L. et Mokhtari, M. 2001. Graph-based representations in different application domains. In *Proc. of 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition, Ischia (Italy)*, pp. 115–124. CUEN.

Burke Hubbard, B. 1995. *Ondes et ondelettes - la saga d'un outil mathématique.* Pour la science - diffusion Belin.

Burt, P. et Adelson, E. 1983. The laplacian pyramid as a compact image code. *IEEE Trans. on Communications* 31 :482–540.

Canny, J. 1986. A computational approach to edge detection. *IEEE trans. on Pattern Analysis and Machine Intelligence 8*(6) :679–698.

Caselles, V., Coll, B., et Morel, J.-M. 1998. Topographic maps. Technical report, Preprint Ceremade - Univ. Paris Dauphine.

Caselles, V., Masnou, S., et Morel, J.-M. 2001. La vision, une machine géométrique ? *La Recherche, numéro spécial : grandes et petites énigmes mathématiques* 346 :78–82.

Catoni, O. 2003. Mathématiques et reconnaissance des formes. *La gazette des mathématiciens* :40–44.

Chanas, L. 2001. *Méthodes variationnelles pour la restauration de séquences d'images fortement dégradées. Application aux images infrarouges éblouies par laser.* Ph. D. thesis, Université de Cergy-Pontoise.

Charon, I., Germa, A., et Hudry, O. 1996. *Méthodes d'Optimisation Combinatoire.* Masson.

Cho, K. et Meer, P. 1997. Image Segmentation from Consensus Information. *CVGIP : Image Understanding 68*(1) :72–89.

Clarke, R. 1995. *Digital compression of still images and video.* Academic Press (Signal processing and its applications).

Cocquerez, J.-P. et Philipp, S. 1995. *Analyse d'images : filtrage et segmentation.* (Ouvrage collectif sous la direction de Jean-Pierre Cocquerez et Sylvie Philipp) Masson.

Cohen, L. 1991. On active contours models and balloons. *Computer Vision, Graphics and Image Processing : Image Understanding 53*(2) :211–218.

Coifman, R. et Wickerhauser, M. 1992. Entropy-based algorithms for best-basis selection. *IEEE trans. on Info. Theory* 38 :713–718.

Conway, J. et Sloane, N. 1993. *Sphere packings, lattices and groups*. 2nd Ed. Springer-Verlag, NY.

Cormen, T., Leiserson, C., et Rivest, R. 1994. *Introduction à l'algorithmique*. Dunod.

Cornu, P. et Smorlarz, A. 2001. Caractérisation de la signature texturelle d'une image. In *Actes du XVIIIème colloque GRETSI sur le Traitement des Images et du Signal, Toulouse, France*.

Cover, T. et Thomas, J. 1991. *Elements of information theory*. John Wiley, New York.

D'Agostini, G. 1999. Bayesian reasoning in high energy physics - principles and applications. Technical report, Univ. di Roma "La Sapienza" - CERN Yellow Report 99-03.

D'Ales, J.-P., Froment, J., et Morel, J.-M. 1999. Reconstruction visuelle et généricité. *Intellectica 1*(28) :11–35.

Danchin, A. 2001. Le vivant et la machine de Turing. *La Recherche, numéro spécial : grandes et petites énigmes mathématiques* 346 :88–93.

Dang, T., Jamet, O., et Maitre, H. 1994. An image segmentation technique based on edge-preserving smoothing filter and anisotropic diffusion. In *IEEE Southwest Symposium Image Analysis and Interpretation. Dallas, Texas (USA)*.

Delahaye, J.-P. 1999. *Information, complexité et hasard*. Hermes.

Deriche, R. 1987. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision 1*(2) :167–187.

Deriche, R. 1993. Recursively implementing the gaussian and its derivatives. Rapport de recherche N°1893, INRIA.

Desolneux, A., Moisan, L., et Morel, J.-M. 2000. Maximal meaningful events and application to image analysis. Technical report, Preprint ENS Cachan/CMLA.

Desolneux, A., Moisan, L., et Morel, J.-M. 2001a. Automatic image analysis : a challenge for Computer Vision. In *Actes du congrès GRETSI sur le Traitement des Images et du Signal, Toulouse, France*.

Desolneux, A., Moisan, L., et Morel, J.-M. 2001b. Partial Gestalts. Technical report, Preprint ENS Cachan/CMLA.

Diday, E. 1984. Une représentation visuelle des classes empiétantes : les pyramides. Technical report, INRIA N°291, Rocquencourt, France.

Donoho, D. 1997. CART and best-ortho-basis selection : a connection. *Annals of statistics* 25 :1870–1911.

Donoho, D. 1999. Wedgelets : Nearly minimax estimation of edges. *Annals of statistics* 27 :859–897.

Donoho, D. 2000. Counting Bits with Kolmogorov and Shannon. Technical report, Stanford Univ.

Drot, S., Descombes, X., Le Men, H., et Zérubia, J. 2002. Object Point Processes for Image Segmentation. In *proc. of ICPR, Québec City, Canada, 11-15 Août*.

Dufournaud 2001. *Navigation aérienne et guidage terminal à partir de données bidimensionnelles*. Ph. D. thesis, Institut National Polytechnique de Grenoble.

Dufournaud, Schmid, et Horaud 2000. Matching images with different resolutions. In *Proc. CVPR, Hilton Head Island, USA*, Volume 1, pp. 612–618.

Felzenszwalb, P. et Huttenlocher, D. 2001. Efficiently computing a good segmentation. Int'l Journal of Computer Vision (soumis).

Forsyth, D. et Ponce, J. 2003. *Computer Vision - A modern approach*. Prentice Hall.

Fua, P. et Brechbühler, C. 1996. Imposing Hard Constraints on Soft Snakes. Technical report, Tech Note 553, Artificial Intelligence Center, SRI International, Cambridge.

Fuchs, F. 2001. *Contribution à la reconstruction du bâti en milieu urbain, à l'aide d'images aériennes stéréoscopiques à grande échelle. Etude d'une approche structurelle.* Ph. D. thesis, Université Paris 5 - René Descartes.

Fuchs, F. et Le Men, H. 1999. Detecting planar patches in urban scenes. In *Proceedings of SPIE, Visual Information Processing VIII, Orlando, USA*, Volume 3716, pp. 167–176.

Gagalowicz, A. et Monga, O. 1986. A new approach for image segmentation. In *8th ICPR, Paris.*

Garcin, L., Descombes, X., Zérubia, J., et Le Men, H. 2001. Building detection by Markov object processes and a MCMC algorithm. Technical Report 4206, INRIA.

Garey, M. R. et Johnson, D. S. 1979. *Computers and Intractability : A Guide to the Theory of NP-Completeness.* W.H. Freeman and Company.

Geman, S. et Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*(6) :721–741.

Gondran, M. et Minoux, M. 1995. *Graphes et algorithmes (3eme édition revue et augmentée. 1ere edition 1979).* Eyrolles.

Guillois, J.-P. 1996. *Techniques de compression des images.* Hermes.

Haralick et Shapiro 1991. Glossary of computer vision terms. *Pattern Recognition 24*(1) :69–93.

Haris, K., Estradiadis, S. N., Maglaveras, N., et Katsaggelos, A. K. 1998. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. on Image Processing 7*(12) :1684–1699.

Haton, J.-P. et Haton, M.-C. 1993. *L'Intelligence Artificielle.* Presses Universitaires de France - coll. que sais-je ? (3e édition corrigée, 1ere ed. 1989).

Holender, D. 2001. *Notions de psychologie.* Presses Universitaires de Bruxelles.

Horaud, R. et Monga, O. 1995. *Vision par ordinateur. Outils fondamentaux.* Hermes.

Horowitz, S. L. et Pavlidis, T. 1976. Picture Segmentation by a Tree Traversal Algorithm. *J. of the Assoc. for Comp. Mach. 23*(2) :368–388.

Hough, P. 1962. Methods and means for recognizing complex pattern. U.S. Pattent 3069654.

Jasiobedski, P. 1993. Adaptive adjacency graphs. *SPIE Geometric Methods in Computer Vision II* 2031 :294–303.

Jibrini, H. 2002. *Reconstruction 3D de bâtiments à partir de données cadastrales et d'images aériennes.* Ph. D. thesis, Univ. Marne-La-Vallée.

Jimenez, M. 1997. *La psychologie de la perception.* Flammarion - dominos.

Johnson, R. and Wichem, D. 1982. *Applied Multivariate Statistical Analysis.* Prentice Hall, Inc., Englewood Cliffs.

Jolion, J.-M. et Montanvert, A. 1992. The Adaptive Pyramid, a Framework for 2D image analysis. *CVGIP : Image Understanding 55*(3) :339–348.

Jungnickel, D. 1999. *Graphs, Networks and Algorithms.* Springer.

Kanisza, G. 1997. *La grammaire du voir.* Diderot, coll. "Pratique des sciences", Paris.

Kanungo, T., Dom, B., Niblack, W., Steele, D., et Sheinvald, J. 1995. MDL-Based Multi-Band Image Segmentation Using a Fast Region Merging Scheme. Technical Report RJ 9960 (87919), IBM Research Division.

Kass, M., Witkin, A., et Terzopoulos, D. 1987. Snakes : Active Contour Models. *International Journal on Computer Vision 1*(4) :321–331.

Kato, Z., Berthod, M., et Zerubia, J. 1996. A hierarchical Markov random field model and multi-temperature annealing for parallel image classification. *Graphical Models and Image Processing* 58 :18–37.

Kervrann, C., Hoebeke, M., et Trubuil, A. 1999. Segmentation d'image par sélection de courbes de niveau. In *Dix-septième colloque GRETSI, Vannes, France*, pp. 631–634.

Kervrann, C. et Trubuil, A. 2002. Optimal level curves and global minimizers of cost functionals in image. *J. of Mathematical Imaging and Vision 17*(2) :153–174.

Koenderink, J. 1984. The structure of images. *Biol. Cybern.* 50 :363–370.

Koepfler, G., Lopez, C., et Morel, J. M. 1994. A Multiscale Algorithm for Image Segmentation by Variational Method. *SIAM Journal on Numerical Analysis 31*(1) :282–299.

Kropatsch, W. 1995. Equivalent Contraction Kernels and the Domain of Irregular Pyramids. Technical report, Institute for Automation, Wien.

Kruskal, Jr., J. B. 1956. On the shortest spanning subtree of a graph and the travelling salesman problem. *Proc. Amer. Math. Soc.* 7 :48–50.

Kuhn, H. W. 1955. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2 :83–97.

Kurita, T. 1991. An Efficient Agglomerative Clustering Algorithm Using a Heap. *Pattern Recognition 24*(3) :205–209.

Kurita, T. 1995. A Efficient Clustering Algorithm for Region Merging. *IEICE Trans. on Information and Systems E78-D*(12).

Laferté, J. M., Pérez, P., et Heitz, F. 2000. Discrete Markov image modeling and inference on the quadtree. *IEEE trans. on Image Processing 9*(3) :390–404.

Lance et Williams 1967. A general theory of classificatory sorting strategies. *The Computer Journal* 9 :373–380.

Lanterman, A. 2000. Schwarz, Wallace and Rissanen : Intertwining Themes in Theories of Model Order Estimation. Technical report, Coordinated Science Lab., univ. of Illinois at Urbana-Champaign, USA.

Laporterie, F. 2002. *Représentations hiérarchiques d'images avec des pyramides morphologiques. Application à l'analyse et à la fusion spatio-temporelle de données en observation de la Terre.* Ph. D. thesis, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace.

Leclerc, B. 1985. La comparaison des hiérarchies : indices et métriques. *Math. sci. hum. 23è année*(92) :5–40.

Leclerc, Y. 1989a. Constructing Simple Stable Descriptions for Image Partitioning. *International Journal of Computer Vision 3*(1) :73–102.

Leclerc, Y. 1989b. Simplicity of Description as the Basis for Visual Interpretation. Technical report, AI Center, SRI International.

Lee, T. C. M. 1998. Segmenting Images Corrupted by Correlated Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 :481–492.

Lee, T. C. M. 2000. A Minimum Description Length Based Image Segmentation Procedure, and Its Comparison with a Cross-Validation Based Segmentation Procedure. *Journal of the American Statistical Association* 95 :259–270.

Lerman, I. 1981. *Classification et analyse ordinale des données.* Dunod, Paris.

Li, M. et Vitanyi, P. 1993. *An Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, N.-Y.

Li, S. 1995. On discontinuity-adaptive smoothness priors in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence 17*(6) :576–586.

Li, S. Z. 2001. *Markov Random Field Modeling in Image Analysis.* Springer-Verlag.

Lindeberg, T. 1994. *Scale-space Theory in Computer Vision.* Kluwer Academic.

Lions, P.-L. 2000. Analyse, modèles et simulations. In *actes des conférences de l'université de tous les savoirs. Vol. 13 - Les Mathématiques. Odile Jacob.*, pp. 101–109.

Lochak, G. 1994. *La Géométrisation de la Physique.* Flammarion.

Lowe, D. 1985. *Perceptual organization and visual recognition.* Kluwer Academics Publishers.

Luong, Q.-T. 1990. La Couleur en Vision par Ordinateur : 1. Une Revue. Technical report, INRIA Sophia-Antipolis - Rapport de Recherche N°1251.

Mach, E. 1886. *L'analyse des sensations - Le rapport du physique au psychique (Beiträge zur Analyse der Empfindungen 1886, rev. 1905 et 1922).* Ed. Jacqueline Chambon 1995 (trad. de l'édition de 1922).

Maître, H. 1985. Un panorama de la transformation de Hough. *Traitement du signal 2*(4).

Maître, H. 2003. *Le traitement des images.* (Ouvrage collectif sous la direction de Henri Maître) Paris, Hermès sciences pub.

Mallat, S. 1989. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*(7) :674–693.

Marr, D. 1977. Artificial Intelligence : a personal view. *Artificial Intelligence* 9 :37–48.

Marr, D. 1982. *Vision : a computational investigation into the human representation and processing of visual information.* W.H. Freeman and Company, NY 29-61.

Marr, D. et Hildreth, E. 1980. Theory of edge detection. *Proc. Roy. Soc. Lond.* B207 :187–217.

Matheron, G. 1975. *Random sets and integral geometry.* John Wiley, N.Y.

Meer, P. 1989. Stochastic Image Pyramids. *Computer Vision, Graphics and Image Processing* 45 :269–294.

Messmer, B. et Bunke, H. 1995. Subgraph isomorphism detection in polynomial time on preprocessed model graph. In *proc. of 2nd Asian Conf. on Comp. Vision, Singapore*, pp. 373–382.

Meyer, F. 1999a. Graph Based Morphological Segmentation. In *Proc. 2nd IAPR TC-15 workshop on Graph based Representations, Haindorf, Austria*, pp. 51–60.

Meyer, F. 1999b. Morphological multiscale and interactive segmentation. In *Proc. of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, Antalya, Turkey.*

Meyer, F. 2001. An overview of morphological segmentation. *Int. Journal of Pattern Recognition and Artificial Intelligence 15*(7) :1089–1118.

Meyer, F. et Beucher, S. 1991. Morphological Segmentation. *Journal of Visual Communication and Image Representation 1*(1) :21–46.

Meyer, Y. 2000. Les ondelettes et la révolution numérique. In *actes des conférences de l'université de tous les savoirs. Vol. 13 - Les Mathématiques. Odile Jacob.*, pp. 55–73.

Monasse, P. et Guichard, F. 2000. Fast computation of a contrast-invariant image representation. *IEEE Trans. on Image Processing 9*(5) :860–872.

Montanvert, A., Meer, P., et Rosenfeld, A. 1991. Hierarchical image analysis using irregular tesselations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*(4) :307–316.

Morel, J.-M. et Solimini, S. 1995. *Variational methods in image segmentation.* Progress in Nonlinear Differential Equations and Their Applications, vol. 14, Birkhaüser, Boston.

Moulin, P. et Liu, J. 2000. Statistical imaging and complexity regularization. *IEEE trans. on Information Theory, sp. issue on information theoretic imaging 46*(5) :1881–1895.

Muchnik, I. et Mottl, V. 1998. Bellman Functions on Trees for Segmentation, Generalized Smoothing, Matching and Multi-Alignment in Massive Data Sets. Technical report, DIMACS.

Mumford, D. et Shah, J. 1989. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Comm. on Pure and Applied Math. 17*(4) :577–685.

Ndili, U., Nowak, R., et Figueiredo, M. 2001. Coding theoretic approach to image segmentation. In *proc. of IEEE Int. Conf. on Image Processing (ICIP'01), Thessaloniki, Greece.*

Nottale, L. 2001. Relativité d'échelle et morphogénèse. *Revue de synthèse : Objets d'échelles, Tome 122, 4è S. n°1, Janvier-Mars 2001* :93–116.

Oliver, J. et Baxter, R. 1994. MML and bayesianism : similarities and differences. Technical Report TR 206, Dep. Comp. Science, Monash Univ., Clayton, Victoria 3168, Australia.

Oliver, J. et Hand, D. 1994. Introduction to Minimal Encoding Inference. Technical Report TR 205, Dep. Comp. Science, Monash Univ., Clayton, Victoria 3168, Australia.

Olsen, O. 1996. Multi-Scale Segmentation of Grey Scale Images. Master's thesis, Univ. of Copenhagen.

Olsen, O. et Nielsen, M. 1997. Multi-scale gradient magnitude watershed segmentation. In *ICIAP'97 - 9th International Conference on Image Analysis and Processing, volume 1310 of Lecture Notes in Computer Science, Florence, Italy*, pp. 6–13.

Osher, S. et Fedkiw, R. 2001. Level Sets Methods : An Overview and Some Recent Results. *J. Comput. Physics* 169 :463–502.

Osher, S. et Setian, J. 1988. Front propagating with curvature-dependent speed : algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics* 79 :12–49.

Paragios, N. 2000. *Geodesic Active Regions and Level Sets Methods*. Ph. D. thesis, Université Nice - Sophia Antipolis.

Pateux, S. 1998. *Segmentation spatio-temporelle et codage orienté-régions de séquences vidéo basés sur le formalisme MDL*. Ph. D. thesis, Université de Rennes I.

Pavlidis, T. 1977. *Structural Pattern Recognition*. Springer-Verlag.

Perona, P. et Malik, J. 1990. Scale space and edge detection using anisotropic diffusion. *IEEE trans. on PAMI 12*(7) :629–639.

Pierrot Deseilligny, M. 1990. *Contribution à la lecture automatique de cartes*. Ph. D. thesis, Univ. Paris V.

Pierrot Deseilligny, M., Stamon, G., et Suen, C. Y. 1998. Veinerization : A New Shape Description for Flexible Skeletonization. *IEEE Trans. on PAMI 20*(5) :505–521.

Poggi, G. et Ragozini, A. R. P. 1999. Image Segmentation by a Tree-Structured Markov Random Field. *IEEE Signal Processing Letters* 6 :155–157.

Poggio, T., Torre, V., et Koch, C. 1985. Computational Vision and Regularization Theory. *Nature* 317 :314–319.

Poincaré, H. 1902. *La science et l'hypothèse*. Champs Flammarion.

Poincaré, H. 1905. *La valeur de la science*. Champs Flammarion.

Préteux, F. 1992. Watershed and Skeleton by Influence Zone : A Distance-Based Approach. *Journal of Mathematical Imaging and Vision* 1 :239–255.

Prim, R. C. 1957. Shortest connection networks and some generalizations. *Bell System Technical Journal* 31 :1389–1401.

Puzicha, J. et Buhmann, J. 1999. Multiscale Annealing for Grouping and Unsupervised Texture Segmentation. *Computer Vision and Image Understanding 76*(3) :213–230.

Randriamasy, S. 1992. *Segmentation Descendante Coopérative en Régions de Paires d'Images Stéréoscopiques*. Ph. D. thesis, Université Paris IX - Dauphine.

Rissanen, J. 1978. Modeling by shortest data description. *Automatica* 14 :465–471.

Rissanen, J. 1983. A universal prior for the integers and estimation by MDL. *Ann. of Statistics 11*(2) :416–431.

Rissanen, J. 1989. *Stochastic Complexity in Statistical Inquiry*. Scientific World, Singapore.

Rissanen, J. 1993. Information theory and neural nets. In P. Smolensky et al. (Ed.), *Mathematical Perspectives in Neural Networks*. Lawrence Erlbaum.

Rissanen, J. 1999. Lectures on Statistical Modeling Theory. Support de cours disponible sur www.cs.tut.fi/rissanen/papers/lectures2.ps.

Ruas, A. 1999. *Modèle de généralisation à base de contraintes et d'autonomie*. Ph. D. thesis, Univ. de Marne la vallée.

Salembier, P. et Garrido, L. 2000. Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval. *IEEE Trans. on Image Processing 9*(4) :561–576.

Salembier, P., Oliveras, A., et Garrido, L. 1998. Anti-extensive Connected Operators for Image and Sequence Processing. *IEEE Trans. on Image Processing 7*(4) :555–570.

Salembier, P. et Pardàs, M. 1994. Hierarchical Morphological Segmentation for Image Coding. *IEEE Trans. on Image Processing 38*(3) :359–386.

Salembier, P. et Serra, J. 1995. Flat Zones Filtering, Connected Operators, and Filters by Reconstruction. *IEEE trans. on Image Processing 4*(8) :1153–1160.

Sapiro, G. et Tannenbaum, A. 1993. Affine Invariant Scale Space. *Int. J. of Computer Vision* 11 :24–44.

Schmid, C. 1996. *Appariement d'images par invariants locaux de niveaux de gris.* Ph. D. thesis, Institut National Polytechnique de Grenoble.

Schmitt, M. et Mattioli, J. 1994. *Morphologie Mathématique.* Masson.

Schrödinger, E. 1956. *Mind and Matter.* "Tarner Lectures", cours donnés à Trinity College (Cambrige). Edition originale : Cambridge University Press, 1958. Trad/Ed. française : L'esprit et la matière, Seuil, coll. sources du savoir, 1990.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics 6*(2) :461–464.

Serra, J. 1982. *Image analysis and mathematical morphology.* Academic Press.

Serra, J. et Salembier, P. 1993. Connected operators and pyramids. In *SPIE conf. on Image Algebra and Mathematical Morphology, San Diego, CA*, Volume 2030, pp. 65–76.

Shannon, C. 1948. The mathematical theory of communication. *Bell System Techical Journal* 27 :379–423 ; 623–656.

Shannon, C. 1959. Coding theorems for a discrete source with a fidelity criterion. *IRE Natl. Conv. Record,* Part 4 :142–163.

Shapiro 1996. Connected Components Labelling and Adjacency Graph Construction. In C. Konig Rosenfeld (Ed.), *Topological Algorithms for Digital Image Processing, Machine Intelligence and Pattern Recognition*, Volume 19. Elsevier.

Shi, J. et Malik, J. 2000. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*(8) :888–905.

Siegel 1956. *Non parametrics statistics.* Mc Graw-Hill.

Simon, J.-C. 1985. *La Reconnaissance des Formes par Algorithmes.* Masson.

Smirnov, N. 1948. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.* 19 :279–281.

Solomonoff, R. 1997a. The Discovery of Algorithmic Probability. *Journal of Computer and System Science 55*(1) :73–88.

Solomonoff, R. 1997b. Does Algorithmic Probability Solve the Problem of Induction ? Technical Report 97-1, Oxbridge Research, Cambridge.

Tamaki, T., Yamamura, T., et Ohnishi, N. 1999. Image segmentation and object extraction based on geometric features of regions. In *Proc. of VICIP'99 (Visual Communications and Image Processing) : Electronic Imaging '99, IS&T and SPIE, San Jose (CA), USA.*

Tanimoto, S. et Pavlidis, T. 1975. A hierarchical data structure for picture processing. *Computer Graphics and Image Processing 4*(2) :104–109.

Toet, A. 1995. Combinatorial Optimization and Image Analysis : a literature survey. Technical report, CALMA.TNO.WP31.AT.95a.

Tutte, W. 1984. *Graph theory.* Addison-Wesley (Réimprimé par Cambridge University Press en 2001).

Vézien, J.-M. 1995. *Techniques de Reconstruction Globale par Analyse de Paires d'Images Stéréoscopiques.* Ph. D. thesis, Université Paris 7.

Vincent, L. 1990. *Algorithmes morphologiques à base de files d'attente et de lacets : Extension aux graphes.* Ph. D. thesis, Ecole des mines de Paris.

Vinet, L. 1991. *Segmentation et Mise en Correspondance de Régions de Paires d'Images Stéréoscopiques*. Ph. D. thesis, Université Paris IX - Dauphine.

Vitanyi et Li 2000. Minimum Description Length induction, Bayesianism and Kolmogorov complexity. *IEEE trans. on Information Theory 46*(2) :446–464.

Wallace, C. et Boulton, D. 1968. An Information Measure for Classification. *The Computer Journal* 11 :185–194.

Wallace, C. et Dowe, D. 1999. Minimum Message Length and Kolmogorov Complexity. *The Computer Journal - Sp. Issue on Kolmogorov Complexity 42*(4) :270–283.

Wallace, C. et Freeman, P. 1987. Estimation and inference by compact coding. *Journal of the Royal Statistical Society 49*(3) :241–252.

Wang, J.-P. 1998. Stochastic Relaxation on Partitions With Connected Components and its Application to Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*(6) :619–636.

Weisstein, E. 1998. *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall / CRC - voir aussi la version en ligne sur www.astro.virginia.edu/eww6n/math/.

Wertheimer, M. 1924. Über Gestalttheorie. Communication devant la Kant Society, Berlin, 7 Décembre 1924 - Erlangen 1925. Traduction de Willis D. Ellis publiée dans son Source Book of Gestalt Psychology, NY, Harcourt, Brace and Co, 1938. Réimprimé par le Gestalt Journal Press, NY, 1997. Également disponible sur le site de la GTA (Society for Gestalt Theory and Applications) : http ://www.gta.com.

Witkin, A. 1983. Scale space filtering. In *Proc. of 8th Int. Joint Conf. on Artificial Intell., Karlsruhe, West Germany*, pp. 1019–1021.

Woodcock, C. et Strahler, A. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21 :311–332.

Wu, Z. et Leahy, R. 1993. An Optimal Graph Theoretic Approach to Data Clustering : Theory and its Application to Image Segmentation. *IEEE trans. on Pattern Analysis and Machine Intelligence 15*(11) :1101–1113.

Zahn, C. T. 1971. Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters. *IEEE Trans. on Computers* C-20 :68–86.

Zanoguera, F. 2001. *Segmentation interactive d'images fixes et de séquences vidéo basée sur des hiérarchies de partitions*. Ph. D. thesis, Ecole des Mines de Paris - Centre de Morphologie Mathématique.

Zanoguera, F., Marcotegui, B., et Meyer, F. 1999. A toolbox for interactive segmentation based on nested partitions. In *proc. of ICIP'99, Kobe (Japan)*.

Zhu, S., Wu, Y., et Mumford, D. 1996. Filters, Random fields And Maximum Entropy (FRAME) - Towards a Unified Theory for Texture Modeling. Int'l Journal of Computer Vision (version finale, à paraître) - Aussi disponible comme rapport technique du Harvard Robotics Lab.

Zhu, S. et Yuille, A. 1995. Region Competition and its analysis : a unified theory for image segmentation. Technical Report 95-07, Harvard Robotics Lab.

Zhu, S. C. et Yuille, A. 1996. Region Competition : Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*(9) :884–900.

Annexes

Annexe A

Sur-segmentation par ligne de partage des eaux

La méthode de segmentation d'image par **Ligne de Partage des Eaux (LPE)** a été introduite par Beucher et Lantuejoul (1979). Elle s'appuie sur une analogie entre une image et un relief topographique, en envisageant les pixels d'une image mono-canal comme autant de sites géographiques et en interprètant les valeurs des pixels comme les altitudes des lieux. Elle applique alors à ce relief les notions géographiques de ligne de partage des eaux et de bassins versants. La LPE est définie par le comportement d'écoulements d'eau imaginaires sur ce relief. En général, si une goutte d'eau tombe en un point donné d'un relief, elle s'écoulera en suivant la ligne de plus grande pente, et terminera sa course au fond d'une cuvette (un minimum local de la surface). On peut ainsi, à chaque point de la surface topographique, associer le lieu d'arrivée d'un ruissellement imaginaire dont il serait la source. On définit alors les *bassins versants* de la surface comme les classes d'équivalence de la relation entre sites *«avoir le même point d'arrivée pour l'écoulement»*. Ils forment donc une partition des sites. La ligne de partage des eaux du relief topographique est alors la frontière des bassins versants. Intuitivement, la LPE est le lieu des points de bifurcation des écoulements, c'est-à-dire ceux sur lesquels une goutte d'eau idéale serait dans une situation instable, pouvant conduire, par un infime perturbation, à son écoulement dans plusieurs bassins différents[1]. La LPE passe par la plupart des lignes de crête de la surface et se ferme au niveau des cols.

La définition continue de la LPE est purement topologique et s'étend naturellement à toute fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ (en abandonnant toutefois l'analogie géographique : les fonctions pour $n > 2$ doivent être vues comme des densités). Cependant sa transposition discrète est délicate : il n'y a pas de procédé unique pour l'effectuer[2].

Une seconde manière d'envisager la LPE consiste à considérer l'immersion progressive du relief : au fur et à mesure que le niveau d'eau monte, différents lacs apparaissent (à chaque fois que le niveau atteint le point bas d'une cuvette), ou se mélangent (à chaque fois que le niveau d'eau atteint un col) pour finir par ne former qu'une mer unique recouvrant toute la surface de l'image. Si l'on empêche les lacs de se mélanger, mais que l'on mémorise l'ensemble des points successifs où un mélange aurait eu lieu, on retrouve la notion de LPE. Cette approche a donné lieu au premier algorithme efficace de calcul de LPE discrète, par simulation d'immersion (Vincent 1990).

[1] On voit que les zones plates posent problème pour cette définition : chacun de leurs points est stable pour l'écoulement, la LPE y est donc dense et surfacique. On peut cependant toujours, par un mécanisme que nous ne détaillerons pas (Préteux 1992), transformer la surface de telle manière qu'elle ne possède plus aucune zone plate (tel que les extrema locaux de la surface forment un ensemble de dimension 1). La LPE passe alors par le squelette des plateaux.

[2] Il semble que ce soit une constante des problèmes de transposition de notions *topologiques* du continu au discret. Il n'existe bien entendu pas de transposition mathématique stricte, et en définitive ce que l'on cherche à transposer est un concept : on cherche à définir des objets discrets qui se comportent de manière *analogue* à leurs homologues continus. Les candidats à l'*analogie* sont souvent nombreux.

D'autres algorithmes de LPE discrète ont été proposés ultérieurement, qui s'appuient l'idée de noyau homotopique, i.e. de transformation maximale conservant la topologie des sections de l'image (Bertrand, Everat, et Couprie 1997).

L'algorithme que nous utilisons pour calculer une LPE discrète est encore autre. Il est du à Marc Pierrot Deseilligny, qui ne l'a jamais publié, mais nous l'a transmis oralement. Contrairement aux algorithmes précédents, la LPE considérée est une ligne *interpixellaire*. L'algorithme est très simple, intuitif, et efficace. Il consiste à construire explicitement le **graphe de ruissellement** associé au relief. L'idée de considérer ce graphe est intimement liée à celle de considérer le «graphe de veines» dans le problème de squelettisation d'ensembles discrets (Pierrot Deseilligny, Stamon, et Suen 1998).

Soit I une image. Notons P l'ensemble des pixels de l'image et munissons-le de la topologie de 8-voisinage modélisée par le graphe symétrique $G = (P, U)$.

Le graphe de ruissellement $R = (P, B)$ est alors le graphe partiel de G dans lequel chaque pixel $p \in P$ possède au plus une arête sortante, celle qui le connecte à son voisin de plus faible altitude. Cette connexion existe si et seulement si l'altitude de ce voisin est moindre que celle du point lui-même :

$$\forall p \in P \quad \forall b \in \mathcal{V}(p) \qquad (p, b) \in B \quad \Leftrightarrow \quad I(b) = \min_{v \in \mathcal{V}(p)} I(v) \quad \text{et} \quad I(v) < I(p).$$

L'arête (p, b) modélise bien le fait qu'une goutte déposée en p s'écoulerait - si elle n'est pas en équilibre - selon la ligne de plus grande pente. Si un pixel p possède plusieurs plus bas voisins, il est connecté au premier d'entre eux dans un ordre arbitraire mais prédéfini (par exemple l'ordre donné par le code de Freeman des voisins).

Ce graphe de ruissellement $R = (P, B)$ est alors symétrisé pour obtenir le graphe $S = (P, B^s)$, où $(x, y) \in B^s \;\Leftrightarrow\; (x, y) \in B \quad \text{ou} \quad (y, x) \in B$.

Les composantes connexes de S définissent les bassins versants du relief.

La figure A.1 illustre les différentes étapes de sur-segmentation d'une image par application de cet algorithme au module du gradient de l'image.

L'algorithme de Marc Pierrot-Deseilligny est très efficace : Un balayage de l'image permet de construire le graphe de ruissellement (la symétrisation est faite dans la foulée) et un second balayage permet d'en extraire les composantes connexes.

Remarquons finalement avec Pierrot Deseilligny et al. (1998) qu'un graphe défini sur les pixels d'une image, lorsqu'il est un graphe partiel du graphe de 8-connexité, peut être représenté par une image d'octets de même taille que l'image d'origine : chaque bit de l'octet attribué à un pixel indique la présence ou l'absence d'une arête vers un des 8 voisins du pixel (si le graphe est symétrique on peut même n'attribuer que 4 bits par pixel).

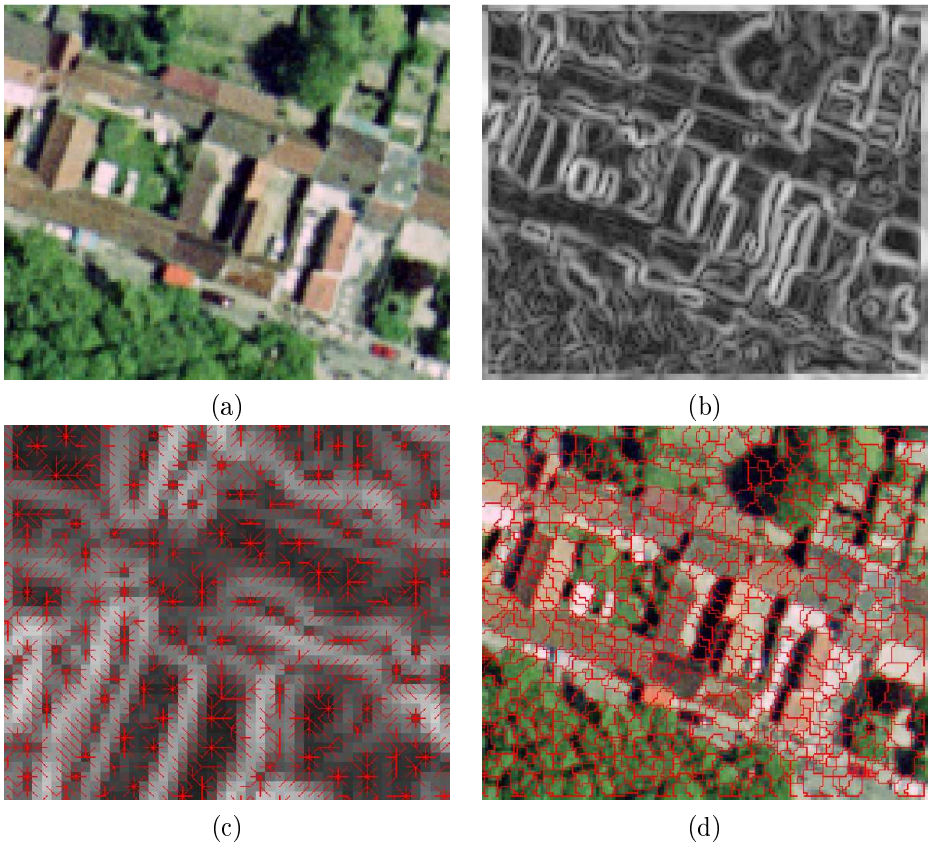


FIG. A.1 – Sur-segmentation par ligne de partage des eaux. (a) Image d'origine. (b) Module de son gradient (Filtre de Canny-Deriche, $\alpha = 2$). (c) Extrait du graphe de ruissellement symétrisé. (d) Ligne de partage des eaux obtenue (frontières des composantes connexes de (c)).

Annexe B

Algorithmes de cocons durs

Cette annexe décrit les algorithmes proposés pour la segmentation d'image par la méthode des cocons durs décrite au chapitre 5. Dans tout ce qui suit, $G = (X, U, d)$ est un graphe attribué par une mesure de dissemblance d.

B.1 Calcul des cocons durs contenant un nœud donné

On note n le nœud de départ de la croissance. Soit Y la pile des nœuds agglomérés à chaque étape. On utilise également une pile d'entiers C destinée à recevoir les numéros des étapes de l'agglomération où l'ensemble Y est un cocon. On note c l'étape courante. Initialement $Y = \{n\}$, $c = 1$ et $C = \{1\}$.

Soit Q la queue de priorité des nœuds voisins de Y triés selon le critère M (équation 5.15 p. 90) et R la queue de priorité des voisins triés selon le critère m (équation 5.16 p. 91). Initialement on insère dans chaque queue l'ensemble des voisins de n avec une priorité égale au poids de l'arête qui les joint. Si v est un nœud présent dans la queue Q (resp. R), la notation $Q(v)$ (resp. $R(v)$) fait référence à sa priorité courante. On note I et E les contrastes interne et externe de Y. Initialement, $I = 0$ et à tout moment E est donné par la valeur du premier élément de la queue R.

Avec ces notations, le coeur de l'algorithme est le suivant :

Algorithme 3 — **Calcul des cocons durs emboîtés sur un nœud**

Tant que la queue Q n'est pas vide, répéter :

1. Retirer le premier élémént p de Q et l'ajouter à l'ensemble Y.
2. Retirer également p de la queue R.
3. Mettre à jour le contraste interne I par l'équation 5.17 p. 91 pour obtenir les cocons durs, ou par l'équation 5.18 p. 91 pour obtenir les cocons complets durs.
4. Pour tout voisin v de p qui n'est pas dans Y, soit $e = [p, v]$ l'arête le joignant.
   (a) Si p est déjà dans Q (et donc dans R) :
      i. Si $d(e) > Q(p)$ affecter $d(e)$ comme nouvelle priorité à p dans Q.
      ii. Si $d(e) < R(p)$ affecter $d(e)$ comme nouvelle priorité à p dans R.
   (b) Sinon insérer p dans Q et R avec la priorité $d(e)$.
5. Incrémenter l'étape courante c de 1.
6. Lire la valeur E du premier élément de la queue R et tester la propriété de cocon : si $I < E$ ajouter c à la pile C des étapes où Y est un cocon.

Fin de l'algorithme

Structures adéquates :
Les opérations élémentaires que l'on souhaite effectuer sur la file d'attente Q sont au nombre de trois :

- Retirer le premier élément de la queue
- Insérer un nouvel élément dans la queue avec une priorité donnée
- Modifier la priorité d'un élément donné

Les deux premières opérations sont implémentées efficacement par une structure d'arbre binaire qui permet des insertions triées en $\log n$. La troisième opération nécessite de pouvoir retrouver rapidement la position d'un élément dans la queue et s'implémente par indexation dynamique des positions des éléments dans l'arbre. Pour des détails concernant la mise en oeuvre de telles structures, voir (Cormen, Leiserson, et Rivest 1994).

Les opérations à effectuer sur la deuxième queue R sont :

- Lire la priorité du meilleur élément de la queue
- Insérer un nouvel élément dans la queue avec une priorité donnée
- Modifier la priorité d'un élément donné
- Supprimer un élément quelconque de la queue

A cause de la nécessité de pouvoir supprimer rapidement un élément aléatoire, il peut sembler plus intéressant d'utiliser pour R une liste chaînée indexée plutôt qu'un arbre binaire indexé. C'est la solution que nous avons adoptée, sans toutefois avoir prouvé sa supériorité.

L'algorithme qui vient d'être décrit permet de déterminer la suite des cocons emboîtés à partir d'un nœud donné, donc de remonter une branche de la hiérarchie des cocons. Pour déterminer l'ensemble de la hiérarchie, il est nécessaire de l'appliquer successivement à l'ensemble des nœuds. Évidemment, en procédant ainsi, on effectuera beaucoup de calculs redondants : à chaque nouveau nœud, de nombreux cocons qui le contiennent ont déja été calculés en remontant depuis des nœuds précédemment visités. Il faut donc détecter les «branchements» sur les chaînes déja calculées de la hiérarchie. Pour cela, le plus simple est certainement de calculer l'ultramétrique associée à la hiérarchie des cocons durs.

B.2 Calcul de l'ultramétrique de cocons

Le calcul de l'ultramétrique de cocons durs δ peut se faire de manière simple en utilisant une variante de l'algorithme B.1 à partir de tous les nœuds du graphe.

Initialement, on affecte à toutes les arêtes une distance ultramétrique égale au max des dissemblances des arêtes, ce qui correspond au contraste interne du graphe entier. Au cours de la croissance, on tient alors à jour la valeur min de la distance ultramétrique sur les arêtes internes de l'ensemble courant. À chaque étape, si le contraste interne de Y excède cette valeur, un branchement sur un cocon qui a déja été atteint auparavant s'est produit et l'on stoppe la croissance. À chaque fois que l'on trouve un nouveau cocon, on affecte aux arêtes rencontrées depuis la construction du dernier cocon une ultramétrique égale au contraste interne courant. On peut pour cela balayer l'ensemble Y du rang $C[c-1]+1$ à $C[c]$ et chercher ses connections vers les éléments du dernier cocon construit (les $Y[i]$ pour $i = 0 \ldots C[c-1]$).

B.3 Calcul des cocons maximaux pour un critère d'homogénéité

Si $G = [X, U, d]$ est un graphe et P est un critère d'homogénéité consistant sur l'ensemble des parties de X, alors la segmentation de G constituée des cocons de G maximaux pour P peut se calculer en utilisant l'algorithme B.1 et un marqueur binaire sur les nœuds du graphe de la manière suivante :

Algorithme 4 **Cocons homogènes maximaux**

1. Initialiser toutes les marques à 0.
2. Pour tout $x \in X$ non marqué, faire :
   (a) Initialiser une nouvelle région $R = \{x\}$. Marquer x.
   (b) Initialiser l'algorithme B.1 de croissance de cocon à partir de x.
   (c) Répéter :
      i. Effectuer un pas d'agglomération de l'algorithme B.1.
      ii. Si aucun nouveau sommet n'est atteint ou si le sommet atteint est marqué aller en 2d.
      iii. Sinon si l'ensemble courant Y est un cocon :
         A. Si Y vérifie le critère P ajouter à R tous les sommets atteints depuis le dernier cocon rencontré et les marquer.
         B. Sinon aller en 2d.
   (d) Ajouter la région R à la partition finale.

Fin de l'algorithme

B.4 Complexité des algorithmes

Cette section analyse la complexité des algorithmes proposés.

On pose $n = |X|$ et $m = |U|$.

L'algorithme B.1 remonte une branche entière de la hiérarchie à partir d'un nœud donné. A chaque étape de croissance, un nouveau nœud est atteint. L'algorithme effectue alors un retrait du premier de la pile Q et un retrait aléatoire dans P. Puis il balaye les arêtes sortantes du nouveau nœud qui n'ont pas été déja vues. Les arêtes du graphe sont donc balayées chacune une et une seule fois. Pour chacune d'elle, il effectue soit deux modifications de priorité, soit deux insertions triées dans des queues qui possèdent autant d'éléments que l'ensemble courant a de voisins. On peut majorer la taille de ces queues par n. Les opérations sur les queues gérées par arbre binaire se font en $\log n$. On arrive donc à une complexité totale de :

$$\mathcal{C}_1 = \mathcal{O}\big((n+m)\log n\big). \tag{B.1}$$

Remarquons que la taille du cocycle de l'ensemble en croissance est majorée par n, ce qui ne peut être atteint que pour des graphes complets et est donc très pessimiste.

L'algorithme B.3 utilise en première approximation n fois l'algorithme B.1. On peut donc majorer sa complexité par

$$\mathcal{C}_2 = \mathcal{O}\big((n^2+nm)\log n\big). \tag{B.2}$$

Toutefois la croissance est interrompue lorsque :

- Des branchements dans la hiérarchie sont détectés. En particulier, lorsqu'un cocon maximal est trouvé par croissance à partir d'un nœud x, toutes les autres branches qui y parviennent depuis d'autres nœuds ne sont pas explorées.
- Le critère P est violé.

La complexité B.2 est donc largement surestimée.

Pour un Graphe d'Adjacence de Régions, la complexité est moindre. En effet, dans un RAG, les nœuds ne sont que très localement connectés. Plus précisément, on sait qu'un graphe *planaire* de n nœuds comporte au plus $6n-3$ arêtes (Jungnickel 1999)[1]. Les graphes d'adjacence topologique de régions de partitions de domaines du plan étant planaires, on a donc

$$\mathcal{C}_1(RAG) = \mathcal{O}(n\log n). \tag{B.3}$$

[1] Cette borne est atteinte pour les maillages hexagonaux dans lesquels les nœuds intérieurs ont tous 6 voisins. Remarquons qu'à l'inverse, les graphes planaires dont les degrés sont très déséquilibrés sont les moins denses.

Le terme en $\log n$ provenait d'une majoration par n du degré des nœuds. Cependant, pour un RAG, le degré moyen des nœuds est majoré par 6. La complexité *moyenne* d'une remontée complète de branche est donc linéaire, c'est-à-dire proportionnelle au nombres d'étapes.

Par ailleurs, dans un certain nombre de problèmes réels de segmentation d'image, on peut raisonablement supposer que la taille des régions homogènes est bornée. Par exemple, pour une image aérienne de résolution donnée, l'emprise du territoire observé a beau croître, la taille des structures homogènes typiques reste limitée (sauf si l'image contient une portion de mer ou un désert...). Cela signifie que dans tous les cas de croissance de cocons, le critère P stoppera la croissance à au plus n_H nœuds, où n_H est la taille maximale d'une région homogène, exprimée en nombre de régions élémentaires. On a alors

$$\mathcal{C}_2(RAG) = \mathcal{O}\big(n(n_H \log n_H)\big) = \mathcal{O}(n). \tag{B.4}$$

Dans de tels cas, la complexité de l'algorithme de segmentation est asymptotiquement linéaire en fonction du nombre de régions élémentaires à fusionner. Comme le nombre de régions élémentaires est également proportionnel au nombre de pixels de l'image, la complexité est asymptotiquement linéaire en fonction du nombre de pixels. Notons que cette linéarité - quand on suppose que la taille des régions homogènes est bornée - provient de la nature *parallèle* de la construction de la hiérarchie : la croissance de cocons s'effectue indépendamment à partir de chaque nœud.

Pour un RAG, la situation la pire au niveau complexité correspond au cas d'une grande région homogène, mais qui n'est pas un cocon, et pour laquelle les cocons inclus sont de petite taille. On obtient une telle configuration dans le cas d'un grand objet homogène sans réelle structure interne, et dont un bord est moins contrasté que certains «bruits» intérieurs. Dans ce cas, toutes les croissances initialisées à l'intérieur de la zone ne trouveront que de petits cocons, mais ne s'arrêteront tout de même pas avant d'avoir parcouru l'ensemble de la grande région. Celle-ci sera donc balayée de nombreuses fois.

Annexe C

Algorithme d'escalade

Cette annexe détaille la méthode de construction d'une description ensembles-échelle par l'algorithme d'escalade binaire connexe, dont le principe est expliqué au chapitre 8. La passe ascendante de la construction s'appuie sur un processus classique de fusion de paires de régions voisines suivant une priorité dynamique. L'implémentation en utilisant un graphe d'adjacence et une queue de priorité correspond globalement à l'algorithme de Takio Kurita (1991, 1995). La différence principale avec l'algorithme de Kurita est que toutes les étapes du processus de fusion sont mémorisées à travers la construction d'un arbre d'inclusions, qui est binaire avant la passe descendante de l'optimisation qui conduit à la hiérarchie persistante. Cette annexe détaille en outre les spécificités liées au principe d'optimisation sur lequel repose l'escalade (calcul par programmation dynamique des énergies partielles). Après la présentation de l'algorithme, nous étudions sa complexité combinatoire.

C.1 Description de l'algorithme

L'algorithme prend en entrée une sur-segmentation S de l'image et un couple (C, D) de fonctions d'énergie sur les régions d'une image. En théorie, C doit être sous-additive, cependant il est toujours possible de «forcer» cette sous-additivité en saturant la valeur calculée de $\mathsf{C}(x \cup y)$ à $\mathsf{C}(x) + \mathsf{C}(y)$ (étape optionnelle 1* de la procédure Calculer_apparition ci-dessous).

Nous considérons un graphe $G = (R, A)$ dynamique dont les nœuds représentent des régions et les arêtes deux types de relations entre régions, des relations d'adjacence et des relation d'inclusion (voir figure C.1). À chaque étape de la segmentation, les relations d'adjacence correspondent aux adjacences topologiques (sans arête multiple) entre régions de la partition courante. Les relations d'inclusion, elles, modélisent les relations fils-père dans la pré-hiérarchie courante.

Remarque 1 : Chaque nœud d'une hiérarchie ayant au plus un père, l'inclusion pourrait se maintenir par une seule référence $\mathsf{p}(x)$ de chaque nœud x vers son père. Cependant, pour l'étape finale de redescente de la hiérarchie, il est nécessaire de connaître l'ensemble des fils $\mathcal{F}(x)$ de chaque nœud. Il est finalement plus simple de considérer les arêtes d'inclusion comme des arêtes à part entière, i.e. des relation bidirectionnelles, et d'utiliser un marqueur booléen sur les arêtes pour distinguer les arêtes d'adjacence de celles d'inclusion.

Remarque 2 : Nous ne décrivons pas ici la modélisation *géométrique* de la partition, par le graphe des frontières de régions (RBG). Un algorithme qui s'appuie sur des énergies géométriques élaborées, par exemple liées à des approximations polygonales des frontières des régions, doit également maintenir la géométrie de la partition courante au cours de la construction de la hiérarchie, par la gestion d'un RBAG dynamique (voir la section 4.3.3 p. 58).

- Chaque nœud x de G possède les attributs suivants

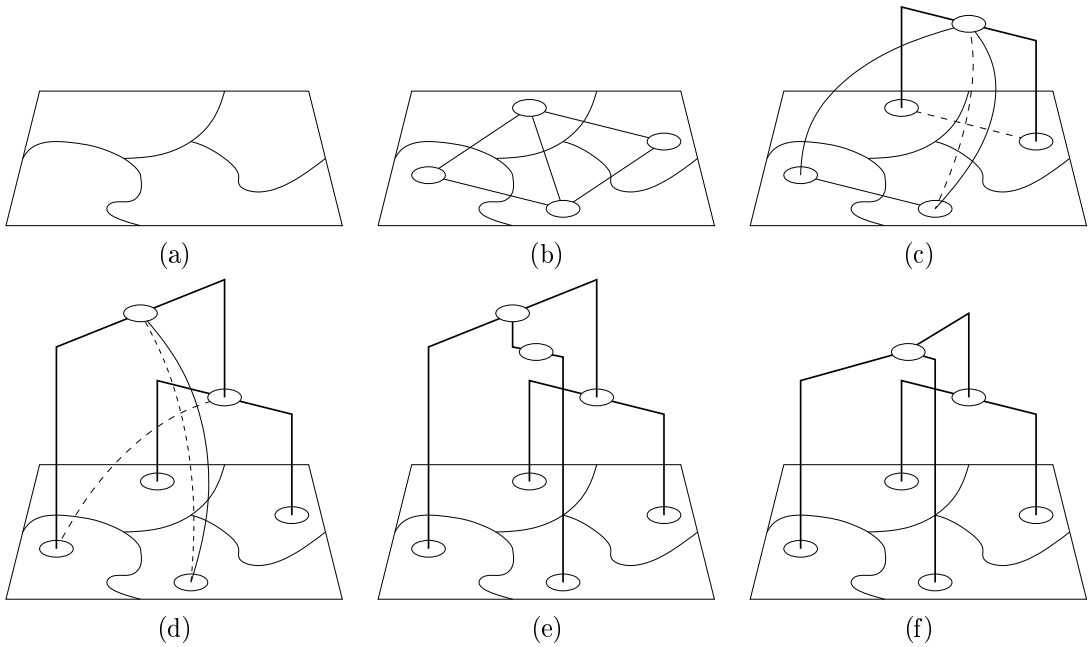


FIG. C.1 – Construction d'une hiérarchie d'escalade. (a) Sur-segmentation initiale (b) RAG correspondant. (c)-(e) Pas d'escalade. (f) Hiérarchie persistante. Les arêtes grasses représentent les arêtes d'inclusion, les arêtes fines celles d'adjacence. Les arêtes en pointillés représentent les arêtes supprimées à chaque étape, dont les arêtes multiples crées par la contraction de deux nœuds en un nœud unique.

- $\mathsf{C}(x)$: énergie de régularisation.
- $\mathsf{D}(x)$: énergie d'attache aux données.
- $\lambda^+(x)$: échelle d'apparition.
- $\mathsf{E}(x)$: fonction affine par morceaux d'énergie partielle (voir ci-dessous).
- Des attributs utilisateur servant à calculer les énergies, e.g. nombre de points, somme des radiométries, somme des carrés des radiométries, périmètre.

• Chaque arête d'adjacence a de G possède les attributs suivants (les arêtes d'inclusion ne portent pas d'attribut)

- $\mathsf{C}(a)$: énergie de régularisation de l'union des deux régions x et y qu'elle joint.
- $\mathsf{D}(a)$: énergie d'attache aux données de $x \cup y$
- $\lambda^+(a)$: échelle d'apparition de $x \cup y$
- $\mathsf{E}(a)$: fonction affine par morceaux d'énergie partielle de $x \cup y$.
- Des attributs utilisateur servant à calculer les énergies, e.g. longueur de la frontière entre x et y.

• Notre implémentation de l'algorithme d'escalade a été réalisée en C++. Nous conservons ici une présentation C++ pour la manipulation des fonctions affines par morceaux, car c'est la façon la plus simple de les décrire.

Une fonction affine par morceaux est une instance d'une classe affine_par_morceaux qui hérite du type list<morceau_affine>, où list est la classe de liste de la STL (standard template library) et morceau_affine est une structure qui comporte simplement trois champs flottants x, y et p. (x, y) représente l'origine du morceau et p sa pente. Nous noterons un morceau_affine comme un triplet (x, y, p).

Deux fonctions sont nécessaires sur ces objets :

– Réaliser l'addition de deux affine_par_morceaux.

La méthode Addition(A, B, S, N) où A, B et S sont des affine_par_morceaux réalise $S = A + B$. Le paramètre N sera expliqué ci-dessous.

– Calculer l'infimum entre une affine_par_morceaux A et un morceau_affine m. L'inf s'effectue «en place», en modifiant l'objet A lui-même.

La méthode Inf(A, m) met à jour A et renvoie l'abcisse du point d'intersection entre A et m, qui représente l'échelle d'apparition d'un nœud.

• La fusion de régions utilise une queue de priorité Q sur les arêtes de G, triée par ordre d'apparition croissante et indexée : chaque arête connait à tout instant sa position dans Q. Comme pour les algorithmes de cocons (cf. annexe B), cette queue de priorité dynamique est efficacement gérée par un arbre binaire (ou tas).

• La construction de la hiérarchie d'escalade peut être stoppée à une échelle maximale λ_{max}.

Si à une étape donnée, toutes les apparitions des complétions possibles sont supérieures à λ_{max}, l'escalade est stoppée.

Algorithme 5 **Escalade**

A. Initialisation

1. Construire le graphe G comme le graphe d'adjacence topologique (sans arête multiple) des régions de S (figure C.1(a)-(b)).
2. Pour chaque nœud x de G :
   (a) Calculer C(x) et D(x).
   (b) E$(x) \leftarrow (0, \mathsf{D}(x), \mathsf{C}(x))$ (L'énergie partielle de x est égale à son énergie propre).
   (c) $\lambda^+(x) \leftarrow 0$ (L'échelle d'apparition des nœuds de base est nulle).
   (d) Calculer et stocker les attributs utilisateurs nécessaires aux calculs d'énergie.
3. Pour chaque arête a de G :
   (a) Calculer_apparition(a).
   (b) Insérer a dans Q avec la priorité $\lambda^+(a)$.

B. Construction

Soit s un nœud, initialement nul.

Tant que Q n'est pas vide, répeter :

1. Retirer la première arête $a = (x, y)$ de Q.
2. Construire un nouveau nœud x, faire $s \leftarrow x$.
3. Si $\lambda^+(a) \geq \lambda_{max}$
   (a) $\lambda^+(s) \leftarrow \lambda_{max}$.
   (b) Pour chaque sommet x d'une composante hiérarchique (nœud sans père) créer l'arête d'inclusion (x, s).
   (c) Aller en **C.**

   Sinon
   (a) $\lambda^+(s) \leftarrow \lambda^+(a)$; C$(s) \leftarrow$ C(a) ; D$(s) \leftarrow$ D(a) ; E$(s) \leftarrow$ E(a)
   (b) Calculer les attributs utilisateur de s qui sont fonction de ceux de x et y, e.g. sommer les cardinaux des régions.
   (c) Détruire l'arête a.

(d) Transformer toutes les arêtes de la forme $(x, .)$ en $(s, .)$ et celles de la forme $(y, .)$ en $(s, .)$.

(e) Éliminer les arêtes multiples sortantes de s. Également les supprimer de Q. Mettre à jour les attributs utilisateurs à chaque fusion de multiples, e.g. sommer les longueurs de frontières.

(f) Créer les arêtes d'inclusion (x, s) et (y, s).

(g) Pour chaque arête b incidente à s :

i. Calculer_apparition(b).

ii. Affecter $\lambda^+(b)$ comme nouvelle priorité de b dans Q.

C. Finalisation

1. Appeler Elaguer_non_persistantes (s,s,$+\infty$).

Fin de l'algorithme

Procédure 1 **Calculer_apparition(arête** a **)**

Soit x et y les nœuds terminaux de a.

1. Calculer $\mathsf{C}(x \cup y)$ et $\mathsf{D}(x \cup y)$.

1*. Éventuellement forcer la sous-additivité de C :
Si $\mathsf{C}(x \cup y) > \mathsf{C}(x) + \mathsf{C}(y)$ faire $\mathsf{C}(x \cup y) \leftarrow \mathsf{C}(x) + \mathsf{C}(y)$.

2. $\mathsf{C}(a) \leftarrow \mathsf{C}(x \cup y)$; $\mathsf{D}(a) \leftarrow \mathsf{D}(x \cup y)$.

3. Addition $(\mathsf{E}(x), \mathsf{E}(y), \mathsf{E}(a), P_{max})$

4. $\lambda^+(a) \leftarrow \mathsf{Inf}(\mathsf{E}(a), (0, \mathsf{D}(a), \mathsf{C}(a)))$

Fin de la procédure

L'étape finale pour parvenir à la hiérarchie persistante pour E_λ est présentée ici de manière récursive :

Procédure 2 **Elaguer_non_persistantes(nœud** n**, nœud** per**, réel** λ_{dis} **)**

1. Si $\lambda_{dis} > \lambda^+(n)$ alors $per \leftarrow n$.

2. $\forall s \in \mathcal{F}(n)$: Elaguer_non_persistantes(s, per, $\min\{\lambda^+(n), \lambda_{dis}\}$)

3. Si $\lambda_{dis} \leq \lambda^+(n)$ shunter n :
   (a) Supprimer l'arête $(n, \mathsf{p}(n))$
   (b) $\forall s \in \mathcal{F}(n)$ dépointer l'arête (s, n) vers (s, per)
   (c) Supprimer n

Fin de la procédure

La procédure intègre le calcul des échelles de disparition et le shuntage des nœuds non persistants. Elle a été optimisée pour que le shuntage d'une chaîne de nœuds non persistants se fasse en une seule fois. Le nœud n est le nœud courant de la récursion. Le nœud per est le plus bas nœud persistant au dessus de n et λ_{dis} est l'échelle de disparition de n.

L'étape 1. change per en n si n est persistant.

L'étape 2. appelle récursivement la fonction sur les fils de n en calculant leur échelle de disparition.

L'étape 3. shunte n s'il n'est pas persistant.

Le shuntage s'effectue lors de la remontée de la récursion, en commençant par les nœuds les plus bas. Un nœud n qui est shunté voit ses fils reconnectés vers le plus bas nœud persistant au dessus de lui, per. On pourrait simplement reconnecter ces fils vers le père f de n, et se passer de la variable per. Cependant, dans les cas où f doit à son tour être shunté, procéder ainsi produit deux reconnexions successives au lieu

d'une seule. Et si toute une chaîne de nœuds doit être shuntée, les arêtes remontent pas à pas jusqu'au plus bas nœud persistant. Dans la procédure ci-dessus, les arêtes sont directement reconnectées à ce nœud.

Décrivons finalement les opérations sur les affine_par_morceaux. Nous donnons ici directement le code C++.

L'addition se fait en commençant par les *derniers morceaux* des affine_par_morceaux et peut être paramétrée pour ne renvoyer au plus que les P_{max} derniers morceau de la somme. Nous verrons que ce point est critique pour l'efficacité de l'algorithme.

```
void Addition ( const affine_par_morceaux& A, const affine_par_morceaux& B,
                affine_par_morceaux& S, int Pmax  )
{
  affine_par_morceau::const_reverse_iterator ia = A.rbegin(),eia = A.rend();
  affine_par_morceau::const_reverse_iterator ib = B.rbegin(),eib = B.rend();
  // Double parcours des morceaux de A et B, en sens inverse
  while ( (ia!=eia) && (ib!=eib) && (S.size()<Pmax) )
  {
    // Si l'origine du morceau courant de A est >= a celle du morceau courant de B
    // Un morceau doit etre cree avec cette origine
    if ( (*ia).x >= (*ib).x ) {
      morceau nouveau = *ia;
      // On ajoute a l'ordonne du nouveau l'ordonnee de (*ib) au point (*ia).x
      nouveau.y += (*ib).y + (*ib).p * ( (*ia).x - (*ib).x );
      // Ajout des pentes
      nouveau.p += (*ib).p;
      // Empilement au debut de la somme
      S.push_front ( nouveau );
      // Si les debuts des morceaux sont egaux : il faut aussi passer
      // au morceau precedent de B
      if ( (*ia).x == (*ib).x ) ++ib;
      ++ia;
    }
    // Memes operations mais en inversant les roles de (*ia) et (*ib)
    else {
      morceau nouveau = *ib;
      nouveau.y += (*ia).y + (*ia).p * ( (*ib).x - (*ia).x );
      nouveau.p += (*ia).p;
      S.push_front ( nouveau );
      ++ib;
    }
  }
  // Remise en 0 de l'origine du premier morceau :
  // si l'addition a ete stoppee par Pmax, le dernier morceau ajoute peut avoir x>0
  if (S.front().x>0) {
    S.front().y -= S.front().p * S.front().x;
    S.front().x = 0;
  }
  // FIN
}
```

Décrivons enfin la fonction Inf :

```
float Inf( affine_par_morceaux& A, const morceau& m )
{
  // utilisation d'un iterateur bidirectionnel comme un reverse_iterator
  affine_par_morceau::iterator i = A.end(),ei = A.begin();
  --i; --ei;
  // cas particulier de fonction de cout additive :
  // les pentes de m et du dernier morceau de A sont alors egales
  if ( m.p == (*i).p ) {
    // ordonnee de m en (*i).x
    float y = m.y + m.p * ( (*i).x - m.x );
    // Si m est au dessus de (*i) : intersection en +infini
    if ( y > (*i).y ) {
      return INFINI_POSITIF;
    }
    // Si m est confondu a (*i) : apparition au debut de (*i)
    else if ( y == (*i).y ) {
      return (*i).x;
    }
    // Sinon, l'intersection se situe sur un morceau precedent,
    // on passe le dernier morceau et on continue
    else {
      --i;
    }
  }
  // Recherche du morceau sur lequel se situe l'intersection
  float xi;
  for ( ; i != ei ; --i ) {
    xi = ( m.x*m.p -(*i).x*(*i).p - (m.y-(*i).y) ) / (m.p - (*i).p );
    if ( xi > (*i).x  ) break;
  }
  // Suppression des morceaux suivant l'intersection
  A.erase ( ++i, A.end() );
  // Insertion nouveau morceau final
  morceau m1( xi, m.y+m.p*(xi-m.x), m.p );
  A.push_back(m1);
  // FIN
  return xi;
}
```

C.2 Complexité de l'algorithme

Notons N le nombre de nœuds du graphe initial. À chaque étape le nombre de nœuds décroit d'une unité. Si l'on construit la hiérarchie jusqu'au sommet, il y a N étapes. Notons M_i le nombre d'arêtes du graphe d'adjacence correspondant à l'étape où il y a i nœuds dans le graphe. Les graphes d'adjacence étant planaires, on a $M_i \leq 6i - 3$ (cf. p.285).

Notons K le coût de calcul de C et de D pour une région x. Il est naturel de supposer que ce coût ne dépend que de la taille de x. Pour toutes les mesures s'appuyant sur des statistiques «aggrégatives», e.g. histogramme, moments, périmètre, ce coût est constant. Par contre pour les mesures nécessitant une analyse complète de chaque région, e.g. une polygonisation ou une régression sur les valeurs image, le

coût de l'analyse dépend au moins linéairement de la taille des régions ou de leur périmètre. Le coût K augmente donc au fur et à mesure que l'on monte dans la hiérarchie. Par ailleurs, si l'on suppose que les tailles des régions de base de la hiérarchie sont à peu près homogènes, alors la taille d'une région dans la hiérarchie est proportionnelle au nombre de régions élémentaires qui la composent. Finalement, nous faisons l'hypothèse que la hiérarchie croît de manière relativement régulière, i.e. qu'à un moment quelconque de l'algorithme, les régions dont on a à mesurer l'énergie ont à peu près la même dimension. Nous exprimons donc le coût K comme une fonction K_i du nombre de nœuds dans le graphe courant (attention K_i décroit en fonction de i, il croit en fonction de la taille $|X|/i$ des régions, où X est le support de l'image). Notons $d(x)$ le degré d'un nœud x dans un graphe d'adjacence courant.

La complexité d'une escalade se décompose alors comme suit :

Opération		Coût
• **Initialisation de l'algorithme :**		
- Construction du RAG	:	linéaire par rapport à la taille de l'image et négligeable, cf. e.g. (Shapiro 1996).
- Attribution des éléments	:	$(N + M_N)K_N = \mathcal{O}(NK_N)$
- Tri de la queue	:	$M_N \log M_N = \mathcal{O}(N \log N)$
• **Construction de la hiérarchie :** Pour $i = N...1$		
- Retrait de la première arête $a_i = (x_i, y_i)$ de Q	:	$\mathcal{O}(\log M_i)$
- Mise à jour du graphe	:	$\mathcal{O}(d(x_i) + d(y_i))$
- Pour chaque arête incidente à $z_i = x_i \cup y_i$	:	
- Calcul de l'énergie	:	K_i
- Calcul de l'apparition de l'union	:	A_i (voir ci-dessous)
- Modification de priorité	:	$T(M_i)$ (voir ci-dessous)

• **Finalisation** : la procédure récursive est de toute évidence linéaire.

On obtient donc une complexité totale en $\mathcal{O}$ de

$$NK_N + N \log N + \sum_{i=1}^{N} (\log M_i + d(x_i) + d(y_i) + d(z_i)(K_i + A_i + T(M_i))) .$$

Les deux premiers termes - qui reflètent le coût d'initialisation - sont négligeables (en théorie et en pratique) devant les autres opérations. Par ailleurs, le terme prépondérant de la somme est son dernier terme. On obtient alors une complexité en $\mathcal{O}$ de

$$\sum_{i=1}^{N} d(z_i)(K_i + A_i + T(M_i)).$$

Posons $C_1 = \sum_{i=1}^{N} d(z_i)K_i$, $C_2 = \sum_{i=1}^{N} d(z_i)A_i$ et $C_3 = \sum_{i=1}^{N} d(z_i)T(M_i)$. C_1 représente le coût lié aux calculs d'énergie, C_2 celui lié à la programmation dynamique fonctionnelle et C_3 correspond au coût de maintien de la liste triée sur les arêtes.

Commençons par le terme C_3.

C.2.1 Part structurelle de la complexité

Pour l'étape à i nœuds, on peut majorer - de manière extrêmement pessimiste - le degré des nœuds par i et les opérations de changement de priorité dans la queue par $\log M_i$ et donc par $\log 6i$. Ainsi,

$$C_3 = \mathcal{O}\left(\sum_{i=1}^{N} i \log i\right) = \mathcal{O}\left(N^2 \log N\right) .$$

Ce cas extrême s'obtient quand l'image comporte (après quelques itérations) une seule grande région F (le fond) constellée de N petites régions qui la «trouent». Le graphe a l'apparence d'une fleur : F est relié aux N régions et chacune d'elle n'est reliée qu'à F. F absorbe alors une à une les régions et chaque absorbtion nécessite de réévaluer les coûts de fusion de F avec toutes les petites régions et de retrier pratiquement complètement la queue. C'est ici le degré $d(z_i)$ de la région en croissance qui est pénalisant.

Le cas le pire pour la complexité d'une escalade correspond finalement à avoir à analyser une image de bruit. C'est vérifié en pratique, voir la section 9.1.3 p. 219 dans le corps du document.

Notons que la majoration de $d(z_i)$ par i correspond également à la complexité effective d'un algorithme qui autoriserait les *régions non connexes*.

Toutefois, sur des images «normales» le degré des nœuds est relativement faible et la majoration de $d(z_i)$ par i est extrêmement pessimiste. Plus précisément, dans le cas de régions connexes (graphe d'adjacence planaire), le *degré moyen* des nœuds est borné par 6. Ainsi, **en moyenne**, $\overline{C_3} = \mathcal{O}(N \log N)$.

En outre, le terme $\log N$ correspond au coût de mise à jour de la position d'une arête dans la queue de priorité. Il est cependant très improbable qu'une arête change radicalement de priorité et se trouve déplacée très loin dans la queue. En pratique, pour des images «normales», la complexité combinatoire liée au terme C_3 est quasi-linéaire.

C.2.2 Coût des calculs d'énergie

Voyons maintenant le terme de complexité C_1, qui correspond aux calculs d'énergie.

Énergies aggrégatives

Si le coût de calcul d'énergie K_i est constant au cours de l'algorithme, la majoration pessimiste de $d(z_i)$ par i donne $C_1 = \sum_{i=1}^{N} d(z_i)K_i = \mathcal{O}(N^2)$. Remplacer $d(z_i)$ par un degré moyen donne $\overline{C_1} = \mathcal{O}(N)$.

C_1 est donc toujours inférieur à C_3 - en théorie comme en pratique, dans le pire des cas comme en moyenne.

Énergies non-aggrégatives

Si le coût K_i varie en fonction de i, il faut faire une hypothèse sur la manière dont il évolue. Typiquement si K_i évolue proportionnellement à la surface des régions à l'étape i, on trouve, pour la majoration pessimiste de $d(z_i)$: $C_3 = \mathcal{O}(N^3)$ et en étant plus réaliste : $\overline{C_3} = \mathcal{O}(N^2)$.

C_1 est donc toujours supérieur à C_3.

C.2.3 Complexité de la programmation dynamique fonctionnelle

Voyons maintenant le terme de complexité $C_2 = \sum_{i=1}^{N} d(z_i)A_i$.

Commençons par nous intéresser au terme A_i, qui correspond au coût d'un pas de programmation dynamique fonctionnelle. Un tel pas fait la somme de deux fonctions affines par morceaux $\mathsf{E}(x_i)$ et $\mathsf{E}(y_i)$ puis calcule l'inf de $\mathsf{E}(x_i) + \mathsf{E}(y_i)$ avec un morceau affine m. Notons $|f|$ le nombre de morceaux d'une fonction f. En examinant les procédures de calcul sur les affines par morceaux données ci-dessus, on constate aisément que

$$A_i = \mathcal{O}\left(|\mathsf{E}(x_i)| + |\mathsf{E}(y_i)|\right)$$

Où $|\mathsf{E}(x)|$ désigne le nombre de morceaux dont est composée $\mathsf{E}(x)$. Le problème est alors que $|\mathsf{E}(x)|$ croît avec la hauteur de x dans la hiérarchie. Plus précisément,

$$|\mathsf{E}(x)| = \mathcal{O}(|x|) \tag{C.1}$$

où $|x|$ désigne le nombre de régions élémentaires dont x est composé (cardinal de la base de la hiérarchie partielle $H(x)$). En effet, $|f+g| \leq |f| + |g| - 1$ (la relation est stricte si certaines extrêmités de morceaux de $|f|$ et $|g|$ coïncident). Par ailleurs, au pire, $|\inf(f,m)| = |f| + 1$ quand l'intersection entre f et m se trouve sur le dernier morceau de f. En propageant les cas d'égalité - les plus défavorables - dans une hiérarchie quelconque, on trouve la relation C.1. C'est illustré par la figure C.2. Pour les hiérarchies

binaires - ce qui correspond à notre cas - les deux bornes extrêmes pour la complexité C_2 correspondent alors à nouveau aux hiérarchies parfaitement équilibrées (BE) et à celles totalement déséquilibrées (BD) (cf. chapitre 7, pp. 158-160).

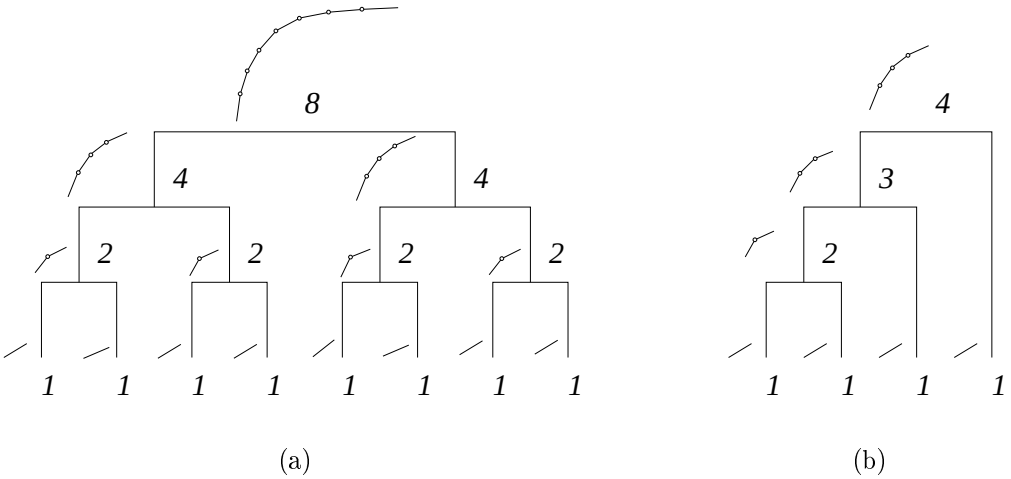


FIG. C.2 – Croissance du nombre de morceaux des fonctions affines lors d'une programmation dynamique fonctionnelle. (a) Hiérarchie binaire parfaitement équilibrée de base 8 (BE_8). (b) Hiérarchie binaire totalement déséquilibrée de base 4 (BD_4). Le nombre figurant à côté de chaque nœud indique le nombre maximal de morceaux affines de la fonction d'énergie partielle du nœud. La forme de ces énergies est schématisée en regard.

Combinons maintenant ce terme A_i avec $d(z_i)$ pour obtenir le terme de complexité C_2.

1. Cas des hiérarchies BD : Majorons comme ci-dessus $d(z_i)$ par i. Quand le graphe a N nœuds, on en fusionne un et on réevalue l'apparition de son union avec les $N-2$ nœuds restants. Ces unions ont trois éléments. On en sélectionne une (car ici on suppose une croissance de type BD de l'arbre), on recalcule l'apparition de $N-3$ nouvelles unions à 4 éléments, etc. On a donc

$$C_2(BD) \quad = \quad \sum_{i=2}^{N-1}(N-i)(i+1) \quad = \quad \mathcal{O}(N^3).$$

Notons que l'on a supposé que le degré du nœud croissant est systématiquement maximal, ce qui est évidemment faux, au moins à l'initialisation. Au contraire au début, sur un graphe de 4-voisinage ou un RAG construit sur LPE, le degré des nœuds vaut au plus 4 ou peut être majoré par une constante faible.

2. Cas des hiérarchies BE : Décomposons la croissance d'une hiérarchie de structure BE_N (avec $N = 2^p$) en p étapes d'agglomération sur des graphes G_i de tailles $t_i = N/2^{i-1} = 2^{p-i+1}$. Supposons que la structure se construit niveau par niveau, i.e. qu'à partir d'un graphe G_i à t_i éléments, $t_i/2$ fusions de paires distinctes s'effectuent, conduisant à un graphe G_{i+1} à $t_i/2$ éléments, etc. La croissance se fait donc en parallèle à chaque niveau. Étudions l'étape à t_i éléments.

Les nœuds de G_i portent des affines par morceaux ayant 2^{i-1} morceaux.

Soit G_i^j le graphe obtenu après j fusions de paires de nœuds de G_i. Après la $j^{\text{ème}}$ fusion d'une paire z de G_i, on recalcule l'apparition de $z \cup x$ pour tout autre élément x de G_i^j.

Parmi ces éléments, $2(j-1)$ sont le résultat d'agrégations d'une paire d'éléments de G_i qui s'est effectuée à une étape précédente ($< j$). L'union $z \cup x$ comporte alors 4 éléments de G_i, 2 provenant de la paire z et 2 de x. La calcul d'apparition de $z \cup x$ a donc un coût de $4 \times 2^{i-1}$ (nombre de morceaux de l'affine résultante). Les autres éléments de G_i^j, en nombre $t_i - 2j$, sont des nœuds de G_i qui n'ont pas été agrégés à une étape avant j. Pour ces éléments, le calcul d'apparition a un coût de $3 \times 2^{i-1}$.

Cela nous conduit donc à un coût total de

$$\begin{aligned}
C_2(BE) &= \sum_{i=1}^{p}\sum_{j=1}^{2^{p-i}} 4\times 2^{i-1}\times 2(j-1)+3\times 2^{i-1}\times(2^{p-i+1}-2j)\\
&= \sum_{i=1}^{p} 2^i \sum_{j=1}^{2^{p-i}} 4(j-1)+3(2^{p-i}-j)\\
&= \sum_{i=1}^{p} 2^i \left(\sum_{j=1}^{2^{p-i}}(j-4)+3\sum_{j=1}^{2^{p-i}} 2^{p-i}\right)\\
&= \mathcal{O}\left(\sum_{i=1}^{p} 2^i\left((2^{p-i})^2+3(2^{p-i})^2\right)\right)\\
&= \mathcal{O}\left(\sum_{i=1}^{p-1} 2^{2p-i}\right) \quad = \quad \mathcal{O}\left(2^p\sum_{i=0}^{p-1} 2^i\right)
\end{aligned}$$

Soit finalement comme $2^p = N$ et $\sum_{i=0}^{p-1} 2^i = \mathcal{O}(N\log N)$

$$C_2(BE) \quad = \quad \mathcal{O}\left(N^2\log N\right).$$

Les hiérarchies construites en pratique se situent quelque part entre le cas BE et BD, le coût C_2 de programmation dynamique fonctionnelle se situe donc entre $N^2\log N$ et N^3.

Notons que pour le calcul de C_2, nous avons à nouveau fait l'hypothèse - pessimiste pour un algorithme à régions connexes - d'une connexion complète entre chaque nœud qui vient d'être fusionné et le reste du graphe. Cette hypothèse introduit un facteur N dans C_2. En analyse moyenne, $\overline{C_2}$ se situe donc entre $N\log N$ et N^2.

Notons que cette complexité *n'est pas une majoration*. Les termes $N\log N$ ou N^2 correspondent au nombre effectif de calculs nécessaires pour l'optimisation par programmation dynamique dans le cas le plus favorable (répartition homogène des degrés).

C.2.4 Optimisation par troncature des sommes d'affines par morceaux

On peut toutefois éliminer la prépondérance du coût C_2 dans la complexité de l'algorithme en ne calculant que des sommes partielles entre affines par morceaux, tel que c'est présenté ci-dessus dans la procédure Addition.

L'idée est que l'intersection entre $\mathsf{E}(x)+\mathsf{E}(y)$ et l'énergie propre de $x\cup y$ se situe normalement au niveau des *derniers morceaux* de $\mathsf{E}(x)+\mathsf{E}(y)$. En effet, si elle se situe sur l'avant dernier morceau, alors ou bien x ou bien y n'est pas persistant ; si elle se situe sur le morceau $n-2$, x et y ne sont pas persistants ; pour $n-3$, c'est un des fils de x et y qui disparaît, etc. A priori, seuls les derniers morceaux sont utiles. Il faut toutefois prendre de la marge car des morceaux précédents peuvent devenir utiles à une étape ultérieure.

En pratique, nous bornons donc le nombre de morceaux calculés dans chaque somme par un paramètre P_{max}, de l'ordre de quelques dixaines de morceaux (voir en section 9.1.3 p. 219 des exemples de nombres maximaux de morceaux utiles sur des cas réels). Le coût A_i de calcul d'apparition devient alors constant. En reprenant les calculs ci-dessus, on constate alors que l'on gagne un facteur N sur C_2, aussi bien pour des hiérarchies BE que BD.

C.2.5 Bilan

Le tableau C.1 dresse le bilan des différentes complexités que nous venons d'évaluer.

Part de complexité	Type d'algorithme	Pire des cas (image de bruit pur)	Cas moyen (image normale)
Calcul d'énergie (C_1)	E aggrégative : K_i constant E non aggrég. : $K_i \propto$ surface ...	N^2 N^3	N N^2
Prog. dyn. (C_2)	Non optimisé Optimisé	$N^2 \log N$ (BE) à N^3 (BD) $N \log N$ (BE) à N^2 (BD)	$N \log N$ à N^2 $\log N$ à N
Structurelle (C_3)	-	$N^2 \log N$	N à $N \log N$

TAB. C.1 – Complexité combinatoire d'une escalade (bilan).

En conclusion, si l'algorithme est optimisé alors pour une énergie aggrégative (e.g. Mumford-Shah), on a au pire une complexité en $N^2 \log N$ et en pratique sur des images «normales», une complexité quasiment linéaire. On voit que l'optimisation de la programmation dynamique fonctionnelle est ici primordiale pour aboutir à un algorithme efficace. Si on ne la fait pas, on obtient dans le meilleur des cas une complexité *effective* qui varie entre $N \log N$ et N^2.

Notons que si l'on souhaite obtenir la fonction complète donnant l'énergie des coupes optimales de la hiérarchie - par exemple pour des problèmes de compression optimale - on *ne peut pas* réaliser cette optimisation, il faut effectuer les sommes exactes des énergies partielles. Remarquons cependant que l'on peut réaliser cette optimisation exacte *après* l'escalade, une fois la hiérarchie construite, ce qui divise le temps de calcul par le degré moyen des nœuds au cours de l'escalade (car si la hiérarchie est déja construite, on n'a pas à évaluer l'apparition des hypothèses de fusion qui ne seront jamais réalisées). Cela ne change pas les complexités théoriques qui sont des «grand O» mais cela divise en pratique le temps par un facteur non négligeable (12 en moyenne).

Si l'énergie n'est pas aggrégative, c'est le calcul d'énergie qui est pénalisant, en théorie comme en pratique, que l'algorithme soit optimisé ou non. Typiquement, si l'on doit polygoniser les frontières de chaque hypothèse de fusion, ce sont ces polygonisations qui nécessitent le plus important temps de calcul. On aboutit en pratique à un algorithme qui est approximativement quadratique sur des images «normales».

La section 9.1.3 p. 219 illustre cette analyse par des exemples réels de temps de calcul.

Annexe D

Algorithme d'approximation polygonale

Nous décrivons ici l'algorithme d'approximation polygonale que nous avons utilisé. On peut le voir comme un algorithme de segmentation à base de prédicats d'homogénéité.

Si $M = (m_0, \ldots, m_{n-1})$ est une séquence ordonnée de points du plan - le polygone à approximer - nous en cherchons une partition en région connexes 1D, i.e. en segments. Notons que si la courbe à approximer est fermée, M possède une topologie torique : m_0 est voisin de m_{n-1}. Nous introduisons alors explicitement les points frontières entre régions. Une partition consiste ainsi en un ensemble de segments séparés par des interfaces ponctuelles, i.e. est de la forme $P = ([m_{n-1}, m_1]m_2[m_3, m_4]m_5[m_6, m_{10}] \ldots m_{n-2})$.

Une telle partition est totalement déterminée par ses points frontières (en 1D, la dualité topologique régions/frontières est complète). Nous recherchons donc finalement $F \subset M$, les frontières de la partition, ou d'un point de vue géométrique, les points retenus pour approximer M. Pour l'exemple de la partition qui précède $F = (m_2, m_5 \ldots, m_{n-2})$.

Notons F sous la forme d'un sous-ensemble d'indices de $\{0 \ldots n-1\}$:

$$F = \{f_0 \ldots f_{k-1}\} \text{ représente les frontières } (m_{f_0} \ldots m_{f_{k-1}}).$$

F définit k régions, frontières inclues, représentées par des plages d'indices :

$$R_i = [f_i, f_{i+1[k]}] \qquad i = 0 \ldots k-1.$$

L'«énergie» ou «distorsion» d'une partition F est alors définie par

$$\mathsf{E}(F) = \max_{i=0}^{k-1} \mathsf{E}(R_i).$$

Si $R = [a, b]$ est une région, son énergie mesure l'écart de ses points intérieurs $\overset{\circ}{R} = \{m_{a+1} \ldots, m_{b-1}\}$ à la droite (m_a, m_b). Cet écart est donné par la variance du nuage $\overset{\circ}{R}$ prise dans la direction normale à (m_a, m_b). Notons Σ_R la matrice de variance/covariance des points de $\overset{\circ}{R}$, centrée sur $\frac{m_a+m_b}{2}$. Notons n_R le vecteur normal à (m_a, m_b). Alors

$$\mathsf{E}(R) = n_R^t \Sigma_R^{-1} n_R.$$

Finalement, l'objectif de l'approximation est de trouver une partition de distorsion bornée par σ^2 et de cardinal minimal.

La stratégie est une stratégie d'escalade qui part de la sur-partition $F = \{0 \ldots n-1\}$ et supprime itérativement la frontière qui conduit à la région union de plus faible énergie.

En pratique, nous définissons une liste chaînée circulaire L d'objets de type frontière. Si $F \in L$, nous notons F^- et F^+ respectivement l'élément précédant et suivant F dans L.

Chaque frontière F représente un point de M et la région qui suit ce point, i.e. la région délimitée par F et F^+. Notons $R(F) \subset M$ l'ensemble des points de M que contient cette région (cet ensemble n'est pas stockée dans les objets frontière, il s'agit juste d'une notation pour nous permettre de décrire l'algorithme).

Les composantes d'un objet frontière sont :

- $F.P$: Point frontière.
- $F.N = |R(F)|$.
- $F.S_{xx} = \sum_{p \in R(F)} p_x^2$.
- $F.S_{yy} = \sum_{p \in R(F)} p_y^2$.
- $F.S_{xy} = \sum_{p \in R(F)} p_x p_y$.
- $F.prio$: Priorité de suppression de F.

L'algorithme utilise alors une queue de priorité Q sur les frontières à supprimer de la liste. La queue est indexée pour avoir directement accès à la position d'une frontière dans la queue.

Algorithme 6 **Approximation polygonale(** $M = (m_0 \dots m_{n-1})$, *torique*, σ^2 **)**

Initialisation

- Pour $i = 0 \dots n-1$: insérer une nouvelle frontière F dans L, avec $F.P = m_i$, $F.N = F.S_{xx} = F.S_{yy} = S_{xy} = F.prio = 0$.
- Si *torique* est faux mettre le champ *prio* du premier et du dernier élément de L à -1 (marquage comme non suppressible).
- Pour chaque $F \in L$: maj_priorite(F).

Fusion

Tant que Q n'est pas vide, répéter :

- Sortir le premier élément F de Q.
- Soit $P = F^-$, faire : $\begin{array}{lcl} P.N & += & F.N + 1 \\ P.S_{xx} & += & F.S_{xx} + F.P_x^2 \\ P.S_{yy} & += & F.S_{yy} + F.P_y^2 \\ P.S_{xy} & += & F.S_{xy} + (F.P_x)(F.P_y) \end{array}$
- Supprimer F de L.
- maj_priorite(P).
- maj_priorite(P^+).

Fin : Renvoyer la séquence de points frontières restant dans L.

Fin de l'algorithme

Procédure 3 **maj_priorite(F)**

- Si $F.prio < 0$: retour.
- Calculer la matrice de variance/covariance Σ associée à $F^- \cup F.P \cup F$. La centrer sur $\frac{F^-.P + F^+.P}{2}$.
- Calculer le vecteur n normal à $(F^-.P, F^+.P)$.
- Calculer $F.prio = n^t \Sigma^{-1} n$.
- Si $F.prio \leq \sigma^2$
  - Si $F \in Q$: changer la position de F dans Q.
  - Sinon insérer F dans Q.
- Sinon
  - Si $F \in Q$: retirer F de Q.

Fin de la procédure

L'algorithme a une complexité en $n \log n$ (coût d'un tri rapide de n éléments).