

UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE  
HEUDIASyC

# THÈSE

présentée en première version en vue d'obtenir le grade de Docteur,  
spécialité « Technologies de l'Information et des Systèmes »

par

Karim YOUSFI

## SEGMENTATION HIÉRARCHIQUE OPTIMALE PAR INJECTION D'*A Priori* : RADIOMÉTRIQUE, GÉOMÉTRIQUE OU SPATIAL

Thèse soutenue le « 3 novembre 2008 » devant le jury composé de :

M.	JEAN MARC CHASSERY	Institut National Polytechnique de Grenoble	(Rapporteur)
M.	MATTHIEU CORD	Université Pierre et Marie Curie	(Rapporteur)
M.	JEAN PIERRE COCQUEREZ	Université de Technologie de Compiègne	(Directeur de thèse)
M.	CHRISTOPHE AMBROISE	Université d'Evry	(Directeur de thèse)
M.	HENRI MAITRE	TELECOM ParisTech	(Examinateur)
M.	THIERRY DENOEUX	Université de Technologie de Compiègne	(Examinateur)



*A ma famille, à Nora*



# REMERCIEMENTS

**U**n travail de recherche a toujours un début mais ne s'achève jamais car la satisfaction de soi-même est rarement atteinte. On n'en voit pas la fin car au fur et à mesure que l'on avance dans notre réflexion, les interrogations et les voies que l'on souhaite explorer se multiplient de manière exponentielle - j'ai envie de dire vertigineuse -. La curiosité de l'esprit et sa volonté à vouloir tout comprendre et tout solutionner nous poussent à nous surpasser. En ouvrant une voie, c'est un arbre de voies qui nous attend et que l'on ne peut explorer tout seul et indéfiniment. Escalader cet arbre pour en étudier une branche est un objectif en soi ambitieux et à notre portée. On y reccueillera inévitablement les fruits qui ont muris au bout de cette branche.

Ce mémoire est une réflexion en amont sur la représentation hiérarchique d'une image. Sa construction s'appuie sur des connaissances haut-niveau liées à l'application. L'objectif est d'améliorer la qualité de la hiérarchie grâce à l'injection d'informations adaptées au contenu de l'image. Les contributions apportées sont le fruit d'un travail à la fois personnel et d'équipe. Durant cette thèse, j'ai rencontré et discuté avec différentes personnes du laboratoire ou à l'extérieur, dans mon environnement professionnel ou personnel. Ils ont de près ou de loin vécu avec moi cette période durant laquelle les certitudes et les doutes s'alternaient. Chacune d'entre elles m'a permis d'avancer d'une manière ou d'une autre. Pour moi il n'y a pas participation faible ou forte, chaque personne a joué un rôle prépondérant durant cette thèse que cela soit par le biais d'une discussion scientifique, banale, ou d'un soutien moral. Ainsi, je souhaite saisir cette occasion qui m'est offerte pour remercier toutes ces personnes et leur témoigner ma gratitude.

Mes premiers remerciements vont tout d'abord à mes directeurs de thèse. A Jean Pierre Cocquerez, je souhaite renouveler mes remerciements pour la confiance qu'il a placé en moi et mes capacités à mener à bien ce travail. A Christophe Ambroise, je voudrai également témoigner toute ma reconnaissance et ma gratitude pour l'attention et l'intérêt qu'il a porté à mon égard et à mes travaux. A mes deux directeurs de thèse je souhaiterai les remercier pour les conseils scientifiques qu'ils m'ont prodigués tout au long de cette thèse et pour leurs remarques pertinentes qui ont guidées ces travaux. Je tiens aussi à souligner les grandes qualités humaines qui les caractérisent dont la patience et la compréhension sont les plus importantes à mes yeux.

Je remercie Jean Marc Chassery, Matthieu Cord, Thierry Denoeux et Henri Maitre pour avoir accepté de participer à mon jury. Je souligne également l'intérêt et l'oeil critique qu'ils ont porté à ce travail. J'ai été très honoré de conclure ce travail de recherche devant ce jury et je souhaiterai tout particulièrement relever la qualité des remarques faites par mes rapporteurs Jean Marc Chassery et Matthieu Cord.

Durant ce travail de recherche, le soutien d'une autre personne m'a été plus que précieux. A Djamal Boukerroui, maître de conférence au sein du même laboratoire,

je souhaite le remercier tout particulièrement pour sa grande disponibilité et son ouverture d'esprit. Nos discussions m'ont permis d'enrichir mes connaissances et je ne le remercierai jamais assez pour avoir partagé avec moi ses connaissances. Inconsciemment ou conscientement, ses remarques et ses suggestions ont contribué à éclaircir mes réflexions.

Je souhaite remercier tout mes collègues et amis du laboratoire **HeuDiaSyC** qui se sont toujours montrés disponibles, à l'écoute et d'un soutien permanent tout au long de ces années. La plupart ont quitté le laboratoire et d'autres poursuivent toujours leurs recherche. Je pense tout paticulièrement à Hassan, Wala, Vlad, Andra, Fares, Manel, Mounia mais également à David, Samis et Alina.

Il est important pour moi de remercier toute ma famille pour leur soutien. Je pense à ma mère et aux inquiétudes qui l'ont habitées durant ces années. Je souhaite lui dédier cette thèse, elle qui a été mon premier professeur de mathématiques, de physique, de français, ...!!!. Elle qui m'a transmis le goût et l'importance de s'instruire. Je voudrai remercier également ma soeur pour sa disponibilité et sa générosité, elle qui à l'heure où j'écris ces mots prépare également la soutenance de sa thèse. A mon père qui n'a jamais douté de moi et s'est toujours préoccupé de l'état d'avancement de mon travail. Je ne voudrai pas oublier de remercier mes deux frères qui ont été d'un grand soutien moral. Ils ont égayé mes étés et m'ont permis de relativiser les choses.

Enfin, je pense qu'il y a une personne pour laquelle mes remerciements seront éternels. Elle n'a jamais sourcillé lorsque je me suis montré égoïste durant mes années de recherches. Elle a toujours fait preuve de patience et de compréhension. Elle a toujours été d'un grand réconfort et sa présence me redonnait confiance et motivation. Je pense à Nora, sans elle j'aurai certainement suivi un autre chemin

...

# RÉSUMÉ

La représentation hiérarchique d'une image fournit une description riche du contenu de la scène. La construction d'une hiérarchie de régions est généralement un processus ascendant guidé par un mécanisme de fusion itératif. La décision de fusion s'appuie sur des critères bas-niveau tels que la similarité radiométrique des régions ou la régularité géométrique de leurs frontières. Cependant, face à des images complexes telles que les images médicales ou les images urbaines, ces informations bas-niveau sont insuffisantes pour guider correctement le processus de fusion vers la construction de régions d'intérêt.

Dans le cadre de cette thèse, nous nous sommes intéressés à la construction d'une hiérarchie de régions par injection d'un *a priori* afin de mieux guider le processus de fusion. Cet *a priori* repose sur les connaissances haut-niveau issues de l'application. Nous proposons une première approche de segmentation hiérarchique dite supervisée. Dans cette approche, nous modélisons le contenu de l'image sous la forme de  $K$  classes. Un classifieur est construit à partir d'une base d'apprentissage. L'objectif est de sélectionner les attributs qui discriminent le mieux les classes et d'estimer par la même les paramètres du classifieur. Notre approche propose de combiner deux critères d'agrégation bas-niveau avec un critère d'agrégation dit *contextuel*. Son rôle est de pénaliser la fusion entre régions ayant une forte probabilité d'appartenir à des classes différentes.

Nous nous sommes également intéressé à l'injection d'un *a priori* spatial dans le processus de construction d'une hiérarchie. Cette information spatiale se présente sous la forme d'un ensemble de blobs. L'approche proposée repose sur la minimisation d'une énergie affine combinant deux termes. Le premier terme est une énergie radiométrique bas-niveau qui mesure l'homogénéité de la région. Le deuxième terme est une énergie spatiale sous-additive construite à partir de l'*a priori* spatial introduit.

**Mots Clés :** Segmentation multi-échelle, représentation hiérarchique, critères d'agrégation, énergie spatiale, classification supervisée, blobs.



# ABSTRACT

The hierarchical representation of an image provides a rich description of the content of the scene. The construction of a hierarchy of regions is generally an ascending process guided by an iterative mechanism of fusion. The decision of fusion is based on low-level criteria such as the radiometric similarity of the regions or the geometrical regularity their boundaries. However, with complex images such as the medical images or the urban images, this low-level information is insufficient to correctly guide the process of fusion towards the construction of regions of interest.

In the frame of this thesis, we were interested in the construction of a hierarchy of regions by injection of prior knowledge in order to better guide the process of fusion. This prior relies on high level knowledges resulting from the application. We propose a first approach of hierarchical segmentation called supervised. In this approach, we model the content of the image in the form of  $K$  classes. A classifier is built by using a training database. The purpose is to select the attributes which discriminate best the classes and also to estimate the parameters of the classifier. Our approach proposes to combine two low-level aggregation criteria with an aggregation criterion called *contextual*. Its role is to penalize fusions between regions having a strong probability of belonging to different classes.

We were also interested in the injection of spatial prior in the process of hierarchy construction. This spatial information is represented through a set of blobs. We propose to build a hierarchy of regions by using this spatial information, the purpose is to preserve the structures, contained in the blobs, from early fusions. The approach suggested relies on the minimization of an affine energy which combines two terms : the first is a low-level radiometric energy which measures the homogeneity of the region ; the second is a sub-additive spatial energy built from the spatial prior introduced before.

**Key words :** multi-scale segmentation, hierarchical representation, aggregation criteria, spatial energy, supervised classification, blobs.



# PUBLICATIONS

## JOURNAUX INTERNATIONAUX

Yousfi K., Cocquerez J-P., Ambroise C. « Scale-sets via spatial prior ». (submitted to International Journal on Computer Vision 2008)

## CONFÉRENCES INTERNATIONALES AVEC COMITÉ DE LECTURE ET PUBLICATION DES ACTES

Yousfi K., Cocquerez J-P., Ambroise C. « Hierarchy construction via spatial prior : semi-automatic approach », (submitted to International Conference on Image Processing - ICIP 2008 - )

Yousfi K., Ambroise C. ,Cocquerez J-P., Chevelu J. « Supervised learning for guiding hierarchy construction : Application to Osteo-Articular medical images database », in International Conference on Pattern Recognition, Hong Kong, China, August 20-24, 2006.

Yousfi K., Ambroise C. ,Cocquerez J-P., Chevelu J. « Driving hierarchy construction via supervised learning : Application to Osteo-Articular medical images database », in International Conference on Image Processing (ICIP), Atlanta, October 8-11, 2006.

## COMMUNICATIONS ORALES SANS PUBLICATION D'ACTES

Yousfi K., Injection d'un a priori dans la construction d'une hiérarchie de régions : Applications aux images Ostéo-Articulaire. Séminaire annuel du laboratoire Heudiasyc, Septembre 2005.

Yousfi K., Indexation multi-échelles des images médicales : Application à la base d'images tomographiques des reins. Séminaire annuel du laboratoire Heudiasyc, Septembre 2004.

# TABLE DES MATIÈRES

TABLE DES MATIÈRES	xii
LISTE DES FIGURES	xv
LISTE DES TABLEAUX	xviii
PRÉFACE	1
<b>1 ANALYSE MONO ET MULTI-ÉCHELLE DES IMAGES</b>	<b>5</b>
<b>1.1 ANALYSE VISUELLE ET INTERPRÉTATION . . . . .</b>	<b>6</b>
1.1.1 Analyse Haut-Niveau . . . . .	7
1.1.2 Analyse bas-Niveau . . . . .	7
1.1.3 Segmentation . . . . .	8
1.1.4 Echelle d'analyse . . . . .	9
<b>1.2 ANALYSE MONO-ÉCHELLE . . . . .</b>	<b>9</b>
1.2.1 Approches frontière . . . . .	9
1.2.2 Approches région . . . . .	12
1.2.3 Approches Hybrides . . . . .	13
<b>1.3 ANALYSE MULTI-ÉCHELLE . . . . .</b>	<b>14</b>
1.3.1 Théorie Espace-Echelle . . . . .	14
1.3.1.1 Echelle(s) d'analyse . . . . .	16
1.3.1.2 Espace-échelle et vision par ordinateur . . . . .	17
1.3.2 Approches hiérarchiques . . . . .	18
1.3.2.1 Méthodes de segmentation par fusion/division de régions . . . . .	18
1.3.2.2 Approches pyramidales . . . . .	21
1.3.2.3 Approches énergétiques . . . . .	22
1.3.2.4 Approches morphologiques . . . . .	23
1.4 APPROCHE ENSEMBLE-ÉCHELLE . . . . .	23
1.4.1 Définitions et Axiomes . . . . .	23
1.4.2 L'algorithme d'escalade . . . . .	28
1.4.3 Résultats expérimentaux . . . . .	29
1.4.3.1 Choix d'une énergie affine . . . . .	29
1.4.3.2 Application à l'imagerie satellitaire et aérienne . . . . .	32
1.4.3.3 Application à l'imagerie médicale . . . . .	36
1.4.3.4 Application en vision par ordinateur . . . . .	39
1.4.3.5 Discussion . . . . .	40
CONCLUSION . . . . .	43
<b>2 CONSTRUCTION SUPERVISÉE D'UNE HIÉRARCHIE DE RÉGIONS</b>	<b>45</b>
<b>2.1 SÉMANTIQUE ET DÉCISION LOCALE . . . . .</b>	<b>46</b>

2.2	D'UN FORMALISME ÉNERGÉTIQUE VERS UN FORMALISME AGRÉGATIF . . . . .	47
2.3	COÛT D'AGRÉGATION CONTEXTUEL . . . . .	49
2.3.1	Modélisation du contenu de l'image . . . . .	49
2.3.2	Injection d'un <i>a priori</i> . . . . .	51
2.3.3	Distances entre distributions de probabilités . . . . .	52
2.3.4	Discussion . . . . .	54
2.4	RÉSULTATS EXPÉRIMENTAUX . . . . .	55
2.4.1	Cas de 2 classes . . . . .	55
2.4.1.1	Pré-traitements des images . . . . .	55
2.4.1.2	Choix du modèle paramétrique . . . . .	56
2.4.1.3	Phase d'apprentissage . . . . .	57
2.4.1.4	Sélection des fonctions de coûts . . . . .	57
2.4.1.5	Rehaussement du contenu sémantique de la hiérarchie . . . . .	58
2.4.1.6	Persistante des régions . . . . .	62
2.4.1.7	Temps d'exécution . . . . .	66
2.4.2	Cas de 3 classes . . . . .	69
2.4.2.1	Exploration visuelle de la hiérarchie . . . . .	69
2.4.2.2	Visualisation par carte de contours . . . . .	70
2.4.2.3	Persistante des régions . . . . .	73
2.4.2.4	Temps d'exécution . . . . .	77
2.5	DISCUSSION . . . . .	78
	CONCLUSION . . . . .	80
3	ENSEMBLE-ÉCHELLE PAR INJECTION D'UN <i>a priori</i> SPATIAL . . . . .	83
3.1	PRÉSENTATION DE LA DÉMARCHE . . . . .	83
3.1.1	Principe d'un potentiel d'attraction . . . . .	85
3.1.2	Pour une énergie spatiale sous-additive . . . . .	86
3.2	MODÉLISATION DE L'ÉNERGIE SPATIALE . . . . .	87
3.3	CONSTRUCTION D'UNE ÉNERGIE SPATIALE SOUS-ADDITIVE . . . . .	87
3.3.1	Potentiel spatial relatif à un blob . . . . .	88
3.3.2	Energie spatiale relative à un blob . . . . .	91
3.3.2.1	Energie bi-potentielle relative à un blob : . . . . .	91
3.3.2.2	Energie multi/mono-potentielle relative à un blob : . . . . .	93
3.3.2.3	Convexité d'une énergie spatiale relative à un blob : . . . . .	94
3.3.3	Energie d' <i>a priori</i> spatial . . . . .	94
3.4	CONSTRUCTION DE POTENTIELS SOUS, ULTRA ET/OU SUPRA-ADDITIFS . . . . .	96
3.4.1	Potentiel supra-additif . . . . .	96
3.4.2	La convexité . . . . .	99
3.4.3	La sous-additivité . . . . .	104
3.4.4	Conclusion . . . . .	105
3.5	RÉSULTATS EXPÉRIMENTAUX . . . . .	106
3.5.1	Cas d'un blob . . . . .	106
3.5.1.1	Energie spatiale mono-potentielle . . . . .	106
3.5.1.2	Combinaison linéaire d'énergies sous-additives . . . . .	131
3.5.1.3	Energie spatiale multi-potentielle . . . . .	132
3.5.1.4	Energie spatiale bi-potentielle . . . . .	142
3.5.2	Cas de plusieurs blobs . . . . .	151
3.5.2.1	Energie mono-potentielle . . . . .	151
3.5.2.2	Energie bi-potentielle . . . . .	160

CONCLUSION . . . . .	163
<b>4 CONCLUSION GÉNÉRALE ET PERSPECTIVES</b>	<b>165</b>
<b>4.1 CONCLUSIONS</b>	<b>165</b>
4.1.1 Construction hiérarchique supervisée . . . . .	166
4.1.2 Construction d'un ensemble-échelle via un <i>a priori</i> spatial . . . . .	166
<b>4.2 PERSPECTIVES</b>	<b>167</b>
4.2.1 Injection d' <i>a priori</i> radiométrique et/ou géométrique . . . . .	167
4.2.2 Injection d' <i>a priori</i> spatial . . . . .	167
4.2.3 Perspectives générales . . . . .	168
<b>A DESCRIPTEURS DE RÉGIONS</b>	<b>173</b>
<b>B SUR-SEGMENTATION PAR LIGNE DE PARTAGE DES EAUX</b>	<b>177</b>
<b>BIBLIOGRAPHIE</b>	<b>179</b>

# LISTE DES FIGURES

1.1	Espace-échelle linéaire (noyau gaussien) . . . . .	15
1.2	Espace-échelle non-linéaire (diffusion anisotropique) . . . . .	16
1.3	Illustrations : hiérarchie/dendrogramme, coupe, hiérarchie partielle .	24
1.4	Etape d'optimisation par programmation dynamique fonctionnelle .	28
1.5	Ensembles-échelles de l'image d'une façade d'un bâtiment . . . . .	31
1.6	Ensembles-échelles de l'image de lena . . . . .	33
1.7	Ensembles-échelles d'une image aérienne . . . . .	34
1.8	Ensemble-échelle d'une image aérienne : Escalade binaire . . . . .	35
1.9	Deux coupes du métacarpe de deux patients . . . . .	37
1.10	Fusions aberrantes d'une escalade binaire sur une coupe du métacarpe	37
1.11	Fusions aberrantes d'une escalade binaire sur une coupe du métacarpe	38
1.12	Ensemble-échelle d'une image de piétons . . . . .	41
1.13	Ensemble-échelle d'une image d'une voiture . . . . .	42
2.1	Séquence de pré-traitements d'une coupe du métacarpe : seuillage, érosion, dilatation . . . . .	55
2.2	Extraction du rectangle englobant . . . . .	55
2.3	Image du métacarpe redimensionnée (256x256) après pré-traitements	56
2.4	Qualité des fusions entre différentes méthodes avec ou sans <i>a priori</i> : cas 2 classes . . . . .	60
2.5	Qualité des fusions obtenues en intégrant un <i>a priori</i> : cas 2 classes .	61
2.6	Illustration du rôle du coût contextuel dans le processus de fusion .	63
2.7	Illustration du rôle du coût contextuel dans le processus de fusion .	64
2.8	Exemples de métacarpes extraits de la base d'image . . . . .	65
2.9	Box-Plot comparant persistance, échelle d'apparition et de disparition des régions osseuses . . . . .	66
2.10	Comparaison du temps d'exécution entre l'algorithme d'escalade binaire et l'algorithme d'agrégation contextuelle . . . . .	69
2.11	Illustration du rôle du coût contextuel dans le processus de fusion .	71
2.12	Illustration du rôle du coût contextuel dans le processus de fusion .	72
2.13	Illustration du rôle du coût contextuel dans le processus de fusion : vision contour . . . . .	74
2.14	Illustration du rôle du coût contextuel dans le processus de fusion : vision contour . . . . .	75
2.15	Box-Plot comparant persistance, échelle d'apparition et de disparition des régions osseuses . . . . .	76
2.16	Comparaison du temps d'exécution entre l'algorithme d'escalade binaire et l'algorithme d'agrégation contextuelle . . . . .	78
3.1	Exemple d'un <i>a priori</i> spatial sous forme de blobs elliptiques . . . . .	84

3.2	(a) une image du périphérique ainsi que l'image des blobs illustrant deux zones d'intérêt (b). . . . .	85
3.3	Exemple d'un blob elliptique contenant un objet d'intérêt . . . . .	85
3.4	Objet complexe avec des structures relativement homogènes. . . . .	86
3.5	Trois types d'énergies spatiale. . . . .	87
3.6	Décomposition du voisinage de deux régions adjacentes . . . . .	92
3.7	Image de synthèse ainsi que l'image blob . . . . .	107
3.8	Ensemble-échelle d'une image de synthèse . . . . .	109
3.9	Suites de coupes monotones pour différentes énergies spatiales . . .	110
3.10	Suites de coupes monotones pour différentes énergies spatiales . .	110
3.11	Ensembles-échelles d'une image de trafic routier . . . . .	113
3.12	Coupes de 20 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	114
3.13	Ensembles-échelles d'une image de trafic routier . . . . .	116
3.14	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	117
3.15	Ensembles-échelles d'une image de trafic routier . . . . .	118
3.16	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	119
3.17	Ensembles-échelles d'une image de trafic routier . . . . .	120
3.18	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	121
3.19	Ensembles-échelles d'une image de trafic routier . . . . .	123
3.20	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	124
3.21	Ensembles-échelles d'une image de trafic routier . . . . .	125
3.22	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	126
3.23	Ensembles-échelles d'une image de trafic routier . . . . .	127
3.24	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	128
3.25	Ensembles-échelles d'une image de trafic routier . . . . .	129
3.26	Coupes de 60 régions extraites de deux hiérarchies (Mumford-Shah/Énergie spatiale) : comparaison gris/couleur . . . . .	130
3.27	Ensembles-échelles d'une image de trafic routier : combinaison d'un <i>a priori</i> spatial avec la longueur du périmètre . . . . .	133
3.28	Ensembles-échelles d'une image de trafic routier : combinaison d'un <i>a priori</i> spatial avec la longueur du périmètre . . . . .	134
3.29	Ensembles-échelles d'une image de trafic routier : combinaison d'un <i>a priori</i> spatial avec la longueur du périmètre . . . . .	135
3.30	Ensembles-échelles d'une image de trafic routier . . . . .	136
3.31	Ensembles-échelles d'une image de trafic routier . . . . .	137
3.32	Ensembles-échelles d'une image de trafic routier . . . . .	138
3.33	Ensembles-échelles d'une image de trafic routier . . . . .	139
3.34	Ensembles-échelles d'une image de trafic routier . . . . .	140
3.35	Ensembles-échelles d'une image de trafic routier . . . . .	141
3.36	Ensembles-échelles d'une image de trafic routier . . . . .	143
3.37	Ensembles-échelles d'une image de trafic routier . . . . .	144
3.38	Ensembles-échelles d'une image de trafic routier . . . . .	145
3.39	Ensembles-échelles d'une image de trafic routier . . . . .	146

3.40	Ensembles-échelles d'une image de trafic routier . . . . .	147
3.41	Ensembles-échelles d'une image de trafic routier . . . . .	148
3.42	Ensembles-échelles d'une image de trafic routier . . . . .	149
3.43	Ensembles-échelles d'une image de trafic routier . . . . .	150
3.44	Ensemble-échelle d'une image de synthèse . . . . .	153
3.45	Suites de coupes monotones pour différentes énergies spatiales . . .	154
3.46	Suites de coupes monotones pour différentes énergies spatiales . . .	154
3.47	Ensembles-échelles d'une image de trafic routier . . . . .	156
3.48	Ensembles-échelles d'une image de trafic routier . . . . .	157
3.49	Ensembles-échelles d'une image de trafic routier . . . . .	158
3.50	Ensembles-échelles d'une image de trafic routier . . . . .	159
3.51	Ensembles-échelles d'une image de trafic routier . . . . .	161
3.52	Ensembles-échelles d'une image de trafic routier . . . . .	162

# Liste des tableaux

2.1	Statistiques sur la persistance et les échelles d'apparition et de disparition : Approche ensemble-échelle <i>vs</i> Approche par agrégation contextuelle (cas 2 classes) . . . . .	65
2.2	Tableau comparatif des temps d'exécution entre l'algorithme de l'escalade binaire et l'algorithme par agrégation contextuelle (cas 2 classes) . . . . .	69
2.3	Statistiques sur la persistance et les échelles d'apparition et de disparition : Approche ensemble-échelle <i>vs</i> Approche par agrégation contextuelle (cas 3 classes) . . . . .	77
2.4	Tableau comparatif des temps d'exécution entre l'algorithme de l'escalade binaire et l'algorithme par agrégation contextuelle (cas 3 classes) . . . . .	78
2.5	Exemple numérique illustrant le comportement de la distance de Kullback-Leibler symétrisée : 1 . . . . .	79
2.6	Exemple numérique illustrant le comportement de la distance de Kullback-Leibler symétrisée : 2 . . . . .	80
2.7	Exemple numérique illustrant le comportement de la distance de Kullback-Leibler symétrisée : 3 . . . . .	80
3.1	Illustration du comportement de l'énergie spatiale mono-potentielle sur une image de synthèse . . . . .	108

# INTRODUCTION . . .

**L**e présent manuscrit décrit les résultats de recherches que nous avons menées au laboratoire **HeuDiaSyC** de l'Université de Technologie de Compiègne (U.T.C).

Ces recherches ont été co-dirigées par Jean Pierre Cocquerez, professeur et chercheur au laboratoire **HeuDiaSyC** de l'U.T.C et par Christophe Ambroise, professeur et chercheur au laboratoire **Statistique et Génome** au Génopole d'Evry.

Le sujet abordé dans cette thèse s'inscrit dans un thème plus général de l'analyse d'image, à savoir la **segmentation d'images**.

La segmentation d'images est un problème complexe et largement abordé dans le domaine du traitement d'images. Cette opération fondamentale a pour objectif de synthétiser l'information contenue dans l'image par le biais d'une description concise et structurée. La description d'une image peut prendre la forme d'une partition en régions, d'une carte de contours, de points d'intérêt, etc. Cette description servira de base à toute interprétation ultérieure de la scène. La segmentation n'est donc pas une fin en soi, mais une étape incontournable dans toute chaîne de traitement. L'opération de segmentation conditionne la bonne marche et l'efficacité de tout système de vision. Ainsi, les performances d'un système d'analyse de scènes d'intérieur sont étroitement liées à la qualité de la description d'une image issue de la segmentation. On peut citer d'autres exemples dans lesquels la segmentation joue un rôle important tels que la vision artificielle en robotique, la détection de piétons, etc.

La bibliographie est riche de méthodes, approches et algorithmes divers et variés dont l'objectif est de segmenter une image. Toutes ces approches peuvent dépendre d'un ou de plusieurs paramètres qui permettent de régler le niveau de la segmentation. Il s'agit en réalité de l'échelle d'analyse de l'image qui permet de se focaliser sur les zones pertinentes pour l'application.

Une récente thèse Guigues [2003] a montré que la détermination de l'échelle de description d'une image est un problème clé qui concerne n'importe qu'elle méthode de segmentation. Un algorithme de segmentation classique, peut permettre, suivant un certain paramétrage, d'extraire une partie des régions d'intérêt tandis que d'autres régions resteront sur-segmentées. Dès lors, une plus forte régularisation, par le biais d'un nouveau paramétrage de l'algorithme, permet de correctement segmenter ces zones sur-segmentées. Cependant, des zones d'intérêt moins contrastées seraient alors absentes (sous-segmentées) de la nouvelle segmentation à cause de cette nouvelle régularisation. Ces problèmes de réglages sont reliés à l'échelle d'analyse de l'image. En occultant l'échelle d'analyse, la description d'une image ne peut répondre correctement aux besoins d'une application.

Cette idée de réglage reliée à une notion d'échelle de description a conduit [Guigues, Cocquerez, et Le Men 2006] à la conclusion qu'**une analyse - bas-niveau - en régions, complète** de l'image doit être présentée sous la forme d'une famille multi-échelle de segmentation qui décrit la scène du détail le plus fin au plus grossier. Les auteurs ont introduit une nouvelle méthode de segmentation multi-

échelle, orientée région, appelée **Représentation Ensemble-échelle d'une image** (scale-set), qui est en quelque sorte le pendant ensembliste de la théorie scale-space introduite et développée par [Witkin 1983, Koenderink 1984].

L'approche « ensemble-échelle » décrit l'image sous la forme d'une hiérarchie de régions (suite de partitions monotone). Cette description est plus riche mais également plus volumineuse qu'une simple partition ou carte de contours. L'ensemble-échelle est construit par minimisation d'une énergie affine. Cette énergie met en concurrence deux énergies antagonistes. La première exprime la fidélité aux observations (un terme d'attache aux données), tandis que la deuxième mesure la complexité de la segmentation (un terme de régularisation). L'approche ensemble-échelle offre une analyse en échelle **continue** à l'inverse d'autres approches où l'échelle est discrétisée. Les coupes horizontales de la hiérarchie, indiquée par l'échelle, représentent les partitions optimales (aux échelles correspondantes) au sens de l'énergie affine. Chaque région est caractérisée par ses échelles d'apparition et de disparition ainsi que par sa persistance. L'utilisateur peut naviguer à travers la hiérarchie afin de sélectionner la ou les coupes qui lui semblent pertinentes pour son application. Le principe algorithmique de cette approche, consiste à simuler une croissance continue de l'échelle à partir d'une sur-segmentation de l'image. On aboutit ainsi à un **processus de fusion optimisé**, appelé « **Escalade binaire** », dans lequel la paire de régions fusionnées est celle qui possède l'échelle d'apparition la plus faible.

Nous avons appliqué l'approche « scale-set » à différents types d'images. Cette étude expérimentale a montré son efficacité dans de nombreuses applications touchant à l'imagerie satellitaire ou aérienne. Le modèle énergétique utilisé est celui de la fonctionnelle constante par morceaux de Mumford et Shah [1989] ou des variantes de cette dernière. Cependant, dans d'autres domaines tels que l'imagerie médicale ou la vision, cette approche a montré certaines limites. Différentes raisons expliquent ces déconvenues et seront à la base des choix scientifiques entrepris durant cette thèse. La principale est la variabilité et la complexité des régions d'intérêt à segmenter dans certains types d'images. Les énergies telles que celles issues de la fonctionnelle constante par morceaux de Mumford-Shah, sont elles indiquées dans les images contenant des objets complexes ? Notons qu'au fur et à mesure que le processus de fusion avance, les régions à reconstituer deviennent plus complexes, rendant la décision de fusion, qui s'appuie uniquement sur des caractéristiques primaires, peu fiable. Le travail expérimental a montré que de nombreuses fusions aberrantes entre régions appartenant à des classes sémantiques différentes se produisaient, empêchant la construction des régions d'intérêt. Une énergie affine, qui n'est en somme que la combinaison linéaire de deux critères, peut elle refléter à elle seule la variabilité (et la sémantique) de la classe d'une région ?

Dans d'autres classes d'image, l'expérience a montré que l'approche ensemble-échelle ne confère pas à certaines régions d'intérêt l'importance qu'elles devraient avoir au sein de la hiérarchie de régions. Ainsi, certaines régions qui, spatialement et visuellement, occupent une place prépondérante dans la scène sont absorbées rapidement et n'apparaissent pas (ni leurs structures) dans les hautes échelles de la description. La raison est la même que celle évoquée ci-dessus, à savoir la complexité de la région à segmenter. Cependant, une autre difficulté s'ajoute à la précédente. Certains objets ne sont pas hiérarchisés par nature. Ils se composent de différentes structures éventuellement homogènes mais de nature différente entre elles (couleur, texture, forme, etc.). Le processus de fusion ne pourra pas regrouper ces structures en un tout représentant une entité symbolique. Néanmoins, ces

structures possèdent une qualité commune, celle de la cohésion spatiale qui permet à l'œil humain de les regrouper sous une même bannière. Cette cohésion spatiale est une connaissance de haut-niveau dont le processus de fusion en ignore l'existence.

Ainsi, si l'on considère les deux énergies, radiométrique et géométrique, comme deux attributs caractérisant chaque région candidate à la fusion, elles ne peuvent refléter la nature complexe et variable de la classe d'une région<sup>1</sup>. Dans ce cas, n'est-il pas préférable d'adoindre à ce processus des **connaissances *a priori*** afin de mieux guider la construction hiérarchique ?

La segmentation guidée par les connaissances est une approche descendante puisqu'elle s'appuie sur la spécificité de l'application pour résoudre le problème de segmentation. Elle permet de coupler l'opération de délimitation avec celle de descrimination des zones d'intérêt. L'injection des connaissances *a priori* dans le domaine du traitement d'images peut se présenter de différentes manières selon les spécificités de l'application. Quand la classe d'images étudiée est suffisamment contrainte, comme c'est souvent le cas dans le domaine de l'imagerie médicale, l'information *a priori* prend la forme d'un ensemble de classes désignant la sémantique des zones d'intérêt. Quand l'application s'intéresse à certaines zones de l'image, l'*a priori* peut prendre la forme d'une localisation spatiale grossière de ces zones, on parle alors de blobs. Cette information spatiale peut être fournie soit manuellement soit par le biais de détecteurs couplés avec le système d'acquisition. La connaissance *a priori* peut aussi s'intéresser à la géométrie des objets étudiés, elle se présentera alors sous la forme d'un modèle de forme (Shape Model). Ce modèle est le fruit d'une étude spécifique de la forme de l'objet. Il tient compte de l'échelle et des différentes transformations linéaires de la classe de l'objet.

Nous proposons dans cette thèse d'étudier la manière d'injecter dans l'approche ensemble-échelle des connaissances *a priori* dans le processus de fusion afin d'améliorer la qualité des hiérarchies construites. Le principe est de coupler à la fois les informations bas-niveau et haut-niveau dans le but de mieux guider le processus de fusion. Notre motivation est d'une part, de rendre cette hiérarchie plus stable face à la variabilité des images. D'autre part, de tenir compte de l'application visée afin de favoriser la construction des régions pertinentes.

On abordera l'injection de connaissances haut-niveau sous deux formes : celle d'un *a priori radiométrique et/ou géométrique* et celle d'un *a priori spatial*. Ce présent manuscrit se divise en 4 chapitres. Le **premier chapitre** est un état de l'art non-exhaustif des techniques de segmentation. L'objectif principal est de visiter les méthodes de segmentation hiérarchique existante et de rappeler le formalisme de la théorie ensemble-échelle puisqu'elle sera le point de départ de nos recherches. Le **deuxième chapitre** présente une nouvelle approche d'injection d'*a priori* radiométrique et/ou géométrique pour la construction d'une hiérarchie de régions. Une étude comparative à la fois qualitative et quantitative entre l'approche mère « ensemble-échelle » et cette nouvelle approche est dressée. Le **troisième chapitre** s'intéresse à l'injection d'un *a priori spatial* dans le formalisme énergétique de l'approche ensemble-échelle. On étudiera les différentes manières de construire une énergie spatiale **sous-additive**, condition nécessaire pour la construction d'un ensemble-échelle. Les différentes expérimentations menées ainsi que les résultats

1. Par ailleurs, Jean-Michel Morel rapporteur du travail de recherche Laurent [2003] a souligné l'importance que l'on doit accorder à la construction d'une énergie affine. Les conditions imposées à la forme de l'énergie affine réduisent considérablement le choix. Il a par conséquent insister sur le fait qu'il est important de mener un travail de recherche sur ce problème.

obtenus seront également exposés dans ce chapitre. Le **quatrième chapitre** se distingue des précédents chapitre puisqu'il s'intéresse à l'extraction d'une partition optimale à partir d'une hiérarchie de régions. Il rejoint toutefois les chapitres 3 et 4 dans l'esprit puisque l'objectif est d'utiliser encore un *a priori* spatial pour l'extraction d'une coupe dans une hiérarchie par l'intermédiaire d'une énergie spatiale. On terminera ce rapport en dressant nos conclusions générales sur les différentes solutions et approches proposées dans cette thèse. On tentera également de poser les différentes voies de recherches qui restent à explorer.

# ANALYSE MONO ET MULTI-ÉCHELLE DES IMAGES

1

## SOMMAIRE

3.1	PRÉSENTATION DE LA DÉMARCHE . . . . .	83
3.1.1	Principe d'un potentiel d'attraction . . . . .	85
3.1.2	Pour une énergie spatiale sous-additive . . . . .	86
3.2	MODÉLISATION DE L'ÉNERGIE SPATIALE . . . . .	87
3.3	CONSTRUCTION D'UNE ÉNERGIE SPATIALE SOUS-ADDITIVE . . . . .	87
3.3.1	Potentiel spatial relatif à un blob . . . . .	88
3.3.2	Energie spatiale relative à un blob . . . . .	91
3.3.3	Energie <i>d'a priori</i> spatial . . . . .	94
3.4	CONSTRUCTION DE POTENTIELS SOUS, ULTRA ET/OU SUPRA-ADDITIFS . . . . .	96
3.4.1	Potentiel supra-additif . . . . .	96
3.4.2	La convexité . . . . .	99
3.4.3	La sous-additivité . . . . .	104
3.4.4	Conclusion . . . . .	105
3.5	RÉSULTATS EXPÉRIMENTAUX . . . . .	106
3.5.1	Cas d'un blob . . . . .	106
3.5.2	Cas de plusieurs blobs . . . . .	151
	CONCLUSION . . . . .	163

L'objectif de ce chapitre est de donner une vue d'ensemble des différentes approches et techniques de segmentation dont la littérature foisonne. Il ne s'agira pas de faire un état de l'art exhaustif de toutes les méthodes de segmentation ou d'estimer la qualité de ces dernières mais plutôt d'appréhender les fondements des grandes familles de segmentation. La richesse et l'abondance des techniques de segmentation sont dues principalement à l'absence d'une définition claire et précise de la segmentation. Les différents types d'images auxquels nous sommes confrontés (imagerie satellitaire, imagerie radar, imagerie médicale, etc.) ainsi que les diverses applications potentielles sont à la source de cette explosion de travaux. Nous tenterons ici, d'exposer d'une manière concise les différentes approches. On

reliera également ces méthodes aux différentes applications auxquelles elles ont été confrontées, afin de mieux comprendre leurs avantages et inconvénients. Nous aborderons aussi bien les méthodes dites mono échelle que multi-échelle. On explorera tout particulièrement les approches de segmentation dites hiérarchique, pyramidale ou multi-échelle, qu'elles soient orientées contour ou région.

## 1.1 ANALYSE VISUELLE ET INTERPRÉTATION

Analyser ou Interpréter une image est un acte qui paraît anodin et tellement simple pour l'être humain que l'on pourrait croire qu'il est facilement transposable à la machine. Prenant l'exemple d'une image représentant un paysage quelconque. On est capable en quelques millièmes de secondes d'y associer des mots **décrivant** la scène. Décrire une image, c'est préciser à la fois les objets qui la composent (i.e leur attribuer des valeurs symboliques), mais également les « relations » entre ces différents objets : position spatiale, relation de voisinage, contexte. La définition d'une zone d'intérêt est complexe et subjective. Plusieurs situations peuvent amener l'observateur à s'intéresser à un objet en particulier. Un objet n'a d'intérêt pour l'œil humain que s'il occupe une place prépondérante dans la scène (notion d'échelle), ou s'il contribue de façon significative à la sémantique globale de la scène. Il est qualifié d'objet d'intérêt, si ses caractéristiques de couleur et de forme le singularise par rapport à l'environnement qui l'entoure (contraste, géométrie). L'interprétation d'une image s'opère une fois que tout ou partie des objets qui la composent sont reconnus. Cette reconnaissance fait forcément appel à nos connaissances *a priori* mémorisées par notre cerveau. Il est fort probable qu'une personne n'ayant jamais observé une image IRM du cerveau, soit incapable d'expliquer la scène. Néanmoins, elle pourra la décrire avec des termes géométriques. Elle pourra également séparer les régions entre elles à travers la texture et la couleur qu'elles portent. Cependant, reconnaître des objets dans une scène, ne suffit pas pour interpréter l'image. Les relations spatiales entre les objets permettent de donner un sens à la scène. Certains objets sont même reconnus grâce à leurs relations avec leurs voisins. Les recherches menées dans le domaine de la neuro-perception, ont permis de montrer que le cerveau humain utilise également un mécanisme de « probabilisation » afin d'interpréter une scène. Ce mécanisme conditionne l'observateur de la scène en fonction du contexte culturel, professionnel, etc. dont il est issu. Les connaissances cognitives et les apprentissages perceptuels auxquels est entraîné l'observateur déterminent en grande partie sa décision. On ne peut reconnaître ou identifier une représentation picturale d'un objet sans avoir fait l'expérience préalable d'en construire une représentation. En d'autres termes, nos perceptions font aussi appel à des scènes observées et mémorisées dans le passé afin d'envisager une interprétation possible de l'image. Sans ce mécanisme, une scène sera sujette à plusieurs interprétations. L'exemple de l'image IRM du cerveau illustre parfaitement nos propos.

Cette introduction sommaire nous a permis d'exhiber les quelques notions fondamentales qui sont à la base de la vision, de l'analyse de l'image et de concepts tels que : l'analyse bas-niveau d'une image, l'analyse haut-niveau d'une image, la localisation, la détection, ....

### 1.1.1 Analyse Haut-Niveau

Il est rare qu'une image renvoie à une seule interprétation, on parle alors de *polysémie* de l'image. Elle est polysémique car son interprétation dépend avant tout de l'observateur, de ces connaissances et du contexte d'observation. Un exemple illustrant cette variabilité dans l'interprétation est le cas d'une figure accompagnée ou non d'une légende. On s'aperçoit rapidement qu'une légende réduit considérablement la polysémie de la figure, elle oriente l'observateur vers une interprétation ou un ensemble d'interprétations restreint. L'analyse haut-niveau d'une image est généralement considérée comme la couche supérieure d'une chaîne de traitement. Elle permet de donner une interprétation ou un sens à une image à partir des informations transmises par la chaîne de traitement. Le rôle d'une analyse haut-niveau est de répondre aux besoins d'une application spécifique ce qui implique que l'interprétation sera orientée en fonction de cette dernière. L'information reçue par cette couche doit être synthétique et pertinente. Elle se présente généralement sous la forme d'une description de l'image en termes d'objets d'intérêt (contours, régions, points caractéristiques, etc.). Le contexte pictural dans lequel se trouve un objet améliore nettement son identification d'ou l'importance des interactions qu'entretiennent les objets entre eux. Pour un observateur, identifier un objet, revient à lui associer un modèle sémantique déjà observé (le passé). Ce modèle a de plus la propriété d'être invariant à différentes transformations (contraste, homothétie, angle d'observation, espace de projection, etc.). Par conséquent, identifier un objet à partir d'informations sur sa forme et/ou sa luminance implique que l'on ait construit au préalable un prototype (on parle aussi de modèle ou de concept) de cet objet et qu'on puisse comparer la représentation picturale avec le modèle stocké en mémoire. *La notion du modèle est donc plus souple, plus économique et surtout plus robuste. Elle possède une forte caractéristique de constance.* L'analyse haut niveau tente de mimer ce comportement en réalisant également une mise en correspondance entre objets et valeurs sémantiques (i.e classes ou modèles), on parlera souvent d'appariements entre modèles et observations. Relier des modèles à des observations, est un des objectifs de disciplines telles que « La Reconnaissance de Formes ». En traitement d'image, l'analyse haut-niveau tente également d'utiliser les connaissances *a priori* pour construire un modèle fiable et robuste, et d'injecter ces connaissances haut-niveau dans un processus pour localiser et identifier les objets sémantiques recherchés.

### 1.1.2 Analyse bas-Niveau

Avant d'aborder plus en détail le rôle d'une analyse bas-niveau en vision, faisons d'abord un détour par la neuro-perception et la psychologie expérimentale. Comme aiment à le rappeler les chercheurs de ce domaine, la base du processus de la vision humaine se fonde avant tout sur une réduction drastique de la quantité d'informations potentiellement perceptible. **C'est un des premiers objectifs d'une description bas-niveau.** Des travaux comme ceux de Francke (1977) ont montré que cette réduction est due à la capacité qu'a notre conscience de traiter l'information : entre 8 et 25 bits/secondes<sup>1</sup>. Ce phénomène de réduction, nécessite une série de mécanismes tels que l'abstraction, la sélection, la catégorisation, la

<sup>1</sup>. soit en moyenne 160 bits si l'on considère que 10 secondes constituent "l'épaisseur du présent". Au niveau neurophysiologique, on observe déjà une importante réduction de l'information puisque l'œil humain contient 130 millions de récepteurs et un million de cellules ganglionnaires qui constituent le nerf optique.

probabilisation, etc. Grâce à ces mécanismes, la perception brute des stimuli peut donner naissance à des primitives telles que la ligne, la surface, la région, etc., et finalement à des entités plus globales, càd des objets.

Il est acquis que l'information que contient une image n'est pas également répartie car l'œil accorde plus d'importance aux contours qu'aux surfaces, aux textures qu'aux zones plates, aux changements de directions et jonctions qu'aux arcs réguliers, aux intersections, etc.<sup>2</sup>. Paillart [1976] appela cette fonction la « capture informationnelle » et souligna que cette fonction était innée chez l'être humain. Par conséquent, toute image présentée à l'œil possède des *points de fixation remarquables qui renferment des informations plus riches*. L'extraction de primitives remarquables est également **une tâche essentielle dans une description bas-niveau**. C'est à partir de ces primitives que débute le processus de reconstruction des formes et zones d'intérêt. La *Théorie Gestalt*<sup>3</sup>, introduite par Max Wertheimer [1944], Wolfgang Köhler [1929] et Kurt Koffka [1935], s'est attachée à définir ce qu'est une forme et à démontrer qu'une forme est une représentation structurée. Cette structure prend forme à partir de primitives remarquables. Cette théorie insiste sur le fait qu'une forme dans un tout est autre chose que cette même forme isolée ou incluse dans un autre tout car elle tire des propriétés particulières de sa place et de sa fonction dans chacun d'entre eux. Nous citerons les propres termes d'un des fondateurs de cette théorie, Wolfgang Kohler : « *Il y a un Tout qui au lieu d'être la somme des parties existant indépendamment, donne leur fonctions ou propriétés spécifiques de parties qui peuvent seulement être définies par rapport au Tout en question* ». L'une des principales notions de la théorie Gestalt est la notion de « regroupement » (Max Wertheimer 1922). La vision humaine, afin de percevoir une forme, tente toujours de regrouper les points si ils sont proches spatialement, si ils sont similaires, si la forme est discontinue, etc. Le mouvement gestaltiste a décrit une série de lois structurales (plus d'une centaine) qui règlent notre perception.

L'analyse bas-niveau d'une image tente donc de donner une description succincte et compacte de la scène en termes de primitives (on parle alors de caricature ou de sketch). Elle vise à caractériser les formes géométriques des zones qui composent la scène, ainsi que leurs propriétés d'apparence en termes de couleurs et/ou textures.

Une description bas-niveau n'est donc qu'une première ébauche du contenu de l'image. Elle ne revêt aucune signification sémantique ou symbolique. Une description bas-niveau peut être destinée à une analyse haut-niveau, pour répondre aux besoins d'une application spécifique. Cependant, cette description peut ne pas intégrer un processus d'interprétation et juste servir à des fins de compression d'informations.

### 1.1.3 Segmentation

Il n'existe pas de définition précise (mathématique) et unique de la segmentation et par conséquent pas de théorie unifiée. La définition de la segmentation fait appel à notre sens commun de la perception des choses. En somme, segmenter revient à diviser l'image en zones d'intérêt afin de réduire sa complexité. Savoir

2. On peut également ajouter à cette liste l'horizontalité, la verticalité, le parallélisme.

3. Le mot allemand Gestalt est traduit par « forme » (ainsi Gestalt theory signifie « théorie de la forme »), mais il s'agit en réalité de quelque chose de beaucoup plus complexe, qu'aucun mot ne traduit exactement dans aucune langue. Aussi, a-t-on conservé ce terme de gestalt aussi bien en français (où il est entré dans le dictionnaire), qu'en anglais, en russe ou en japonais.

si ces zones ont un sens représente la frontière entre ce qu'on appelle l'analyse bas-niveau et haut-niveau.

La segmentation est par conséquent envisagée ou abordée par la communauté de la vision comme une opération purement bas-niveau. D'un point de vue ensembliste, la tâche de segmentation revient donc à découper l'espace, représenté par l'image, en régions deux à deux disjointes et dont l'union permet de reconstituer tout le support de la scène. En se basant sur cette définition, on remarque que l'entité **région**, est l'élément central de la description délivrée par une segmentation. Cette démarche a donné naissance aux approches de segmentation dites « **orientées-région** ».

L'existence physique d'un objet est également liée à ses limites, autrement dit aux frontières qu'il partage et le sépare de son environnement. En résumé, segmenter une image c'est délimiter au sens propre les zones d'intérêt qui la composent. Ainsi, cette définition a poussé les chercheurs à s'intéresser aux frontières qui composent l'image. Elle a donné naissance aux approches, dites « **orientées-frontière** »

Ces deux approches qu'on qualifie souvent de *duales* seront abordées plus en détail dans les sections suivantes. Contentons nous juste de souligner que lorsque l'approche est orientée-région alors elle tentera d'exploiter les similarités radio-métriques entre les éléments picturaux pour construire des zones remarquables tandis que si l'approche est orientée-frontière elle tentera de déceler des changements d'intensités (discontinuités) dans l'image afin de détecter et de localiser les contours.

#### 1.1.4 Echelle d'analyse

L'interprétation d'une image est aussi attachée à une notion d'échelle. Prenons l'exemple d'une image représentant une maison. L'observateur peut s'intéresser au toit de cette maison, mais il peut aussi manifester un intérêt pour les tuiles composant ce toit. Une photo d'une jarre greco-romaine peut n'être qu'une simple jarre pour un observateur lambda, mais pour l'œil d'un expert son intérêt se portera plutôt sur les motifs et les dessins figurant sur la jarre.

Il paraît donc évident qu'il n'existe pas un seul niveau de description de l'image mais plusieurs niveaux allant des détails les plus grossiers aux détails les plus fins. Augmenter l'échelle d'analyse doit permettre d'exhiber des détails plus fins de l'image. Les objets d'intérêt observés à une certaine échelle, renferment des détails (i.e d'autres objets) que l'on devrait distinguer à une échelle plus petite. Cette notion d'échelle, renforce l'idée que les régions d'intérêt sont hiérarchisées. Par conséquent, une représentation complète du contenu de l'image doit prendre la forme d'une hiérarchie de régions, ou d'une suite de partitions monotone de l'image. Elle peut également se présenter sous la forme d'une suite d'images contours où les frontières disparaissent au fur et à mesure que l'on augmente l'échelle d'analyse.

## 1.2 ANALYSE MONO-ÉCHELLE

### 1.2.1 Approches frontière

Les approches orientées frontière, s'attachent à détecter les discontinuités dans l'intensité des niveaux de gris de l'image. On admet alors l'hypothèse que le pas-

sage d'une région à une autre (adjacente) se traduit obligatoirement par une transition (saut), ou une variation rapide dans l'intensité lumineuse. Ceci n'est évidemment vrai que si l'on suppose que les régions adjacentes sont relativement homogènes (faiblement texturées) au sens d'une certaine mesure radiométrique. Ces approches exploitent donc une dissimilarité qui existe entre les objets pour mesurer l'intensité du saut et détecter ainsi la présence d'un contour.

Afin de détecter ces transitions, les premiers travaux ont conduit les chercheurs à s'intéresser au gradient d'une image discrète mais également à son laplacien. Le premier détecteur de contour qui approxime le gradient a été proposé par Roberts [1965]. Le calcul se fait avec deux masques de convolution représentant la dérivée dans deux directions orthogonales. Par la suite plusieurs approximations du gradient discret sont apparues telles que les masques de Perwitt, les masques de Sobel, ceux de Kirsh, sans oublier les masques qui approximent le laplacien. Mathématiquement, ces approches consistent à convoluer l'image avec des filtres de petites dimensions. Elles sont donc dépendantes de la taille des objets traitées et fortement sensibles au bruit. Dès lors, certains auteurs se sont focalisés sur la définition de critères d'optimalité pour la détection de contours. Cette approche équivaut à effectuer une opération de filtrage. Ainsi, Canny [1986] pose trois critères<sup>4</sup> afin d'aboutir à un filtre optimal à réponse impulsionale finie. Par la suite Deriche [1987], Shen et Castan [1986], Chehikian [1997], ont proposé des filtres à réponse impulsionale infinie qui respectent les critères de bonne détection. Une excellente comparaison et analyse des filtres de détection des contours est donnée dans Cocquerez et Philipp [1995]. Malgré le grand développement qu'ont connu les détecteurs de contours, le défaut majeur qui les caractérisent est resté sans réponse. Ces approches produisent des contours incomplets et rarement fermés. Ces inconvénients compliquent souvent l'étape de reconnaissance des formes qui doit suivre. Pour corriger ces inconvénients, l'image des frontières subit une suite de post-traitements comportant par exemple : une élimination des points de contours dus au bruit et une fermeture de contours. Cette dernière a pour objectif de rétablir la connexité des frontières de l'image. Différentes méthodes de fermeture de contours ont été proposées, elles recherchent toutes le chemin optimal au sens d'un certain critère, on pourra consulter les travaux de Martelli [1972], Montanari [1971], Nevatia [1980] ou Giraudon [1987] pour plus de détails sur cette opération. Les détecteurs de contours sont déconseillés dans le cas des images texturées ou fortement bruitées telles que les images acquises en thermographie, les images ultrasonores ou les images radars. En outre, Torre et Poggio [1997], ont montré que le problème de détection de contours est généralement mal posé (au sens de la résolution des systèmes).

Avant de clore notre discussion sur les détecteurs de contours, qu'on peut classer dans la famille des méthodes dérivatives, on notera juste que des auteurs tels que Hueckel [1971], Haralick [1984] se sont intéressés à la détection de contours en considérant l'image comme une surface. L'approche adoptée par Haralick [1984] débouche aussi sur la construction d'opérateurs. D'autres auteurs, issus de la mouvance de la morphologie mathématique, se sont penchés sur la détection de contours en utilisant le gradient morphologique. Ce dernier est défini comme la différence entre l'image dilatée multipliée par un facteur 1/2 et l'image érodée. Si le contour est épais, le gradient morphologique donnera un contour. Si le contour

---

4. Ces critères sont : **la bonne détection** (détecter un maximum de contours), **la bonne localisation** (que les points détectés doivent être le plus proche possible du vrai contour), **la réponse unique** (minimiser le nombre de contours détectés plusieurs fois).

est saillant, celui rendu par l'opérateur sera plus large, la largeur dépendant de la taille de l'élément structurant utilisé. Cependant, dans le cas d'images bruitées ou texturées cet opérateur est contre-indiqué du fait de la sensibilité des opérations morphologiques telles que l'érosion et la dilatation au bruit et à la texture.

En raison du caractère local des détecteurs de contours et du problème de fermeture, des approches globales ont émergé introduisant la notion de modèle de contour déformable. Le premier modèle a été proposé par Kass, Witkin, et Terzopoulos [1988]. Connu sous le nom de contour actif (*snake*), ce modèle est formé d'une série de points mobiles, répartis sur une courbe en deux dimensions. La courbe (qui peut être fermée) est placée dans la zone d'intérêt de l'image ou autour d'un objet. Plusieurs équations décrivent son évolution : la courbe se déplace et épouse lentement les contours des objets en fonction de divers paramètres comme l'élasticité, la tolérance au bruit, etc. Le déplacement du contour initial vers l'objet d'intérêt se fait grâce à la minimisation d'une énergie (somme d'une énergie interne et d'une énergie externe) sous certaines contraintes qui garantissent la régularité de la courbe tout en autorisant des déformations. Les modèles de contours déformables sont particulièrement adaptés pour détecter les formes des objets dans une scène. Le modèle de contour actif a par la suite connu un grand développement aboutissant à divers algorithmes tels que ceux proposés par Amini, Weymouth, et Jain [1990], Cohen et Cohen [1993], Fua et Leclerc [1988], Blake et Isard [1998], etc.

De cette première approche du contour actif, dite physique, d'autres approches telles que les level-sets (Osher et Sethian [1988], Sethian [1996]), les modèles géodésiques actifs (Caselles, Kimmel, et G. [1997], Aubert et Blanc-Féraud [1999]) sont apparues qui sont également des approches variationnelles conduisant à des solutions sous forme de courbes évoluant dans le temps. A l'inverse de l'approche *snake*, la méthode des ensembles de niveaux (Osher et Sethian [1988], Sethian [1996]) représente le contour de façon implicite sous la forme d'un ensemble de niveau d'une fonction scalaire de dimension supérieure. La courbe évolue en fonction de mesures géométriques résultant d'une évolution. L'introduction de cette approche a permis de résoudre le problème de changement de topologie (fusion, séparation, etc.) de façon automatique. Pour une vision globale sur l'évolution de cette approche, nous invitons le lecteur curieux à consulter l'état de l'art dressé dans Osher et Fedkiw [2001]. Les modèles des contours actifs et des level-sets ont ensuite connu une nouvelle évolution grâce au travail de Paragios [2000]. Ce dernier a permis la prise en compte d'informations relatives aux régions délimitées par les contours actifs. Ainsi, on ne se restreint plus à l'information portée par le contour et l'énergie d'attache aux données porte sur la région d'où le nom de **région active** au lieu de contour actif. Une présentation des outils mathématiques utiles pour ces méthodes est dressée dans (Paragios [2005], Aubert et Kornprobst [2006]).

Néanmoins, ces approches dépendent de plusieurs paramètres rendant la phase d'initialisation assez contraignante et par conséquent non-automatique. De plus, ces méthodes sont dédiées à la segmentation d'un seul objet ou d'un ensemble d'objets de faible cardinalité et fixe. A noter que d'autres modèles globaux ont été proposés tels que les modèles perceptifs, les modèles stochastiques, etc. Une classification complète est proposée dans Cocquerez et Philipp [1995].

Les chaînes et les champs de Markov ont aussi apporté leurs pierres angulaires à la détection et l'extraction des contours de l'image, à travers les travaux de Blake et Zisserman [1987] , Geman [1990], Dias et Leitao [1996], etc. L'image est consi-

dérée comme la réalisation d'un champ aléatoire ou le niveau de gris de chaque pixel est supposé être une réalisation bruitée du champ originel. Le label d'un pixel dépend de son voisinage. Dans la détection des contours, ce label indiquera alors la présence ou non d'une frontière. Le principe est de définir des énergies locales (potentiels de clique) entre pixels ayant un certain degré de voisinage et qui reflètent les interactions entre leurs niveaux de gris. La modélisation markovienne du processus aboutit alors à la minimisation d'une énergie globale fonction des énergies locales définies sur chaque site. Ce type d'approches permet l'extraction de contours fermés mais le reproche souvent fait à ce type de modélisation est le coût du calcul excessif. Néanmoins de toutes les approches évoquées ci-dessus, les approches markoviennes demeurent les plus indiquées dans les images texturées.

### 1.2.2 Approches région

A l'inverse des approches frontières, les méthodes orientées région se focalisent sur la recherche de zones partageant des attributs radiométriques communs (luminosité, ...). On distingue les méthodes dites de classification dont l'objectif est d'affecter à chaque pixel une unique classe, i.e un label. On retrouve dans cette famille de segmentation les méthodes de seuillage (Weszka [1978]; Kitler et al. [1985]), la méthode des nuées dynamiques (plus connues sous le nom des K-Moyennes). On retrouve aussi les méthodes de classification floue (Hernandez [1995]; Ogawa et al. [1998]; Ahmed et al. [1999]) ainsi que les méthodes de classification bayésienne (Pappas [1992]; Vemuri et al. [1995]). Dans les méthodes de segmentation par classification, les pixels de l'image sont considérés comme un ensemble d'observations, où chaque individu (pixel) est doté de caractéristiques, mais n'entretenant aucune relation spatiale avec le reste des observations, i.e la topologie de l'image est généralement ignorée. Les régions construites par ce type de modélisation ne respectent pas en général la condition de connexité et les frontières de ces dernières sont de qualité médiocre. Ces défauts nécessitent généralement des pré et post-traitements de l'image afin d'améliorer la qualité de la segmentation.

Les méthodes de type croissance de régions (Adams et Bishof [1994]; Chang et Li [1994]; Revol et Jourlin [1997]) sont aussi des méthodes orientées région. Elles s'appuient sur des prédicats d'homogénéité pour construire les régions d'intérêt. La procédure débute par le choix de germes dans l'image, à partir desquels des pixels ou des régions s'agrègent à condition qu'ils vérifient un prédicat. Les techniques de segmentation par fusion et/ou par division sont aussi des méthodes orientées région qui peuvent soit exhiber une simple partition en guise de segmentation, à travers la définition d'un critère d'arrêt, soit exhiber une suite de partitions monotone ce qui équivaut à une hiérarchie de régions si l'on fait abstraction de ce critère d'arrêt. Pour notre part, nous avons décidé de répertorier ces méthodes dans les approches dites « hiérarchiques » ou « pyramidales ».

La segmentation d'une image peut être vue comme un problème de partitionnement d'un graphe. Chaque pixel de l'image représente un nœud du graphe et les arêtes modélisent les relations d'adjacence entre les pixels. Les arêtes sont caractérisées par une valeur de dissemblance. L'objectif est de rechercher une coupe<sup>5</sup> du graphe qui divise le graphe en deux ensembles complémentaires de nœuds par minimisation d'un critère. Shi et Malik [2000] se sont intéressé à ce problème de division d'une image en un nombre fixé de régions. Il ont proposé un nouveau

---

5. ensemble d'arêtes qui divisent le graphe en deux composantes connexes.

critère global, appelé *critère de coupe normalisée*<sup>6</sup> qui mesure à la fois la dissimilarité totale entre les régions ainsi que la similarité à l'intérieur de ces régions. Cependant, cette approche est gourmande en ressources et en temps d'exécution. En outre, on doit fixer le nombre de régions souhaitées au préalable. Notons également qu'augmenter le nombre de régions n'implique pas forcément un emboîtement des partitions obtenues.

### 1.2.3 Approches Hybrides

On qualifiera d'approche hybride toutes les méthodes de segmentation qui combinent à la fois l'information issue des régions et l'information issue des contours de ces régions. Ainsi certaines approches telles que (Besl et Jain [1988], Pavlidis et Liow [1990]) utilisent à la fois des méthodes de détection de contours et des méthodes de fusion de régions ou de croissance de régions, pour améliorer la qualité de la segmentation. Ces méthodes ne reposaient sur aucune modélisation de la segmentation recherchée et donc la combinaison était purement algorithmique. Avec l'apparition d'approches de segmentation orientée-modèle, la combinaison entre information basée-région et information basée-contour est devenue plus naturelle. Ainsi, dans les approches basées sur les champs de Markov (Geman et Geman [1984] ; Marroquin, Mitter, et Poggio [1987] ; Dubes et Jain [1989] ; Geman [1990]), l'image est considérée comme une réalisation d'un champs de Markov suivant une distribution qui modélise le contexte spatiale de la scène. Le problème de segmentation se transforme en un problème de minimisation d'une énergie combinant à la fois un terme d'attache aux données (information région) et un terme de régularisation qui généralement porte sur la géométrie de la partition (information contour). Cependant, ce type d'approche requiert une bonne connaissance de la distribution *a priori* de l'image label et la plupart de ces techniques sont gourmandes en temps de calcul.

A l'encontre des approches markoviennes, d'autres approches de segmentation basées sur l'inférence d'un modèle et exploitant un formalisme variationnel et déterministe ont été étudiées. Ainsi, Mumford et Shah [1989] proposent de modéliser le problème de segmentation sous la forme d'une minimisation d'une fonctionnelle intégrant également un terme d'attache aux données et un terme de régularisation. Le modèle d'images recherché à travers cette fonctionnelle est celui d'une fonction constante par morceaux. La fonctionnelle proposée par Mumford-Shah peut modéliser des modèles de segmentation plus complexes d'ordre 1 ou plus mais engendre un coût en temps de calcul pour l'optimisation, plus conséquent.

Pour finir, il existe d'autres approches qui combinent à la fois une formalisation aléatoire et déterministe. Elles sont issues des théories dites d'*inférence par codage minimal* telles que la théorie **MDL**<sup>7</sup> ou la théorie **MML**<sup>8</sup>. Ainsi, Leclerc [1989a], est le premier à proposer une méthode de segmentation basée sur la théorie de Description de Longueur Minimale. L'approche consiste aussi à minimiser une énergie dont le premier terme représente la qualité de régression, qui est un terme aléatoire, et le second terme mesure la complexité du modèle, qui est un terme déterministe. Ce dernier terme permet de modéliser la géométrie des régions de la partition. D'autres travaux ont été développés s'appuyant sur la théorie MDL tels que Kanungo et al. [1995], Zhu et Yuille [1996], Lee [2000].

6. Normalized cut.

7. Pour Minimum Description Length. Cette théorie est l'œuvre de Jorma Rissanen

8. Pour Minimum Message Length. Elle a été proposée par Wallace et Freeman en 1987

## 1.3 ANALYSE MULTI-ÉCHELLE

### 1.3.1 Théorie Espace-Echelle

La notion d'analyse en échelle **continue**, d'une image a été introduite pour la première fois par Witkin [1983]. Par la suite, divers auteurs (Koenderink [1984]; Yuille et Poggio [1986]; Lindeberg [1994]; Florack [1997]) se sont penchés sur cette approche, en étudiant son formalisme mathématique. La théorie espace-échelle<sup>9</sup>, s'est considérablement développée ses vingt dernières années. Elle a donné naissance à différentes représentations espace-échelle tels que l'espace-échelle linéaire Koenderink [1984]; les espaces-échelles non linéaires Perona et Malik [1990], Alvarez et al. [1992], Florack et al. [1995], Black et al. [1998]; les espaces-échelles morphologiques (Alvarez et Morel [1994]), ....

L'espace-échelle **gaussien** proposé par Witkin [1983], et dont le formalisme mathématique est dû à Koenderink [1984], est construit à partir d'une convolution de l'image originale avec une gaussienne d'écart type  $\sigma$ . Le paramètre de déviation  $\sigma$  joue le rôle d'échelle d'analyse. Plus formellement, si  $I$  est une image de dimension  $N$ ,  $f : \mathbb{R}^N \rightarrow \mathbb{R}$ , alors sa représentation dans l'espace-échelle gaussien,  $L : \mathcal{R}^N \times \mathbb{R}^+ \rightarrow \mathbb{R}$ , est définie par :

$$L(x; \sigma) = \int_{x \in \mathbb{R}^N} I(y) g_\sigma(x - y) dy = I \star g_\sigma(x)$$

tel que  $L(:, 0) = f$  est l'image originale et  $g_\sigma$  représente le noyau gaussien :

$$g_\sigma(x) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \exp[-(xx^T)/2\sigma^2]$$

Une définition équivalente peut être obtenue par la résolution de l'équation de la chaleur :

$$L(:, t) = c \cdot \Delta L(:, t)$$

avec  $L(:, 0) = I$  comme condition initiale,  $c$  étant une constante.

Intuitivement, l'augmentation du paramètre échelle  $\sigma$  induit un lissage de l'image de plus en plus fort et simplifie ainsi ses structures en supprimant les frontières les moins saillantes. Ainsi, le niveau des détails dans une scène décroît comme si l'on simulait un éloignement progressif (en termes de distance) du point d'observation de la scène. A titre d'exemple, on peut imaginer un observateur regardant les feuilles d'un arbre, puis s'éloigner afin d'observer les arbres et enfin prendre une distance assez grande pour pouvoir observer toute la forêt.

L'unicité du noyau gaussien peut être dérivée de différents ensembles de contraintes définies sur  $L$ . L'ensemble des contraintes introduites par Koenderink [1984], comporte trois conditions : *la causalité*, *l'isotropie*, *l'homogénéité*. Le principe de causalité stipule que les frontières à plus grandes échelles trouvent leurs origines à des échelles plus fines. Autrement dit, augmenter l'échelle d'analyse, entraînera forcément la disparition de détails et ne peut par conséquent créer d'information. Le principe de causalité garantit alors la non création de nouveaux *extrema* locaux et la non amplification des *extrema* existants. Les principes d'isotropie et d'homogénéité signifient que toutes les positions spatiales et les niveaux d'échelles doivent être traités de façon identique.

L'inconvénient de l'espace-échelle linéaire est la délocalisation des frontières lors du lissage progressif de l'image, autrement dit lors d'une augmentation du

---

9. En anglais « scale-space ».

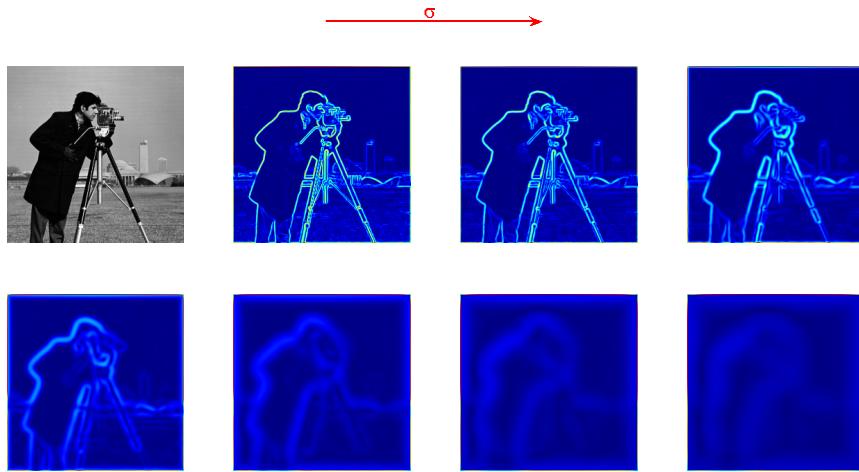


FIGURE 1.1 – La figure représente le module du gradient de l'image après une suite de convolutions avec un noyau gaussien pour  $\sigma = 0.5, 1, 2, 4, 8, 12, 16$ . Elle illustre l'effet de lissage des discontinuités provoqué par cette opération lors d'une croissance de l'écart type  $\sigma$ .

paramètre d'échelle  $\sigma$  (voir figure 1.1). Perona et Malik [1990] ont proposé alors un schéma de diffusion **anisotropique** dans lequel le facteur, dit de conductance, ne serait plus constant mais fonction décroissante du module du gradient :

$$L(., t) = c(., t) \cdot \Delta L + \nabla c \cdot \nabla L,$$

tel que :

$$c(., t) = h(\|\nabla L(., t)\|), \quad \text{où } h \text{ est une fonction décroissante.}$$

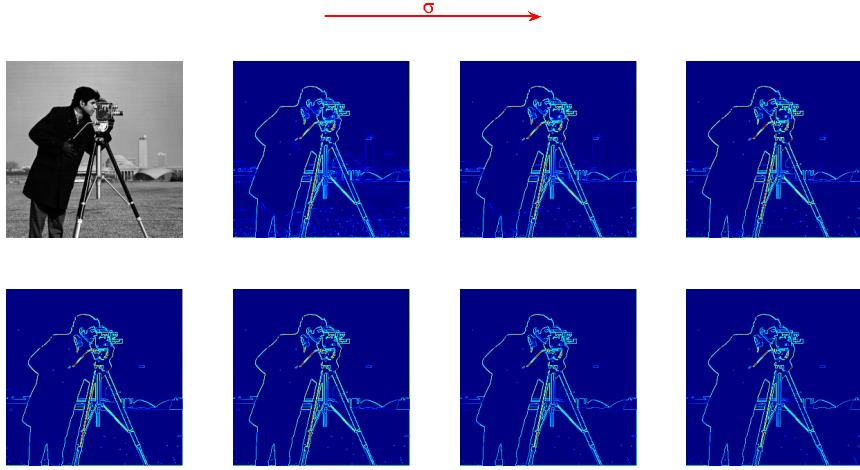
Le choix d'un facteur de conductance *variable*, rend la diffusion **anisotropique**. La diffusion dépend alors du contenu de l'image. Plus précisément, la diffusion est favorisée dans les régions assez homogènes et réduite au voisinage des frontières de ces régions, d'où le choix d'une fonction décroissante selon le module du gradient. Perona et Malik [1990] ont suggéré deux fonctions de diffusion :

$$h_1(\nabla L) = \exp -\left(\frac{\|\nabla L\|}{K}\right)^2,$$

$$h_2(\nabla L) = \frac{1}{1 + \left(\frac{\|\nabla L\|}{K}\right)^2},$$

La constante  $K$  fixe le seuil en deçà duquel le module du gradient n'intervient plus. Ainsi, selon les termes des auteurs, le coefficient de diffusion encourage la régularisation intra-région plutôt que la régularisation inter-régions, ce qui permet de préserver l'intégrité des frontières. L'inconvénient de cette approche réside dans l'absence d'une solution explicite à l'équation. La solution proposée est calculée à travers un schéma numérique. D'autre part, cette contrainte implique un réglage manuel du nombre d'itérations sans oublier qu'il faut aussi choisir ou estimer (à travers une estimation du bruit de l'image) la constante  $K$ . Les itérations successives produisent des partitions en zones plates qui respectent la localisation des frontières entre les régions. La figure 1.2 montre les modules des gradients de l'image après pour différents nombres d'itérations. On voit clairement que la

délocalisation des frontières aux échelles grossières produites par l'espace-échelle gaussien est, dans le cas d'un représentation espace-échelle non linéaire, complètement absente.



**FIGURE 1.2 – Espace-échelle non-linéaire :** la figure représente le module du gradient de l'image après une suite de diffusions anisotropiques : Nbr d'itérations = 10, 30, 60, 120, 180, 240, 300. Les coupes successives du module du gradient montrent clairement la préservation de la localisation des frontières ainsi qu'une qualité de frontières supérieure à celle obtenue par le scale-space gaussien.

A travers ces deux exemples d'espace-échelle, on voit clairement la relation préviligierée qu'entretient l'analyse multi-échelle avec les équations aux dérivées partielles. Dès lors, différentes représentations espace-échelle ont été proposées en s'appuyant sur le formalisme des EDP. Grâce à la caractérisation de l'analyse multi-échelle par Alvarez, Guichard, Lions, et Morel [1993] et tout récemment dans Guichard et Morel [2002], une classification des différents espaces-échelles a été établie. Ces représentations multi-échelles se distinguent par les hypothèses d'invariances posées, et qui par ailleurs permettent de garantir l'unicité de la solution. On peut citer pour exemple l'analyse linéaire isotrope, l'invariance par changement de contraste, l'invariance affine, .... On peut encourager le lecteur intéressé par ce sujet à consulter des ouvrages tels que Guichard et Morel [2002], Weickert [1998].

### 1.3.1.1 Echelle(s) d'analyse

La question de la continuité de l'échelle d'analyse dans la représentation espace-échelle est théoriquement acquise et indiscutable. Néanmoins, dans la pratique, l'échelle est discrétisée et l'image est analysée à travers différentes coupes, extraites à différentes échelles. Il se pose alors le problème du choix des échelles pertinentes qui permettent d'analyser les structures significatives de l'image. Cette remarque est très bien résumée dans Lindeberg [1998] par un extrait de son introduction : "Une représentation multi-échelle en elle même ne contient aucune information explicite sur les structures de l'image qui devraient être perçues comme pertinentes et elle ne nous renseigne en rien sur les échelles appropriées pour les traiter. Ainsi, à moins que l'on ait une connaissance à priori sur les structures qui devraient être considérées comme importantes, la représentation espace-échelle nous délivre une importante quantité de données, à l'inverse d'une analyse mono-échelle, dont l'interprétation est remise à une étape

*ultérieure incluse dans la chaîne de traitements". En résumé, faute de connaissances à priori de haut-niveau sur le contenu de l'image, la manipulation d'une représentation espace-échelle demeurera fortement interactive.*

Les premiers travaux dans cette direction ont été réalisés par Bischof et Caelli [1988], qui ont proposé une analyse de stabilité spatiale, basée sur l'étude du laplacien, afin de reconstruire les frontières d'une région d'intérêt. Dans Lindeberg et Eklundh [1992] et Lindeberg [1991], une approche différente est adoptée. Elle prend en compte les structures dites **blob-like** à différentes échelles dans l'espace-échelle. Cette méthode aboutit à la construction d'une représentation multi-échelle d'un arbre (tree-like), appelée "scale-space primal sketch". En se basant sur une mesure de pertinence, exprimée par le volume qu'occupe les blobs dans la représentation, les auteurs extraient les échelles où ces blobs atteignent leur maximum.

Dans Lindeberg [1998], l'auteur propose un principe de sélection d'échelles qui s'appuie sur les dérivées des filtres gaussiens. L'idée repose sur l'étude des propriétés d'évolution à travers l'échelle des descripteurs différentiels normalisés. Plus précisément, il suggère que les *extrema* locaux le long de l'axe des échelles de ces entités différentielles normalisées, sont les plus à même de correspondre à des structures pertinentes de l'image. Cette approche a été appliquée avec succès dans des problèmes de détection de blobs, de détection de jonctions et d'estimation de la fréquence Lindeberg [1996]. Cependant, l'auteur conclut qu'il n'y a pas de notion d'échelle optimale pour l'interprétation d'une image, si l'analyse demeure de bas-niveau. L'approche qu'il a développé doit être interprétée comme une méthode systémique adaptée au cas où aucune information *a priori* sur le contenu de la scène n'est fournie. Sporring, Colios, et Trahanias [2000] se sont aussi intéressés au problème de la sélection automatique d'échelle. Ils ont proposé une méthode générique basée sur les fonctions locales de Lyaponov ainsi que sur les histogrammes locaux. Ils démontrent que leur approche généralise la méthode des détecteurs de blobs proposé par Lindeberg et Eklundh [1998].

#### 1.3.1.2 Espace-échelle et vision par ordinateur

Le socle théorique solide sur lequel repose le(s) représentation(s) espace(s)-échelle, ont permis de développer des applications réelles dans des domaines divers et variés. Ainsi, dans l'imagerie médicale, on peut citer des travaux tels que Jang Ben [1994], Khashman et Curtis [1997], Lorenz et al. [1997], Undeman et Lindeberg [2002], Lindeberg et Roland [1999], Haber et Modersitzki [2005], Henn et Witsch [2005], etc. Ces travaux touchent aussi bien à la segmentation, au recalage qu'à l'analyse des formes sans oublier la détection des contours et des points caractéristiques. Dans le domaine de la vidéo, on retrouve aussi de nombreuses références. On pourra se référer aux travaux réalisés dans Izquierdo [2000], Aldershoff et Salden, Kitahara et Ohta [2003] ou Myung-Cheol et al. [2006], .... Le domaine de l'indexation a également connu quelques contributions issues de la théorie des espaces-échelle telles que Mokhtarian et al. [1996], Dengsheng et Guojun [2003], Ceccarelli et al. [2005], ....

Il serait vain de lister tous les domaines où cette théorie a contribué de manière directe ou indirecte à résoudre des problèmes spécifiques à des applications. La preuve indéniable du succès de cette approche est la tenue de conférences spécialement dédiées à la théorie des espaces-échelle et à ses récents développements.

### 1.3.2 Approches hiérarchiques

Avant d'exposer dans le détail l'approche ensemble-échelle, pendant ensembliste de la théorie espace-échelle, il est nécessaire de s'arrêter sur les différentes approches qui ont pour objectif de décrire l'image non pas sous forme d'une simple partition mais sous la forme d'une hiérarchie de régions, d'une suite de partitions monotone ou encore sous la forme d'une pyramide de segmentation. Il est à noter que la majorité de ces approches ne peuvent pas être qualifier de *multi-échelle* au sens strict du terme comme dans l'approche espace-échelle, mais plutôt de *multi-niveau* car elles tentent d'exhiber différents degrés de détails de l'image sur plusieurs niveaux.

#### 1.3.2.1 Méthodes de segmentation par fusion/division de régions

Les techniques de segmentation par fusion et/ou division de régions sont parmi les premières approches proposées pour résoudre le problème de segmentation d'une image (Horowitz et Pavlidis [1976]). Le formalisme heuristique de ces approches se base sur l'idée intuitive qu'une « bonne » segmentation d'une image doit être composée de régions homogènes. C'est cette notion d'homogénéité ou d'hétérogénéité de régions que les premiers travaux se sont attachés à définir et à quantifier.

Les premières techniques proposaient de définir l'homogénéité d'une région à travers un prédicat booléen sur les régions de l'image. La définition du prédicat repose sur le choix d'une mesure d'homogénéité sur la région et sur la valeur d'un seuil. Une fois le seuil fixé, toute région dont la mesure d'homogénéité dépasse ce seuil est considérée comme non-homogène et par conséquent le prédicat est faux. Des exemples de mesures d'homogénéité de régions ne sont pas nombreux et les plus couramment utilisées sont (voir Cocquerez et Philipp [1995]) :

1. le contraste global :

$$C_g(R) = \max_{x \in R} I(x) - \min_{x \in R} I(x),$$

2. le contraste local :

$$C_{loc}(R) = \max\{|I(x) - I(y)| : (x, y) \in R^2, x \leftrightarrow y\},$$

3. la variance d'une région :

$$Var(R) = \frac{1}{|R|} \sum_{x \in R} (I(x) - \mu_R)^2,$$

tel que  $|R|$  représente la superficie de la région  $R$  et  $\mu_R$  représente la moyenne des niveaux de gris de la région,

4. l'entropie d'une région :

$$S(R) = - \sum_{l \in I^{-1}(R)} p(l) \log_2(p(l)),$$

tel que  $I^{-1}(R)$  représente l'ensemble des niveaux de gris de la région  $R$  dans l'image  $I$ ,

5. l'étendue interquartile :

$$W(R) = f_{3/4} - f_{1/4},$$

tel que  $f_\alpha$  représente le quartile d'ordre  $\alpha$  de l'histogramme,

6. les différences limitées :

$$D_l(R) = \max_{x,y \in R} |I(x) - I(y)|,$$

tel que les pixels  $x$  et  $y$  sont voisins,

7. etc.

Le principe algorithmique des premières techniques de fusions est assez trivial. Il s'agit, à partir d'une sur-segmentation de l'image, de parcourir toute l'image et de fusionner des paires de régions voisines qui vérifient le prédictat. Les pixels sont visités systématiquement et deux à deux. Un prédictat « P » est testé et s'il est vérifié, les pixels sont regroupés en une région. Une fois l'image parcourue, les régions résultantes se voient appliquer le même test et éventuellement les mêmes conséquences. Ce processus est itéré jusqu'à ce que le prédictat ne soit plus vérifié par aucune fusion de paire de régions. On obtient alors une partition dite, **2-maximale**<sup>10</sup>.

On peut consulter les travaux de Horowitz et Pavlidis [1976]<sup>11</sup>, Gagalowicz et Monga [1986], Rosenfeld et Kak [1982], Jain [1989], etc.

A travers la définition d'un critère d'arrêt (que cela soit le critère de 2-maximalité, le cardinal de la partition, la taille de la plus petite région, etc.), l'objectif est d'aboutir à une unique solution au problème de segmentation. Mais en réalité, cette stratégie permet également de construire une suite de partitions monotone et par conséquent une hiérarchie de régions. Cette hiérarchie permet d'exhiber une description **multi-niveau** de l'image.

Le choix délicat qui se pose dans ce type de méthodes (approches par prédictats) est celui du seuil, qui dépend fortement de la dynamique de chaque image, ce qui rend ces méthodes peu robustes.

L'inconvénient de ces méthodes réside également dans la faiblesse de l'information sur laquelle elles s'appuient pour fusionner des régions. En effet, seul un attribut radiométrique caractérise chaque fusion possible. De plus, des mesures telles que le contraste global ou local sont peu robustes au bruit. Par ailleurs, ces critères d'homogénéité se focalisent sur la similarité intra-région, sans se préoccuper de la dissimilarité inter-régions qui permet de caractériser localement le contraste entre une région et ses voisines (contraste relatif). Par ailleurs, aucune information géométrique sur les frontières des objets n'est intégrée dans la mesure d'homogénéité. Il est clair que ce type de méthodes ne se soucie pas de rechercher un modèle de segmentation et il en résulte des partitions comportant des régions très hétérogènes.

Un autre prédictat également utilisé est celui qui se focalise sur la dissimilarité inter-régions : *La région  $R$  est distincte de ses voisines*. Dans ce cas là, on utilise des mesures d'hétérogénéité, appelées aussi mesures de dissimilarité, telles que :

1. l'écart inter-régions moyen :

$$D_{moy}(R_i, R_j) = |\mu_{R_i} - \mu_{R_j}|,$$

10. Une partition est dite 2-maximale, si i) toutes ses régions vérifient le prédictat d'homogénéité et ii) toute nouvelle fusion de paires de régions ne vérifie plus ce même prédictat

11. La méthode proposée par Horowitz et Pavlidis repose sur la représentation d'une image par une pyramide « *quad-tree* » et combine fusion et division pour arriver à une segmentation de l'image

2. la distance inter-médianes :

$$D_{med}(R_i, R_j) = |f_{R_i,0.5} - f_{R_j,0.5}|,$$

tel que  $f_{R,0.5}$  désigne le deuxième quartile, appelé aussi la médiane, des niveaux de gris de la région  $R$ .

- 3. les contrastes moyens ou minimums aux frontières,
- 4. etc.

Cependant, ces prédicts souffrent également des mêmes inconvenients, à savoir le choix du seuil d'hétérogénéité, une information radiométrique faible pour caractériser les régions, pas d'informations géométriques intégrées et pas de modélisation du processus « image ».

D'autres auteurs se sont principalement focalisés sur la minimisation directe de critères de dissimilarité entre régions. Des travaux tels que ceux de Haris et al. [1998] ou de Brox et al. [2001], ont ainsi défini différents critères de dissimilarité entre régions adjacentes. Les critères de dissimilarité, appelés aussi fonctions de coût ou coûts d'agrégation, les plus utilisés sont issus de l'analyse de données et de la statistique inférentielle.

Ainsi, dans l'analyse des données multi-dimensionnelles, la classification hiérarchique ascendante (**C.H.A**) propose de construire une hiérarchie indicée d'un ensemble d'observations  $\Omega$  sur lequel une mesure de dissimilarité a été pré-définie. Elle considère dans un premier temps que chaque singleton de l'ensemble des données est une classe. Elle enchaîne ensuite une série de fusions entre classes dites « similaires » jusqu'à l'obtention d'une seule classe. Pour se faire, il est nécessaire de définir une distance de dissimilarité entre parties de l'ensemble  $\Omega$  à partir de la distance de dissimilarité définie sur les éléments de  $\Omega$ . Cette distance, appelée « critère d'agrégation », a fait l'objet de nombreuses investigations et les plus utilisées sont le critère du lien minimum, celui du lien maximum, le critère de la distance moyenne et le critère d'inertie (critère de Ward). Il est à noter que la notion de voisinage (topologie) est inexistante dans la classification hiérarchique ascendante.

Les techniques de segmentation par fusion ont emprunté bon nombre de critères d'agrégation de la **C.H.A** et en ont introduit de nouveaux, afin de construire une hiérarchie ou une suite de partitions monotone de l'image. On citera pour exemples le critère de Ward [1963], le critère de la différence de moyenne de luminance<sup>12</sup>, le critère Mean-Ward<sup>13</sup> ou le critère de luminance-linéaire. D'autres mesures de dissemblance se sont appuyées sur des tests statistiques pour faire une comparaison entre les distributions des régions à fusionner ou certains de leurs moments. On peut citer pour exemple le test du **Chi2** ( $\chi^2$ ), le test de student d'égalité d'espérances, le test de Fisher-Snedecor d'égalité des moyennes et des variances, le test de Fisher-Hotelling, etc. Brox et al. [2001], après avoir étudié les avantages et les inconvenients de certains critères, proposent un processus de fusion multi-étape (Multi-Stage). A chaque phase, un critère d'agrégation différent est utilisé dans le processus de fusion. Le passage d'un critère à un autre s'appuie sur un seuil fixé par avance portant sur les valeurs du critère. Les auteurs préconisent de passer à un autre critère dès que le critère d'agrégation croît d'une

---

<sup>12</sup>. Il s'agit du carré de l'écart entre les moyennes de deux régions

<sup>13</sup>. Il s'agit tout simplement de la combinaison entre le critère de Ward et le précédent critère, à travers une simple multiplication.

façon abrupte ce qui traduit selon eux la non-adaptabilité du critère à la complexité des régions. Cependant, la sélection des critères de fusion reste purement empirique (Ils préconisent 3 critères qu'on cite dans l'ordre de leur utilisation : Ward, Mean-Ward, Border). Par ailleurs, le choix du seuil demeure une opération assez délicate et peu robuste. La validation expérimentale se contente uniquement d'une application sur des images contenant des formes très régulières.

D'autres auteurs, conscients que l'information sur les contours peut améliorer la qualité de la segmentation, ont proposé de combiner des critères géométriques avec des prédictifs d'homogénéité comme dans Tamaki, Yamamura, et Ohnishi [1999]. L'approche proposée par ces auteurs introduit néanmoins un nombre de paramètres et de seuils assez conséquents, ce qui la rend dépendante d'un lourd réglage. Les critères géométriques reposent généralement sur l'amplitude du gradient le long de la frontière partagée par les deux régions ou sur le périmètre le long de ces dernières.

Ces algorithmes de segmentation ont été le plus souvent appliqués sur quelques images synthétiques ou réelles mais sans être réellement évalués qualitativement. Il n'existe pas d'exemples, à notre connaissance, d'applications réelles exploitant ces méthodes, dans leurs chaînes de traitement.

A l'inverse des méthodes par fusion de régions, les techniques de division sont des méthodes descendantes (top-down) où le processus est également itératif. L'image originale est dans un premier temps considérée comme une seule région rectangulaire. A chaque étape, toute région déclarée hétérogène par le biais d'un prédictif est divisée en quatre régions rectangulaires. Horowitz et Pavlidis [1976] ont été les premiers à présenter une méthode de segmentation qui permet de décrire l'image sous forme d'une hiérarchie quad-tree. L'approche par *Split and Merge* de ces auteurs s'appuie sur une formulation par prédictifs d'homogénéité consistants mais intègre à la fois un mécanisme de fusion et de division. L'inconvénient de cette approche est l'absence d'information sur la géométrie des régions. De plus, la forme contrainte du quad-tree ne permet pas d'épouser les lignes de contraste. A noter que cette pyramide, exhibe certes des niveaux de détails différents, mais la notion d'échelle dans ce cas là n'est rien d'autre que le niveau de la pyramide. Le processus est achevé quand toutes les régions sont homogènes. Les méthodes de division n'ont pas connu de succès car les contours obtenus ne respectent pas la géométrie de l'image d'une part et le coût d'exécution est assez prohibitif. La difficulté vient du fait que les calculs effectués à une étape ne sont généralement pas exploitables dans les étapes suivantes ce qui ralentit considérablement le processus de segmentation. Par ailleurs, diviser une région revient à créer une information tandis que fusionner des régions revient juste à éliminer une information déjà existante.

### 1.3.2.2 Approches pyramidales

Certains auteurs ont proposé des méthodes de segmentation visant à représenter l'image sous forme d'une séquence de partitions monotones (pyramide). Les algorithmes de pyramides stochastiques de Meer [1989], irrégulières de Montanvert, Meer, et Rosenfeld [1991] ou adaptatives de Jolion et Montanvert [1982] s'appuient sur un procédé de décimation stochastique des noeuds d'un graphe d'adjacence de régions utilisant la notion d'ensemble stable (Meer 1989). Le processus est itéré afin de réduire la taille du graphe. 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-tree et dans les pyramides multi-résolution. L'idée est toutefois de considérer des ensembles de géométrie irrégulière afin de pallier les importants défauts des pyramides à géométrie rigide. Là aussi, le paramètre d'échelle dans les approches conduisant à la construction d'une pyramide irrégulière ou adaptative, est simplement le niveau de la pyramide.

### 1.3.2.3 Approches énergétiques

Jusqu'à présent, la plupart des approches citées dans cette section ne respecte aucun schéma de modélisation. Elles envisagent le problème de segmentation de manière très locale en se focalisant sur l'entité région pour construire une partition de l'image. Au début des années quatre vingt, plusieurs auteurs ont commencé à concevoir la segmentation en régions d'une image comme un problème d'inférence d'un modèle. La détermination d'une partition passe alors par la définition d'un espace de modèles, dans lequel s'effectuera la recherche de la partition, et par la définition d'une énergie mesurant l'adéquation du modèle aux données. Ainsi, la segmentation se présente sous la forme d'un problème d'optimisation où l'objectif est de minimiser l'énergie sur l'espace des modèles.

De nombreux algorithmes de fusions de régions ont été proposés qui dérivent d'un formalisme énergétique. Ainsi, Koepfler, Lopez, et Morel [1994] pose le problème de segmentation comme un problème d'inférence d'un modèle à travers la minimisation d'une fonctionnelle. Ils utilisent la fonctionnelle constante par morceaux de Mumford-Shah (Mumford et Shah [1989]) comme énergie pour évaluer la qualité de la segmentation par rapport au modèle. Ils s'appuient ensuite sur une méthode de descente de gradient associée à un mécanisme de fusion de régions. L'algorithme est initialisé à partir d'une partition  $P_1$ , supposée être une sur-segmentation de la solution optimale. L'algorithme produit alors une suite de partitions monotone où la partition  $P_{k+1}$  est définie comme la partition d'énergie minimale couvrant la partition  $P_k$ . Le critère d'arrêt peut être implicite, autrement dit le processus de fusion s'arrête quand aucune fusion de régions ne minimise l'énergie. Les auteurs Koepfler, Lopez, et Morel [1994] parlent alors de 2-normalité ou normalité de la partition<sup>14</sup>. D'autres auteurs comme Lee [1998] ou Lee [2000] proposent de poursuivre la fusion même si l'énergie croît à certaines étapes, jusqu'à ce que l'image entière soit fusionnée. Il sélectionne ensuite 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. On pourra citer d'autres travaux reposant sur un formalisme d'inférence de modèle et couplant un mécanisme de descente avec un processus de fusion tels que Fuchs et Le Men [1999], Ballester et al. [1994], Kanungo et al. [1995].

On notera que la première tentative de segmentation - orientée-région - et multi-échelle, s'appuyant sur une minimisation d'une énergie est également due à Koepfler et al. [1994] que l'on vient d'expliciter ci-dessus. Le paramètre de pondération entre le terme d'attache aux données et le terme de régularisation est discrétisé dans un premier temps pour former une suite croissante. Ce terme joue alors le rôle d'échelle d'analyse. Cependant, la pyramide obtenue ne présente en aucun cas une analyse en échelle continue du fait de la discrétilisation du paramètre de pondération. On peut citer d'autres travaux similaires à l'approche adoptée par

---

<sup>14</sup>. Une partition est dite 2-normale lorsque aucune partition couvrante ne minimise l'énergie. On parle de normalité quand aucune sous-partition de la partition courante ne minimise l'énergie.

Koepfler et al. [1994] tels que Ballester et al. [1994], Fuchs [2001]. A chaque fois, l'algorithme utilise le principe de descente énergétique pour réaliser les fusions en partant d'une sur-segmentation de l'image. Le paramètre d'échelle est aussi discrétisé et la suite d'échelle construite est croissante dans le but d'obtenir des partitions de plus en plus grossières.

#### 1.3.2.4 Approches morphologiques

La théorie morphologique des opérateurs connexes a permis dans les années 90 d'aboutir à des approches de segmentation pyramidale. On pourra citer pour exemple les travaux de Serra et Salembier [1993], Salembier et Pardàs [1994], Salembier et Serra [1995], Salembier et al. [1998]. Le principe de ces approches consiste à appliquer des filtres connexes tels que les filtres de reconstruction, qui élargissent graduellement les zones plates d'une image et produisent donc des partitions en zones plates de finesse décroissante. Toutes ces approches dites pyramidales, utilisent une notion d'échelle, mais l'axe des échelles est également discrétisé.

L'approche adoptée par Meyer (voir Meyer [1999a;b]), qui s'inspire de l'algorithme classique de calcul d'une ligne de partage des eaux par simulation d'immersion, permet de produire une analyse multi-échelle continue de l'image. L'approche simule une montée continue du niveau d'eau qui au passage permet de fusionner des régions dès que deux bassins se rencontrent. 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 Meyer [2001], Zanoguera et al. [1999].

## 1.4 APPROCHE ENSEMBLE-ÉCHELLE

Nous allons maintenant exposer de manière brève l'approche « Ensemble-échelle » (en anglais *scale-set*) introduite par Guigues, Cocquerez, et Le Men [2006]. Cette approche permet une analyse en échelle continue de l'image et décrit cette dernière sous forme d'une hiérarchie de régions dont les coupes horizontales sont les partitions optimales au sens d'une énergie affine.

#### 1.4.1 Définitions et Axiomes

Soit  $\mathcal{D}$  un ensemble ouvert représentant le support de l'image  $I$  et notons par  $\mathcal{P}(\mathcal{D})$  l'ensemble de ses parties. On appelle une partition  $P$ , un ensemble de parties de  $\mathcal{D}$  tel que :

$$\forall R, R' \in P, \quad R \cap R' = \emptyset \quad \text{et} \quad \bigcup_{R \in P} R = \mathcal{D}.$$

On notera  $\mathbb{P}(\mathcal{D})$  l'espace des partitions possibles de  $\mathcal{D}$ . On dit qu'une partition  $P$  est plus fine qu'une partition  $Q$ , ce qu'on notera  $P \leq Q$ , si et seulement si :

$$\forall R \in P, \exists R' \in Q : R \subseteq R'.$$

On rappelle qu'une hiérarchie  $H$  sur  $\mathcal{D}$  est un ensemble de parties de  $\mathcal{D}$  qui vérifient les propriétés suivantes :

1.  $\emptyset \notin H$ ,

2.  $\mathcal{D} \in H$ ,
3.  $\forall x \in \mathcal{D} : \{x\} \in H$ ,
4.  $\forall R, R' \in H : R \cap R' \in \{\emptyset, R, R'\}$ .

On appellera la base d'une hiérarchie l'ensemble, noté  $H_B$ , et définie par :

$$H_B = \{\{x\}, x \in H\}.$$

Si l'on se place du point de vue d'un problème de segmentation d'image,  $\mathcal{D}$  représente le domaine spatial de l'image  $I$  et  $H_B$  représente la plus fine partition de l'image ou en d'autre terme la sur-segmentation absolue de  $I$ . Dans la pratique cette dernière peut prendre la forme d'une partition en pixels individuels ou d'un ensemble de régions homogènes.

La relation  $\vee$  définie sur  $H \times H$  désignera le plus petit majorant commun (*supremum*) entre deux éléments  $A$  et  $B$  de la hiérarchie  $H$  :

$$x \vee y = \inf\{z \in \mathcal{D} \mid x \subseteq z \text{ et } y \subseteq z\}.$$

Une hiérarchie  $H$  peut être représentée par un arbre  $\mathcal{T} = (H, \mathcal{E})$ , plus communément appelé : dendrogramme (voir Figure 1.3). Les nœuds de l'arbre  $\mathcal{T}$  représentent les régions de la hiérarchie  $H$  (la racine de l'arbre  $\mathcal{T}$  étant le domaine entier de l'image  $\mathcal{D}$ ), tandis que les arêtes  $\mathcal{E}$  de l'arbre  $\mathcal{T}$  correspondent à la relation de couverture ( $\subset$ , i.e l'inclusion stricte) entre les éléments de  $H$ . On notera par  $\mathcal{F}(x)$  le père d'un nœud  $x$  et par  $\mathcal{S}(x)$  l'ensemble des nœuds fils du nœud  $x$ .

Pour tout nœud  $x$  appartenant à la hiérarchie  $H$ , le sous ensemble de  $H$  définit par  $H(x) \triangleq \{y \in H, y \subseteq x\}$  est une hiérarchie sur  $x$ , appelée "*hiérarchie partielle*" induite par  $x$ , (voir Figure 1.3).

Une *coupure*,  $C$ , dans une hiérarchie  $H$  est un sous ensemble de nœuds de la hiérarchie  $H$  formant une partition du domaine  $\mathcal{D}$  de l'image (voir Figure 1.3).

On notera par  $\mathbb{C}(H)$  (resp.  $\mathbb{C}(H(x))$ ) l'ensemble de toutes les coupes possibles dans une hiérarchie  $H$  (resp. dans une hiérarchie partielle  $H(x)$  induite par  $x$ ).

Le nombre de coupes dans une hiérarchie est en général très élevé mais il reste largement inférieur à celui de l'ensemble de toutes les partitions possibles du support  $\mathcal{D}$ , qu'on notera  $\mathbb{P}(\mathcal{D})$ .

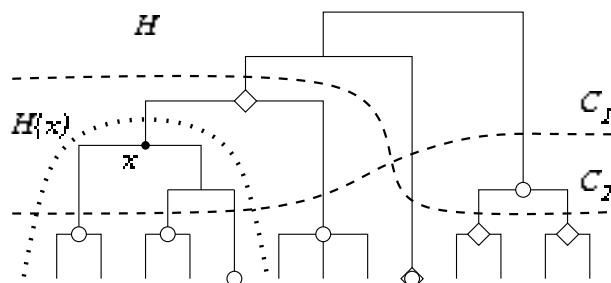


FIGURE 1.3 – Un exemple d'une hiérarchie  $H$  représentée par un arbre.  $H(x)$  est la hiérarchie partielle induite par  $x$ .  $C_1$  et  $C_2$  sont deux coupes quelconques qui forment deux partitions distinctes d'éléments de  $H$ .

On appelle un *critère croissant* sur une hiérarchie  $H$  toute application  $\lambda : H \rightarrow \mathbb{R}^+$  vérifiant la propriété suivante :

$$\forall x, y \in H, x \subseteq y \Rightarrow \lambda(x) \leq \lambda(y).$$

Une *hiérarchie indicée* est un couple  $(H, \lambda)$ , où  $H$  est une hiérarchie et  $\lambda$  est un critère croissant sur  $H$ , nul sur ses feuilles. Dans ce cas, le critère  $\lambda$  est appelé *indice*. L'indice de la hiérarchie permet de tracer les noeuds de la hiérarchie  $H$  en leur affectant des ordonnées dans la représentation.

Une distance ultramétrique  $\delta$  sur un ensemble  $\mathcal{D}$  est une distance qui vérifie toutes les propriétés classiques d'une distance métrique et une inégalité plus forte que l'inégalité triangulaire :

$$\forall x, y, z \in \mathcal{D} \quad \delta(x, z) \leq \max(\delta(x, y), \delta(y, z)).$$

Si  $(H, \lambda)$  est une hiérarchie indicée alors la fonction  $\delta : \mathcal{D} \times \mathcal{D} \longrightarrow \mathbb{R}^+$  définie par

$$\delta(x, y) = \lambda(\{x\} \vee \{y\}), \quad (1.1)$$

est une distance ultramétrique sur  $\mathcal{D}$  où  $\vee$  représente le *supremum* au sens de l'inclusion sur la hiérarchie  $H$ .

Notons par  $E : \mathbb{P}(\mathcal{D}) \longrightarrow \mathbb{R}^+$  l'énergie d'une partition  $P$ . On dit que  $E$  est une énergie **séparable**, si et seulement si :

$$E(P) = \sum_{R \in P} E(R).$$

Dans un premier temps, Guigues et al. [2006] partent de l'hypothèse que la hiérarchie de régions de l'image  $I$  a été construite par un quelconque algorithme. Ils démontrent alors que la coupe optimale (au sens d'une énergie  $E$  séparable) d'une région quelconque de la hiérarchie est reliée aux coupes optimales de ses noeuds fils par une simple relation de récurrence.

Cette propriété remarquable est résumée dans la proposition suivante :

**Proposition 1.1** Soit  $\mathcal{C}^*(x)$  la coupe de la hiérarchie partielle  $H(x)$  qui minimise l'énergie  $E$ , i.e :

$$\mathcal{C}^*(x) = \underset{\mathcal{C} \in \mathbb{C}(H(x))}{\operatorname{argmin}} E(\mathcal{C}) = \underset{\mathcal{C} \in \mathbb{C}(H(x))}{\operatorname{argmin}} \sum_{y \in \mathcal{C}} E(y).$$

et soit  $E^*(x) \triangleq E(\mathcal{C}^*(x))$ , alors :

$$\forall x \in H, E^*(x) = \inf \left\{ E(x), \sum_{y \in \mathcal{S}(x)} E^*(y) \right\} \quad (1.2)$$

tel que  $\mathcal{S}(x)$  est l'ensemble des noeuds fils de la région  $x$  dans la hiérarchie  $H$ .

Cette relation de récurrence permet alors de calculer la coupe optimale d'une hiérarchie  $H$  en utilisant une procédure de programmation dynamique. Elle nécessite uniquement une étape ascendante durant laquelle les noeuds optimaux ainsi que leurs énergies partielles respectives seront collectés. Les noeuds optimaux les plus élevés dans la hiérarchie, formeront alors la partition optimale de la hiérarchie au sens de l'énergie  $E$ .

Insistons sur le fait que la partition optimale obtenue l'est sur l'espace des partitions issues de la hiérarchie  $H$ , qui est supposée préalablement construite par une méthode de fusion ou de division.

Dans un second temps, les auteurs ont considéré une famille d'énergies, appelée *Energies affines séparables*.

Un énergie affine séparable sur une partition  $P$  d'une image, est définie comme suit :

$$E_\lambda(P) = \sum_{R \in P} \lambda C(R) + D(R)$$

tel que  $D$  représente un terme d'attache aux données qui exprime la qualité de régression,  $C$  représente un terme de régularisation et  $\lambda \in \mathbb{R}$  est un paramètre de pondération.

Les deux énergies,  $D$  et  $C$ , sont antagonistes. La première exprime la qualité de régression de la partition aux données observées (l'image originale) tandis que la deuxième mesure le degré de complexité de la partition. La solution optimale pour une énergie d'attache aux données est la sur-segmentation absolue alors que l'énergie de régularisation cherchera à sélectionner les partitions les moins complexes au sens d'une certaine mesure de complexité (régularité des frontières, convexité des contours des régions, coût du codage, etc.).

Ainsi, le paramètre  $\lambda$  joue le rôle d'un coefficient de pondération entre les deux énergies. Pour chaque valeur que prendra le paramètre  $\lambda$ , on peut calculer une partition optimale de la hiérarchie  $H$ . Néanmoins, lors d'une croissance du paramètre rien ne garantit que les partitions optimales seront monotones. L'objectif principal est donc de définir les conditions pour lesquelles les coupes optimales, qui sont indiquées par le paramètre  $\lambda$ , forment une séquence monotone.

Notons par  $\mathcal{C}_\lambda^*(H)$  la coupe optimale (appelée  $\lambda$ -coupe) de  $H$  à l'échelle  $\lambda$ . Par ailleurs, supposons que l'énergie de régularisation  $C$  est une énergie décroissante sur l'ensemble des partitions  $\mathbb{P}(\mathcal{D})$ , i.e :

$$\forall (P, Q) \in \mathbb{P}^2(\mathcal{D}) \quad Q \geq P \Rightarrow C(Q) \leq C(P).$$

Cette propriété de décroissance sur l'espace des partitions est équivalente à la propriété de **sous-additivité** de cette énergie sur l'ensemble des régions  $\mathcal{P}(\mathcal{D})$ , à condition que l'énergie soit séparable, i.e :

$$\forall (R, S) \in \mathcal{P}^2(\mathcal{D}), \quad R \cap S = \emptyset \Rightarrow C(R \cup S) \leq C(R) + C(S).$$

Le théorème suivant décrit les conditions nécessaires et suffisantes pour que la famille de coupes  $(\mathcal{C}_\lambda^*(H))_{\lambda \in \mathbb{R}}$  forme une séquence croissante de partitions.

**Théorème 1.1** (Coupes minimales causales) *Soit  $E_\lambda = (C, D)$  une énergie affine séparable. Si l'énergie  $C$  est sous-additive, alors quelque soit la hiérarchie  $H$  sur  $\mathcal{D}$ , la séquence  $(\mathcal{C}_\lambda^*(H))_{\lambda \in \mathbb{R}}$  des coupes minimales de  $H$  pour l'énergie  $E_\lambda$  est une segmentation multi-échelle, en d'autres termes,  $(\mathcal{C}_\lambda^*(H))_{\lambda \in \mathbb{R}}$  est suite de partitions monotone.*

On pourra consulter la démonstration de ce théorème dans Guigues [2003]. La précédente équation récursive (1.2) est généralisée sous une forme fonctionnelle dans laquelle l'optimisation porte sur des fonctions de  $\lambda$  plutôt que sur une unique valeur d'énergie à échelle fixe :

$$\forall x \in H, \quad E_x^* = \inf \left\{ E_x, \sum_{s \in \mathcal{S}(x)} E_s^* \right\}, \quad (1.3)$$

tel que  $E_x : \lambda \mapsto E_x(\lambda) = \lambda C(x) + D(x)$  représente l'énergie de la région  $x$ , appelée *énergie-propre* et  $E_x^* : \lambda \mapsto E_x^*(\lambda) = E_\lambda^*(H(x))$  représente l'énergie de ses coupes minimales, appelée *énergie-partielle* de la région  $x$ .

Cette équation de programmation dynamique fonctionnelle peut se réécrire d'une autre manière :

$$E_x^*(\lambda) = \begin{cases} \sum_{s \in \mathcal{S}(x)} E_s^* & \text{si } \lambda < \lambda^+(x) \\ E_x & \text{sinon.} \end{cases} \quad (1.4)$$

Que suggère cette relation ? et quel est son intérêt en pratique ? Focalisons nous dans un premier temps sur la région  $x$  et sa hiérarchie partielle  $H(x)$ . Sa partition optimale, au sens de l'énergie  $E(\lambda)$  est l'*union* des coupes optimales des nœuds fils de  $x$  quand l'échelle est inférieure à une certaine quantité, notée  $\lambda^+(x)$ . Cette région,  $\{x\}$  devient la partition optimale quand l'échelle passe au dessus de cette valeur (voir Figure 1.4). Le saut qui survient à l'échelle  $\lambda^+(x)$  correspond alors à la fusion des régions fils du nœud  $\{x\}$ <sup>15</sup>. L'énergie-propre de  $x$  est une fonction affine, non-décroissante et concave. De même, la somme et l'intersection de fonctions affines par morceaux, non-décroissantes et concaves sont également des fonctions affines par morceaux, non-décroissantes et concaves. Par conséquent, l'énergie optimale  $E_x^*(\lambda)$  de  $x$  est une fonction affine, non-décroissante et concave, (voir Figure 1.4).

L'échelle à partir de laquelle la région  $\{x\}$  devient optimale est unique<sup>16</sup> et peut être calculée en résolvant l'équation suivante :

$$E_x(\lambda) = \sum_{s \in \mathcal{S}(x)} E_s^*(\lambda).$$

Ainsi, la fonction  $E_x^*$ , qui fournit les énergies des coupes optimales de  $H(x)$  en fonction du paramètre  $\lambda$ , est une simple fonction affine par morceaux. Elle peut être explicitement mémorisée pour chaque nœud de la hiérarchie  $H$ , typiquement par la liste des points terminaux de ses morceaux affines. Ainsi pour chaque nœud  $x$  le calcul de l'énergie  $E_x^*$  requiert uniquement deux opérations sur les fonctions affines par morceaux , une somme et une intersection avec une fonction affine (voir figure 1.4).

La relation de programmation dynamique fonctionnelle (1.3) permet alors de calculer  $\lambda^+$  (appelé *échelle d'apparition* de la région  $x$ ) et  $E^*$  pour toute région de la hiérarchie  $H$  en une seule phase ascendante (bottom-up) de  $H$ . Cependant, une phase descendante (top-down) de  $H$  est nécessaire afin d'élaguer les régions dites *non-persistantes*, càd celles dont les régions pères apparaissent avant ces mêmes régions.

Le signe de l'échelle d'apparition n'est pas nécessairement positif, excepté quand l'énergie radiométrique est sur-additive (i.e croissante selon la relation d'ordre partiel sur l'ensemble des parties de  $\mathcal{D}$ ), ce qui est généralement le comportement naturel des énergies de régression.

Si l'on note par :

$$\lambda^+(x) = \left\{ \lambda, E_x(\lambda) = \sum_{s \in \mathcal{S}(x)} E_s^* \right\}, \quad (1.5)$$

l'échelle d'apparition de la région  $x$  et par :

$$\lambda^-(x) = \min_{y \in H, x \subset y} \{\lambda^+(y)\}, \quad (1.6)$$

<sup>15</sup>. Rigoureusement, c'est la fusion des partitions optimales des régions fils du nœud  $x$  qui survient.

<sup>16</sup>. En toute rigueur, cette échelle d'apparition existe si et seulement si l'énergie géométrique  $C$  est strictement sous-additive.

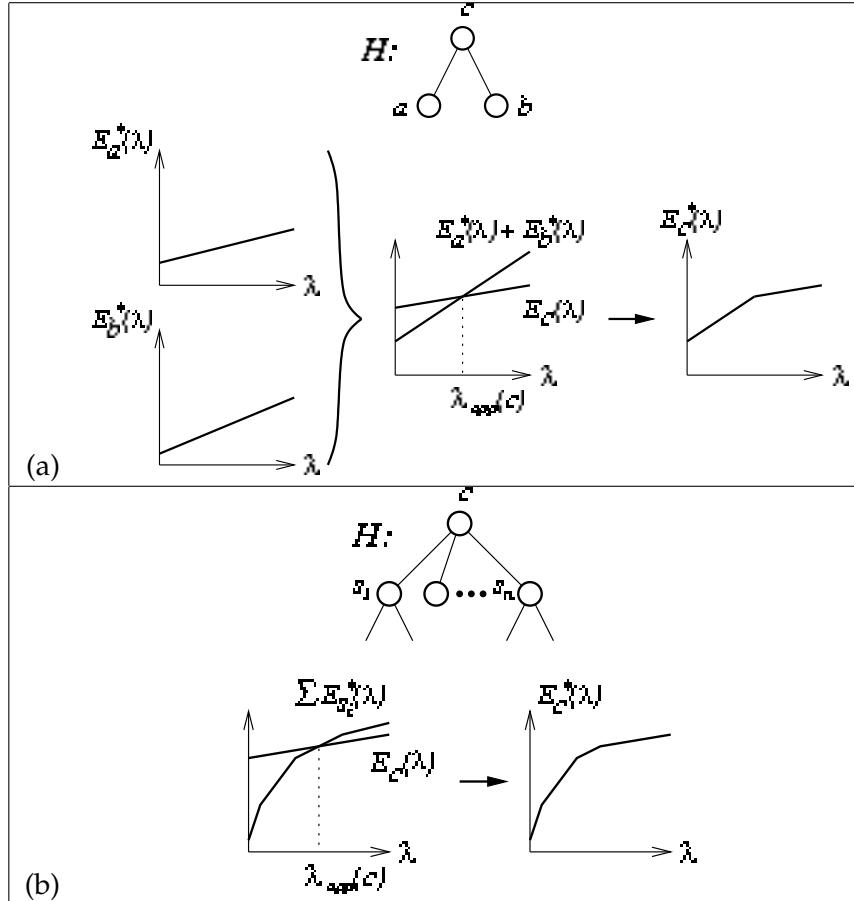


FIGURE 1.4 – Calcul de la fonction d'énergie partielle  $E_c^*$  ainsi que l'échelle d'apparition  $\lambda^+(c)$  d'un nœud  $c$  sachant les énergies  $E^*$  de ses nœuds fils. (a) Premier niveau du processus. (b) Cas général.

l'échelle de disparition de  $x$ , alors la nouvelle hiérarchie obtenue après l'étape d'élagage est définie par :

$$H^* \triangleq H - \{x \in H | \lambda^+(x) \leq \lambda^-(x)\}. \quad (1.7)$$

Ainsi, le couple  $(H^*, \lambda^+)$  est une hiérarchie indicée où l'échelle d'apparition des régions joue le rôle d'indice.

On appellera :

$$\Lambda(x) = [\lambda^+(x), \lambda^-(x)], \quad (1.8)$$

l'intervalle de persistance de la région  $x$  et la quantité :

$$\Delta\lambda(x) = \lambda^-(x) - \lambda^+(x), \quad (1.9)$$

désignera la *persistance* de la région  $x$ .

#### 1.4.2 L'algorithme d'escalade

La précédente section a permis d'expliquer comment construire et calculer une séquence complète de coupes minimales ainsi que l'énergie  $E_\lambda$  d'une hiérarchie construite au préalable par un autre algorithme.

Selon les termes des auteurs de l'approche « Ensemble-échelle », la  $\lambda$ -coupe, **escalade** la hiérarchie quand  $\lambda$  croît. Elle saute à une région  $x$  plus haute dans la hiérarchie à chaque fois que le paramètre  $\lambda$  atteint une échelle spécifique,  $\lambda^+(x)$ . Augmenter  $\lambda$  revient donc à régulariser la solution en durcissant la contrainte sur le modèle de complexité. Ce saut coïncide alors à la fusion d'un certain nombre de régions. Ainsi, en partant d'une partition initiale telle que la sur-segmentation absolue, l'algorithme recherche toutes les fusions possibles entre régions adjacentes à partir de cette partition. Sa stratégie est de choisir la fusion dont l'échelle d'apparition est la plus petite à chaque étape. Ce processus appelé *Escalade pure* est naturellement prohibitif en terme de temps de calcul<sup>17</sup>.

Afin d'aboutir à un algorithme raisonnable en terme de complexité, une autre stratégie dite d'*Escalade binaire* a été proposée. Elle consiste à se focaliser sur les fusions de paires de régions adjacentes. Cette stratégie est, par ailleurs, souvent adoptée dans les algorithmes de fusions de régions. Dans ce cas là, l'escalade binaire recherche un optimum local. L'initialisation de l'algorithme peut se faire par une partition fine construite à l'aide d'un algorithme de Ligne de Partage des Eaux Beucher et Meyer [1993], comme l'ont proposé les auteurs dans Guigues et al. [2006]. L'initialisation d'un algorithme de fusion par une ligne de partage des eaux a aussi été adoptée dans d'autres travaux, on peut citer pour exemple Haris et al. [1998].

Le choix d'utiliser une ligne de partage des eaux en lieu et place de la sur-segmentation absolue ou d'un pavage par dalle pixellaire a plusieurs avantages. D'une part la qualité des frontières obtenues avec cette technique permet de respecter les contours réels des objets de la scène, à l'inverse d'un pavage par dalle. D'autre part le temps de calcul est très significativement réduit contrairement à une initialisation s'appuyant sur une sur-segmentation absolue. La complexité combinatoire de la stratégie d'escalade binaire, est au pire de l'ordre de  $\mathcal{O}(N^2 \log N)$ , où  $N$  représente la taille de la sur-segmentation initiale (i.e le nombre de régions composant la partition de départ).

### 1.4.3 Résultats expérimentaux

#### 1.4.3.1 Choix d'une énergie affine

La construction d'un ensemble-échelle d'une image nécessite seulement le choix d'un couple d'énergies  $E = (C, D)$ , aucun réglage ou paramétrage préalable n'est requis. Cependant, ces deux énergies doivent respecter la condition de séparabilité. Par ailleurs, l'énergie de régularisation  $C$  doit être sous-additive et l'énergie d'attache aux données sur-additive.

Les énergies les plus couramment utilisées dans la segmentation d'images possèdent cette propriété de séparabilité telles que la fonctionnelle de Mumford-Shah (Mumford et Shah [1989]), le potentiel de Pott's pour les Champs de Markov aléatoire (MRF), (Geman et Geman [1984]), ainsi que la plupart des critères de description de la longueur (MDL) comme celui proposé par Leclerc [1989b].

Pour les énergies de régularisations sous-additives, les exemples sont multiples. La quantification de la complexité par le nombre de régions est une mesure décroissante. Les critères à base de somme d'une quantité géométrique positive

---

<sup>17</sup>. Concrètement, on doit évaluer pour chaque région de la partition initiale, toutes les fusions possibles avec ses régions voisines. La taille de l'ensemble des fusions possibles d'une région, dépend du voisinage de la région. Par ailleurs, la fusion peut être binaire, ternaire, quaternaire, etc.

tel que la longueur, la courbure, ..., le long des frontières de la partition forment aussi des énergies de régularisation sous-additive.

Le choix des énergies de régularisation et d'attache aux données a été longuement discuté dans la thèse de Guigues [2003] et nous n'allons pas exposer toutes les énergies explorées par les auteurs. Néanmoins, on peut citer deux énergies affines qui semblent donner des résultats satisfaisants. La première est celle qui s'appuie sur le modèle constant par morceaux de Mumford-Shah. Elle est composée d'un terme qui représente la somme des écarts quadratiques par rapport à la moyenne de la région et d'un terme quantifiant la longueur de la frontière de la région. La deuxième énergie affine possède le même terme d'attache aux données mais mesure la complexité de la région non pas par la longueur de sa frontière mais par sa concavité. Cette énergie introduite par les auteurs favorise la construction d'objets convexes :

$$C(R) = \frac{1}{2\pi} \sum_{i=1}^p |\theta_i|$$

tel que  $\{\theta_i\}_{i=\{1,\dots,p\}}$  représente la suite des angles de la polyligne décrivant la frontière de la région.

L'énergie de concavité n'est pas forcément sous-additive mais les auteurs contournent ce problème en posant  $\Delta C = 0$ , lorsque l'énergie devient sur-additive. On interdit ainsi la fusion entre régions dont l'union cause un accroissement de l'énergie de régularisation puisque l'échelle d'apparition de l'union est l'infinie ( $+\infty$ ).

La figure 1.5 illustre la hiérarchie construite avec l'énergie de Mumford-Shah. La partition initiale a été obtenue en utilisant l'algorithme de partage des eaux<sup>18</sup>. Les coupes horizontales présentées ont été extraites successivement en respectant l'ordre inverse des fusions. La structure de la hiérarchie divise clairement l'image en deux parties, la première décrit le ciel, tandis que la deuxième décrit les bâtiments et la rue. L'exploration en profondeur de la hiérarchie permet de découvrir les détails de l'image les plus pertinents. Ainsi des zones remarquables telles que les portes et les fenêtres du bâtiment principal apparaissent les premières tandis que les détails concernant le deuxième bâtiment, situé à gauche du bâtiment principal et occupant un espace moins prépondérant, apparaissent ultérieurement. Cet exemple illustre clairement l'avantage d'une description multi-échelle de l'image, il n'y a pas une échelle fixe de description mais une famille d'échelles qui permet de représenter le contenu de l'image du détail le plus grossier au détail le plus fin.

La figure 1.6 montre les coupes extraites des hiérarchies de régions de l'image « *Lena* ». La première colonne illustre les partitions de la hiérarchie obtenue avec l'énergie de Mumford-Shah tandis que la deuxième colonne présente les coupes obtenues avec l'énergie affine associant l'écart quadratique de la région et l'énergie de concavité présentée ci-dessus. Deux coupes se trouvant sur la même ligne, sont issues de deux énergies différentes avec un nombre de régions quasi-identique. Les deux descriptions sont très similaires, néanmoins on relèvera que la région décrivant le haut du chapeau a fusionné dans la première colonne avec le fond tandis que dans la deuxième colonne le pourtour du chapeau a résisté plus longuement

<sup>18</sup>. La LPE utilisée est identique à celle utilisée par L. Guigues pour l'initialisation de son algorithme d'escalade. L'algorithme de cette LPE est du à Marc Pierrot Deseilligny, qui ne l'a jamais publié, mais qui l'a transmis oralement à L. Guigues. On appellera les grandes lignes de cet algorithme en annexe (voir B). La ligne de partage des eaux est calculée sur l'image gradient de l'image en utilisant le filtre auto-régressif de Deriche.

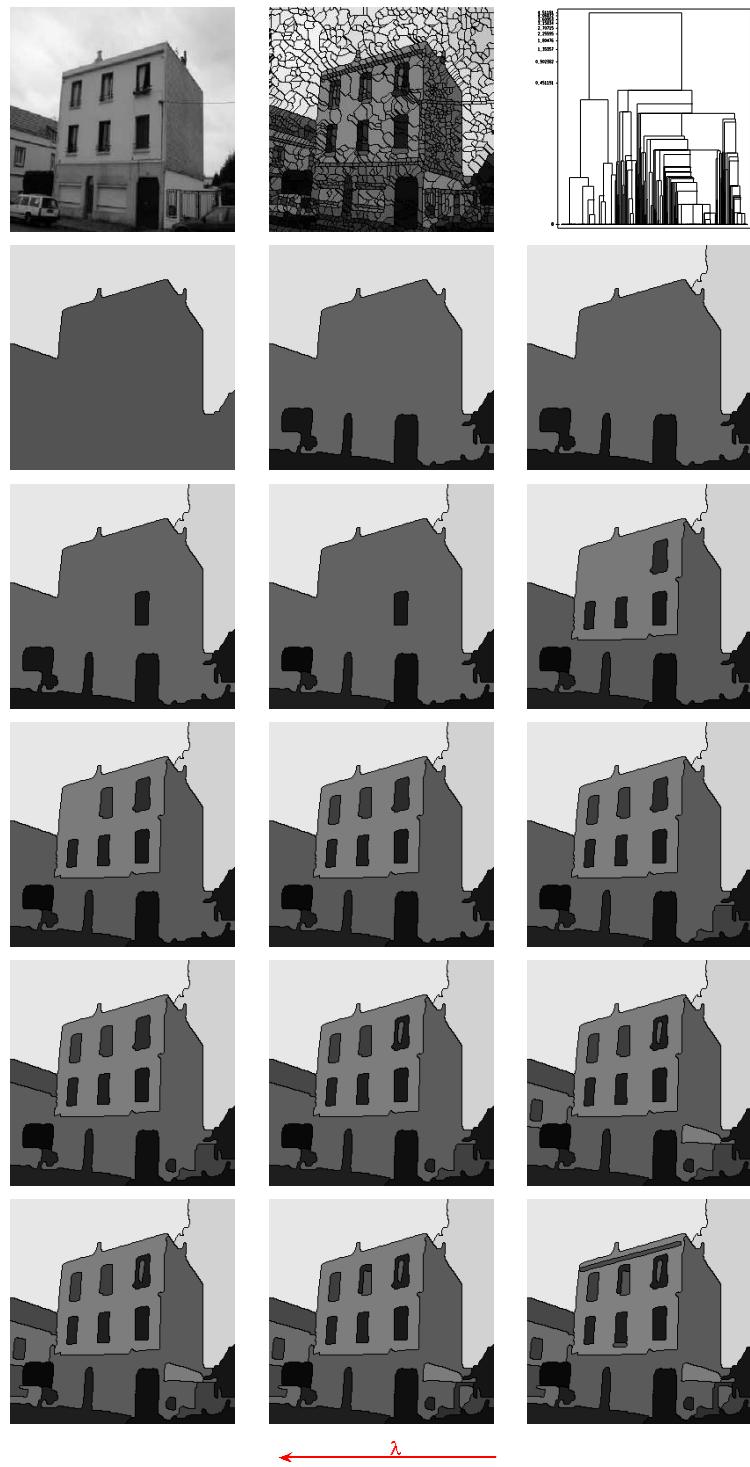


FIGURE 1.5 – Coupes extraites de la hiérarchie construite avec l'énergie de Mumford-Shah. La partition initiale de taille 1559 régions, a été construite avec une LPE.

à la fusion. La raison réside dans le fait que la fusion de cette région avec le fond aurait augmentée la concavité de cette dernière. Néanmoins le choix de l'énergie de concavité a un coût non négligeable en terme de temps d'exécution. A l'inverse de l'énergie de régularisation de Mumford-Shah, l'énergie de concavité ne possède pas une formule de récurrence lui permettant de déduire l'énergie de la fusion à partir des énergies des régions à l'origine de cette fusion.

La figure 1.7 montre les coupes extraites des hiérarchies de régions d'une image aérienne d'un champs de taille  $240 \times 360$ . Ce type d'image convient parfaitement au modèle constant par morceaux de Mumford-Shah car les zones composant la scène sont très proche du modèle par caricatures des régions. Néanmoins, on note que l'introduction d'un terme de concavité à la place du périmètre de la région, permet de mieux segmenter les routes et plus généralement les régions filiformes. En effet, pour une énergie basée sur la longueur de la frontière d'une région, fusionner deux morceaux de routes qui se suivent (selon la largeur) ne se traduit pas par un gain de complexité (simplification) significatif ; ces régions rectangulaires et filiformes seront plutôt fusionnées avec des régions partageant la longueur du rectangle au lieu de sa largeur (ce qui ne permet pas de reconstituer les routes traversant les champs). A l'inverse, fusionner des bouts de routes successifs en utilisant l'énergie de concavité permet de préserver la convexité de la région résultante.

L'originalité de l'approche de Guigues & al réside dans le fait que la hiérarchie est construite par minimisation d'une énergie, avec pour résultat une **analyse en échelle continue**, bas-niveau, de l'image. Elle intègre une information géométrique dans le processus de fusion via le terme de régularisation. Elle introduit également la notion de *persistence* des régions, qui devrait normalement représenter une information pertinente sur la place qu'occupe une région dans une scène. L'intervalle de persistance d'une région, qui traduit sa durée de vie dans la hiérarchie, est une mesure de sa stabilité, dans le sens de Leclerc [1989b]. Elle représente les efforts requis pour la fusionner avec ses régions voisines. Elle quantifie ainsi son degré de prédominance dans la scène.

Une autre propriété remarquable de l'approche ensemble-échelle est l'invariance en échelle par homothétie. Plus précisément, la hiérarchie obtenue reste inchangée si l'on applique une transformation linéaire sur les énergies. L'axe des échelles est dilaté du même facteur que le facteur d'homothétie appliqué aux énergies. Ce point démontre que le réglage *a priori* de l'échelle, dans la définition des énergies, ou *a posteriori*, par coupe dans la description multi-échelle trouvée, est équivalent.

#### 1.4.3.2 Application à l'imagerie satellitaire et aérienne

La représentation ensemble-échelle a été appliquée avec succès dans le domaine des images aériennes et satellitaires. La figure 1.8 illustre le comportement de l'algorithme d'escalade sur une image aérienne d'une ferme. Chaque niveau de coupe exhibe un degré de description différent et met en lumière des structures fines qui n'apparaissaient pas à une échelle plus élevée. Cet exemple illustre parfaitement les problèmes que pose le réglage *a priori* d'une seule échelle de segmentation. Aucun réglage fixe du paramètre d'échelle ne correspond donc à une description complète de la scène. L'algorithme d'escalade parvient à délimiter quasiment toutes les unités pertinentes de l'image (toit, cours, champs) mais à des échelles différentes.

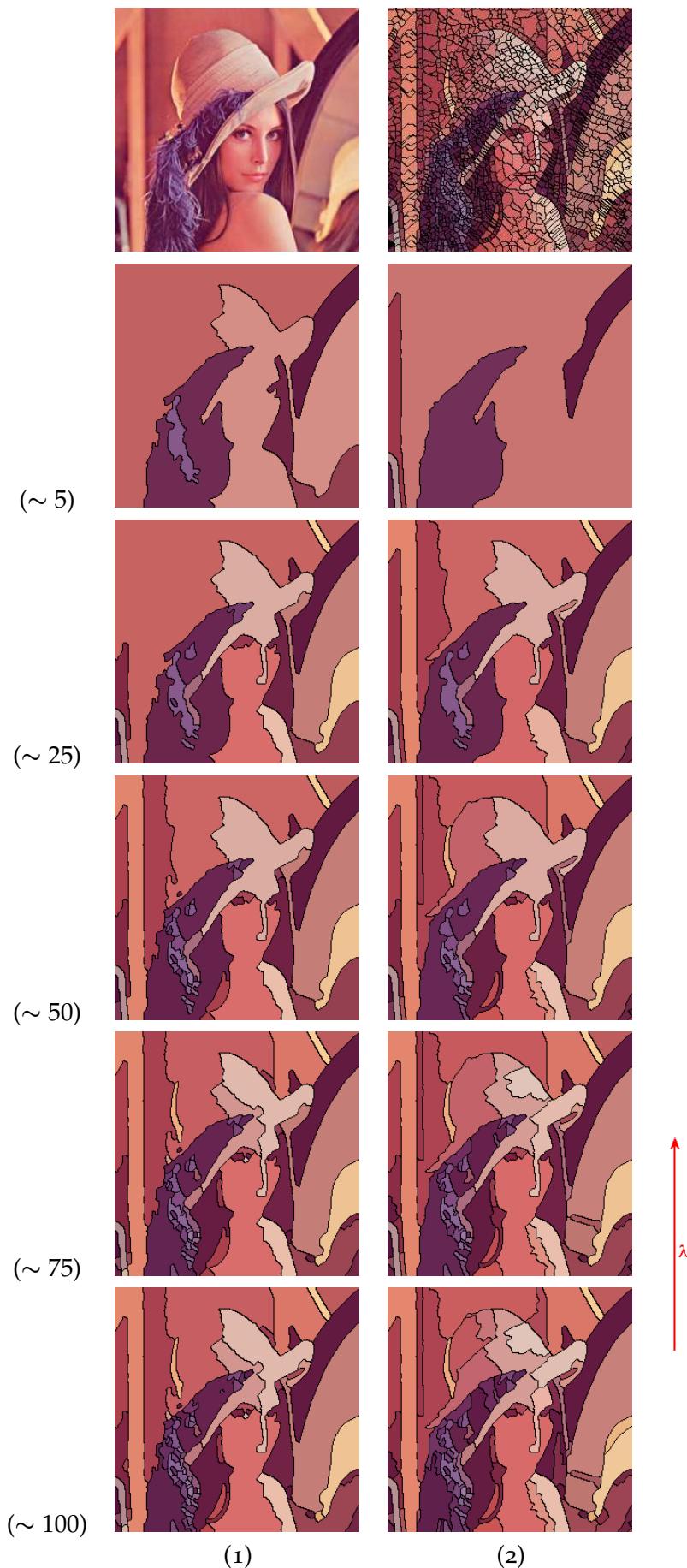


FIGURE 1.6 – Colonne (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : celui obtenu avec l'énergie de concavité.

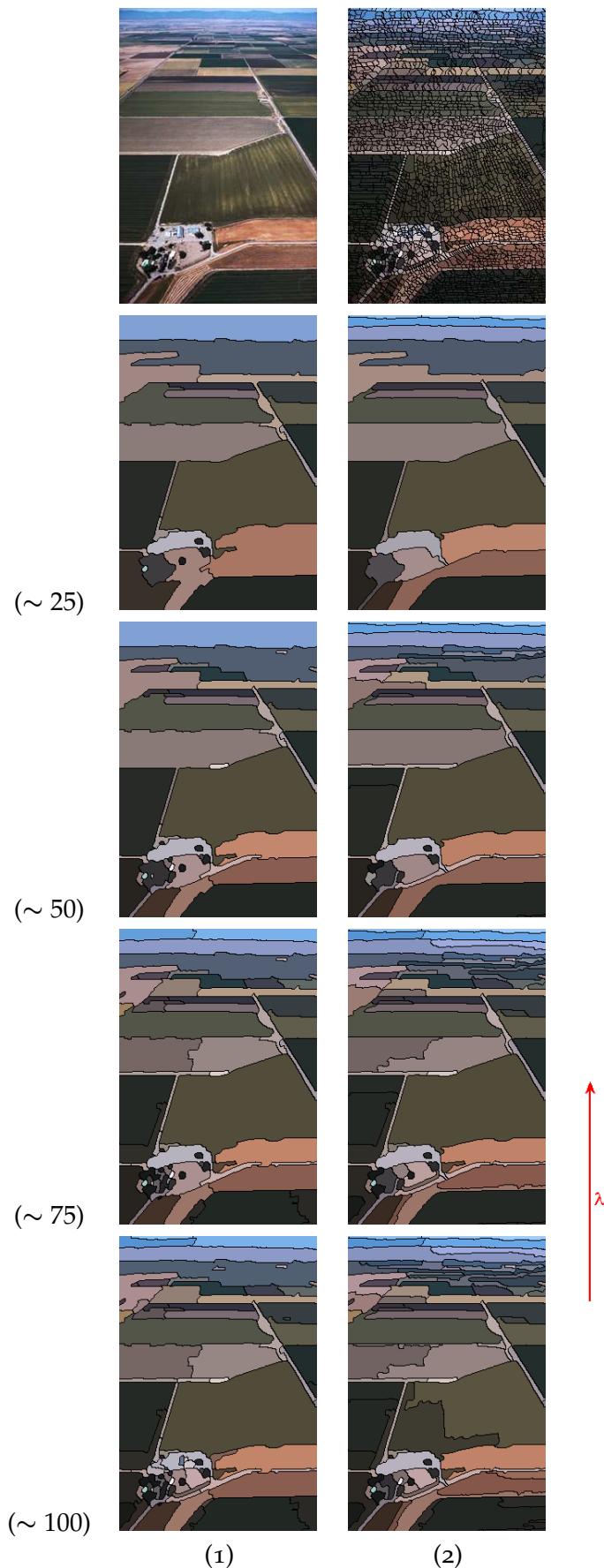
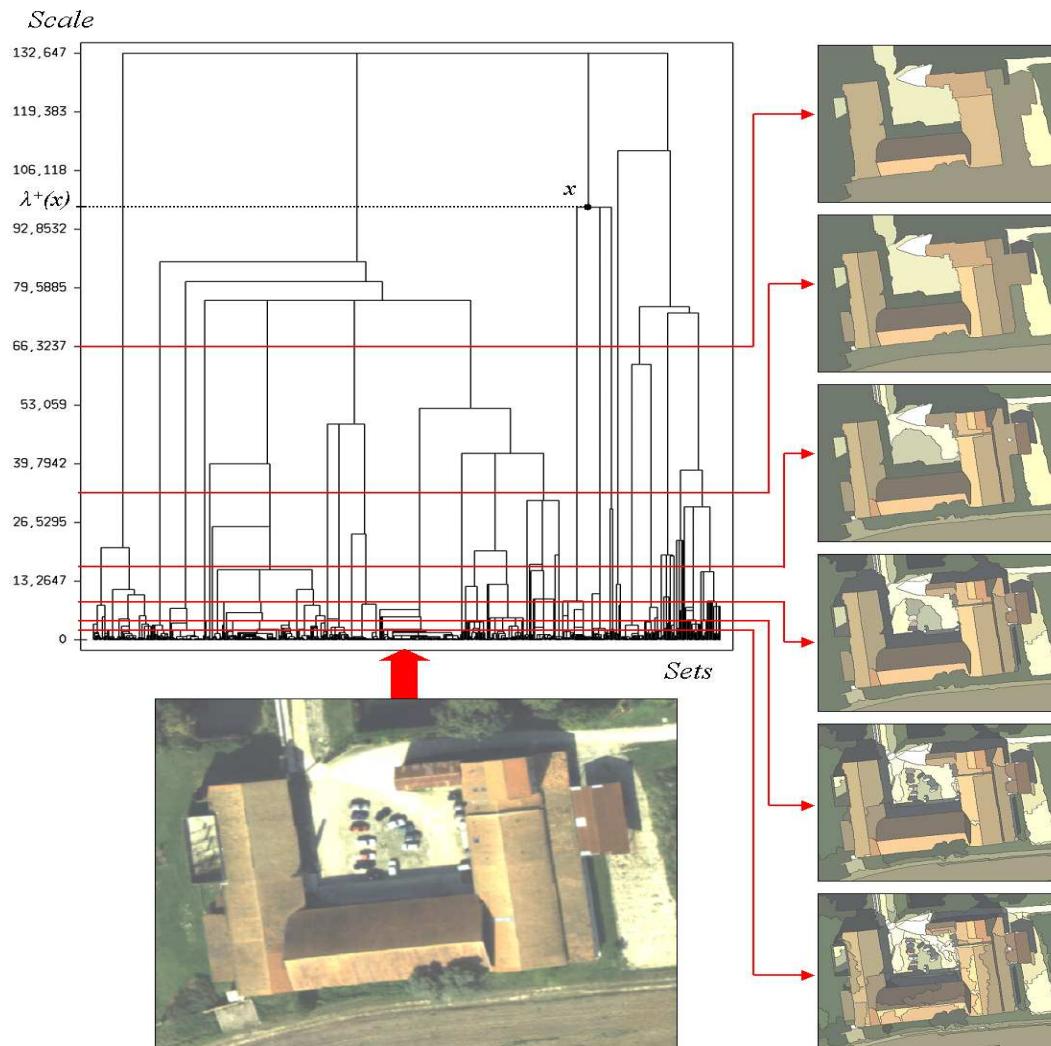


FIGURE 1.7 – Colonne (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : celui obtenu avec l'énergie de concavité.



**FIGURE 1.8 – Une description multi-échelle d'une image aérienne et quelques coupes extraites de sa hiérarchie. Cette représentation a été obtenue par l'algorithme d'escalade binaire, à partir d'une sur-segmentation réalisée avec une ligne de partage des eaux.**

Les résultats des travaux de recherche sur l'approche ensemble-échelle ont donné lieu à plusieurs applications, essentiellement dans le domaine de l'imagerie aérienne et satellitaire. Ainsi, cette méthode de segmentation multi-échelle, est exploitée dans Trias-Sanz et Pierrot-Deseilligny [2004], Trias-Sanz [2004] pour l'interprétation des plans d'occupation du sol en utilisant conjointement des photographies aériennes numériques et des données parcellaires. L'objectif était d'apparier le parcellaire cadastral avec la photographie aérienne. La difficulté principale dans ce type de problème est l'extraction des frontières pertinentes de l'image, celles qui séparent les champs, des bois, des forêts ou des routes. Afin d'éviter le problème que pose les détecteurs de contours comme le détecteur de Canny-Deriche, à savoir le réglage du niveau de détail, les auteurs exploitent la notion de persistance des frontières obtenues par l'algorithme d'escalade pour construire une segmentation de l'image orientée frontière. Le principe est d'affecter, un niveau de gris à une frontière, proportionnel à sa persistance dans la hiérarchie (plus une frontière est haute et plus son niveau de gris est élevé). Le processus

qui consiste à transformer une hiérarchie en une simple représentation-contour, est appelé *Aplatissement Hiérarchique* par Guigues [2003].

Dans Taillandier, Guigues, et Deriche [2003], les auteurs s'intéressent à l' extraction de facettes planes dans un environnement urbain à partir d'images de distance (Modèle numérique de Surface, MNS). Ils utilisent des critères radiométriques intégrés dans le formalisme de L. Guigues pour réaliser la segmentation de l'image à distance. La segmentation est appliquée sur l'ortho-image du gradient, contraintes par des segments 3D qui délimitent les facettes planes. Le terme radiométrique de l'énergie affine est couplé avec un autre terme - calculé sur l'ortho-image du gradient - qui pénalise les régions inhomogène. Citons pour terminer le travail de Chehata, Jung, Pierrot-Deseilligny, et Stamon [2003], qui se sont intéressés à la mise en correspondance d'images satellitaire d'une même scène urbaine. Là aussi, la méthode de segmentation utilisée est la représentation ensemble-échelle. Ils argumentent ce choix par le fait qu'il est impossible d'obtenir des partitions automatiques dont les régions se correspondent bijectivement (Il existe systématiquement des degrés de découpage différents entre les deux images à différents endroits de ces dernières). Ils préconisent ainsi d'utiliser la représentation multi-échelle afin d'avoir plusieurs niveaux de description et de réaliser par la suite des appariements inter-niveaux.

En conclusion, les travaux qui ont suivi l'apparition de l'approche *ensemble-échelle*, se sont focalisés sur l'incorporation de la description multi-échelle de l'image dans une chaîne de traitement afin de résoudre un problème spécifique. Les différents auteurs argumentent leur choix de substituer la représentation d'une image par une hiérarchie en lieu et place d'une simple partition par i) la difficulté du réglage d'une segmentation mono-échelle au niveau des détails recherchés. ii) le phénomène de sur ou sous segmentation des régions d'intérêt dans diverses parties de l'image que produit un algorithme mono-échelle. A contrario, une représentation multi-échelle, offre l'avantage de contenir, différents niveaux de détails. En outre, une fois un couple d'énergies choisi, aucun paramétrage supplémentaire n'est requis.

#### 1.4.3.3 Application à l'imagerie médicale

Comme nous l'avons précisé ci-dessus, l'approche ensemble-échelle produit des résultats très satisfaisant dans le domaine de l'imagerie aérienne. Ce succès est étroitement lié à l'adéquation entre ce type d'image et le modèle choisi.

Nous allons maintenant nous intéresser à des images plus complexes et analyser le comportement de l'algorithme d'escalade binaire dans une situation où le modèle constant par morceaux de Mumford-Shah devient une hypothèse inadaptée à la réalité.

Nous avons appliqué cette approche sur les images médicales du métacarpe, acquises par I.R.M, de taille  $256 \times 256$  et d'une profondeur de 8 bits/pixel. Une image du métacarpe représente une coupe de la main au niveau des cinq doigts. Ce type d'images permet de visualiser les os des doigts de la main afin de rechercher d'éventuels signes de maladies ostéo-articulaires, telles que la synovite ou l'érosion. Ces os sont reconnaissables par leur forme convexe (d'aspect variant entre le carré et le cercle), leur faible niveau de gris ainsi qu'une texture particulière (voir Figure 1.9). Une description idéale d'une image du métacarpe représenterait le pourtour des deux mains d'un patient ainsi que ses régions osseuses. L'énergie utilisée ici est également la forme discrète de la fonctionnelle constante par morceaux

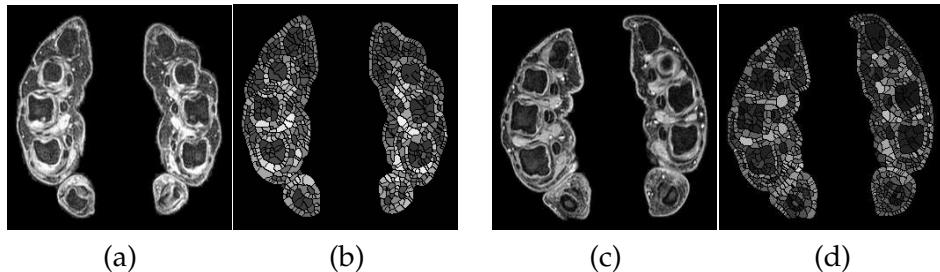


FIGURE 1.9 – Les figures (a) et (c) représentent les coupes du métacarpe de deux patients. Les figures (b) et (d) leur sur-segmentation respectives obtenues avec un algorithme de ligne de partage des eaux. Les régions circulaires et noires représentent les os des doigts de la main.

de Mumford-Shah. La sur-segmentation a été construite par le même algorithme de ligne de partage des eaux, comme le montre les figures 1.9-(b) and 1.9-(d). Observons donc le comportement de l'algorithme d'escalade face à ce type d'images.

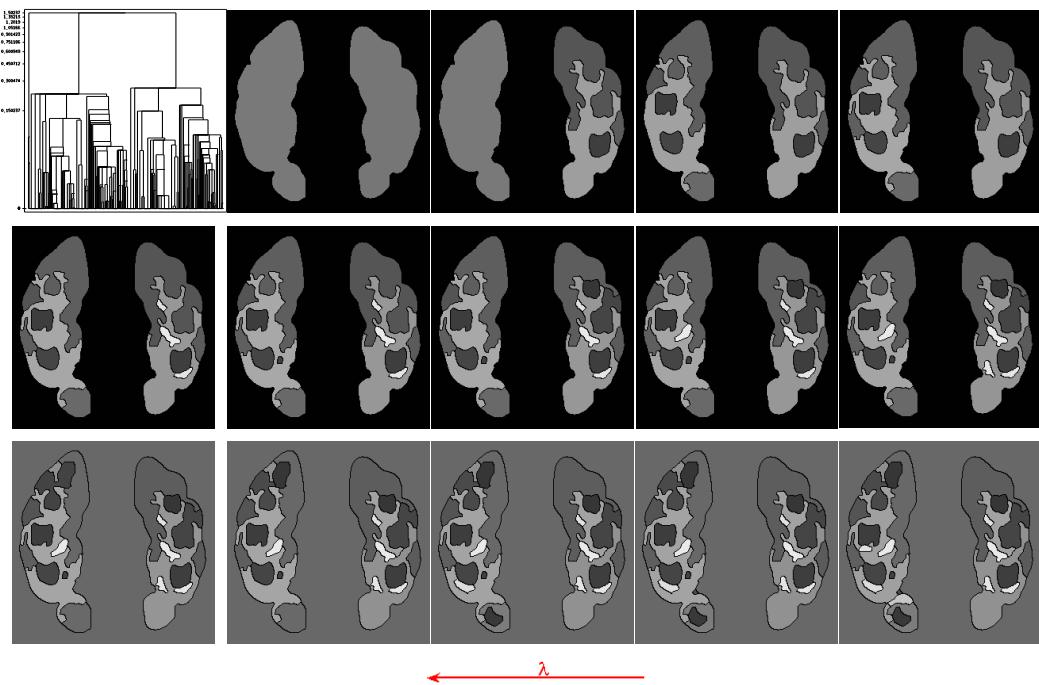
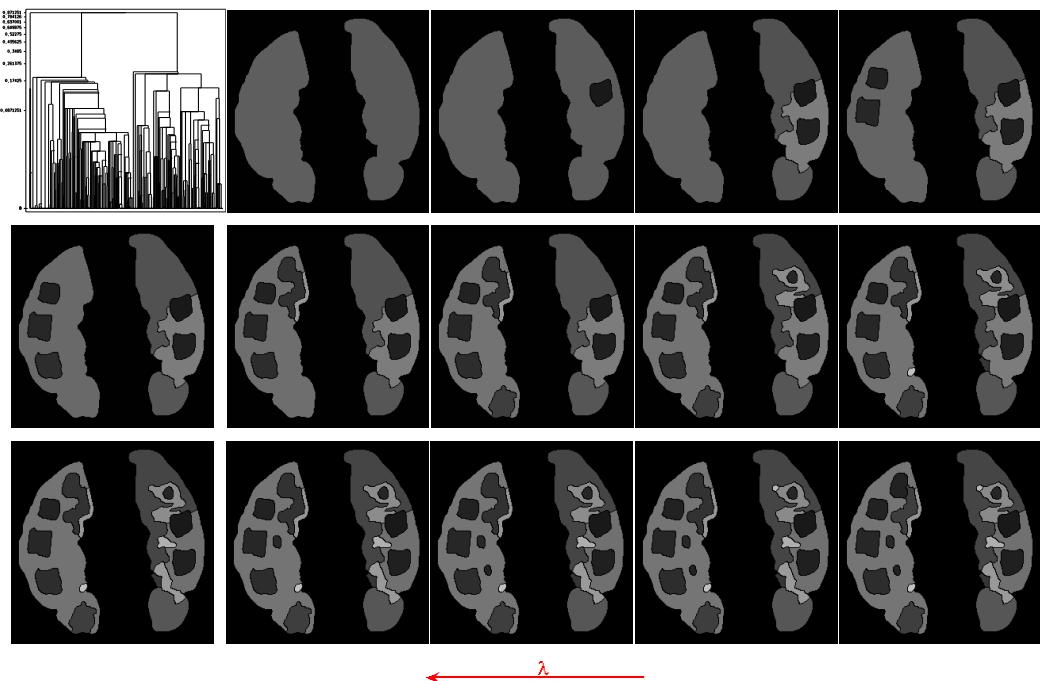


FIGURE 1.10 – Séquence des dernières fusions de la figure 1.9-(a) et sa hiérarchie. On observe que les régions osseuses ont été fusionnées avec les régions non-osseuses bien qu'il subsiste d'autres régions non-osseuses voisines à celles qui ont fusionnées avec les régions osseuses.

La figure 1.10 présente la hiérarchie ainsi que les dernières coupes horizontales de l'image 1.9-(a). Les coupes sont présentées dans l'ordre inverse de celui des fusions des régions, c'est-à-dire de haut en bas. Si l'on analyse la structure de la hiérarchie, on remarquera que le sommet de celle-ci présente trois nœuds fils. Le premier nœud correspond au fond de l'image, les deuxième et troisième nœuds correspondent aux deux mains du patient. La suite de l'exploration de la hiérarchie montre que les deux sous-hiéronymies représentant chacune une main sont asymétriques. Les régions apparaissant dans l'une ne trouvent pas de correspondants dans l'autre pour un niveau de coupe identique. En particulier, les os des mains apparaissent et disparaissent à des échelles distinctes. Quand certains os persistant dans la hiérarchie d'autres ont déjà fusionné avec des régions adjacentes. Quand

certains os sont sur-segmentés d'autres sont sous-segmentés. Des zones appartenant à des classes sémantiques différentes sont fusionnées. On peut voir qu'une région non-osseuse est agrégée avec une région osseuse alors qu'ils subsistaient encore des régions non-osseuses adjacentes à cette dernière. Des fusions aberrantes entre régions appartenant à des classes sémantiques différentes sont inévitables puisque c'est la finalité de l'algorithme d'escalade mais l'ordre des fusions n'est pas compatible avec celui que l'on souhaite. L'idéal aurait été de fusionner des régions de même classe sémantique dans un premier temps avant de terminer la construction hiérarchique.

La décision de fusion s'appuie sur la minimisation d'un compromis entre la radiométrie et la géométrie des paires des régions candidates à la fusion. La décision est par conséquent locale et formellement, la procédure de fusion est guidée uniquement par deux informations (radiométrique et géométrique) locales, dans ce cas précis c'est la somme des écarts quadratiques de la région et le périmètre de cette dernière. Ces deux attributs ne sont que des critères bas-niveau et ne peuvent par conséquent suffire à caractériser la complexité et la variabilité de certains objets de l'image.



**FIGURE 1.11 – Une séquence de partitions monotones représentant les dernières fusions de l'algorithme d'escalade de l'image 1.9-(c). Le processus de fusion échoue à segmenter l'os en haut et à gauche de la main gauche du patient. Ceci est du à la forte similarité radiométrique entre ce dernier et les régions non osseuses qui lui sont adjacentes.**

La figure 1.11 représente le résultat de l'escalade sur l'image 1.9-(c). L'ordre des fusions dans cette hiérarchie est différent par rapport à celui de la hiérarchie de l'image précédente. Topologiquement, les deux structures ne sont pas similaires. D'une part, elles ne comportent pas le même nombre de noeuds et d'arêtes et d'autre part les noeuds se trouvant à un niveau identique<sup>19</sup> ne désignent pas des régions équivalentes. Deux images ayant un contenu sémantique identique mais qui diffèrent dans la taille et le contraste des régions, auront deux hiérarchies diffé-

19. Ici le niveau désigne la hauteur de la hiérarchie vue comme un arbre.

rentes en termes de i) structure topologique des hiérarchies, ii) d'échelles d'apparitions des régions d'intérêt, et iii) de la présence ou non de ces zones d'intérêt. La comparaison structurelle entre deux hiérarchies paraît donc peu réaliste. Des méthodes issues de la théorie de graph-matching (Bomze et al. [1999]) sont inadaptées puisque d'une part, leur complexité combinatoire les rendent inapplicable face à des hiérarchies de taille importante. D'autre part, ces approches sont purement structurelles, elles ne permettent pas de résoudre le problème du décalage entre régions similaires dans des hiérarchies distinctes. Elles ne permettent pas également de résoudre le problème lié à l'absence des régions d'intérêt, à apparier, dans l'une ou l'autre des hiérarchies.

Malgré que certains objets d'intérêt soient correctement segmentés, ils apparaissent et disparaissent à différentes échelles dans la hiérarchie. Cet inconvénient rend difficile l'extraction automatique d'une partition contenant ces objets en se basant uniquement sur leurs échelles d'apparition et de disparition.

Les énergies telles que la fonctionnelle constante par morceaux de Mumford-Shah, suppose que les régions composant la scène, sont relativement homogènes, afin que l'hypothèse du modèle « *constant par morceaux* » soit admissible. Or, ce n'est pas le cas dans des images aussi complexes que celles des images médicales. Le formalisme de la représentation ensemble-échelle modélise un autre aspect essentiel de la vision, dans son processus de fusion. Il s'agit de la dissimilarité radiométrique inter-régions des objets d'intérêt. Cette notion est quantifiée par le numérateur de l'équation temporaire de l'échelle d'apparition d'une région :

$$\lambda^+(x) = -\frac{D(x) - D(P)}{C(x) - C(P)},$$

tel que  $P$  représente la partition optimale de la hiérarchie partielle de la région  $x$ .

Plus la dissimilarité radiométrique est faible et plus la fusion a des chances de se réaliser. Ainsi, un faible contraste, entre deux régions appartenant à des classes sémantique différentes, peut provoquer leur fusion. Alors pourquoi ne pas décrire les zones qui nous intéressent par les attributs appropriés et adaptés à leur variabilité ?

Nos expérimentations sur les images médicales nous ont conduit à constater qu'un certain nombre de fusions, qu'on qualifiera de fusions aberrantes, se produisent entre régions issues de classes sémantiques différentes. Ce phénomène est dû principalement à l'énergie elle-même, qui n'intègre pas des connaissances *a priori* de haut-niveau afin de mieux guider les fusions. Elle ne peut dans ce cas là s'attaquer à la construction de régions de fortes variabilité et complexité en se basant sur le modèle constant par morceaux.

#### 1.4.3.4 Application en vision par ordinateur

De nombreuses applications en vision s'intéressent à la détection d'objets dans des séquences d'images et aux suivis de ces derniers. Ainsi, les détections de piétons, de visage, de panneaux ou de véhicules sont des problèmes récurrents depuis de nombreuses années, mais les difficultés sont nombreuses et font toujours l'objet de nombreux travaux (Cutler et Davis [2000], Viola et al. [2003], Elzein et al. [2003], Navneet et B. [2005]). Par exemple, la détection de piéton doit faire face à la grande variabilité du piéton en échelle, posture et apparence. En outre, un algorithme de détection doit être efficace en temps d'exécution afin que le système de détection puisse s'exécuter en temps réel. La description d'une image est également un point

important et conditionne souvent les résultats de l'algorithme de reconnaissance de forme. Ces images dites urbaines sont très complexes et difficiles à segmenter. Néanmoins, l'application visée ne focalise son attention que sur certaines zones de l'image puisqu'elle doit se prononcer sur la présence ou non des objets d'intérêt et les localiser spatialement dans la scène.

La figure 1.12 illustre une série de coupes extraites de l'ensemble-échelle d'une image contenant des piétons. L'ensemble-échelle a été obtenu en utilisant l'énergie de Mumford-Shah. Un piéton est un objet hétérogène, il se divise en plusieurs régions radiométriquement et géométriquement différentes : cheveux, visage, bras, tronc, jambes, etc. Il paraît donc difficile de segmenter un piéton correctement en s'appuyant sur ce modèle énergétique. Pour découvrir toutes les structures composant un piéton, on est dans l'obligation d'explorer en profondeur la hiérarchie. La partition de 25 régions montre que la partie basse des deux piétons, correspondant à leur pantalon, a été fusionnée avec la chaussée en raison de la similarité radiométrique tandis que le pantalon du piéton se trouvant en arrière plan de l'image et occupant un espace moins prépondérant que les deux piétons précédemment cités est resté persistant à cause d'une plus forte dissimilarité radiométrique avec son environnement. Cela montre que dans certaines situations l'espace occupée par une région est moins important que l'information radiométrique qu'elle porte. Pour extraire toutes les structures composant nos deux piétons, il faut descendre jusqu'à la partition comportant 175 régions. Dans des applications de détection, l'information spatiale est importante et peut être fournie par des détecteurs embarqués. On peut donc légitimement se poser la question de savoir si cet *a priori* spatial peut être intégré dans le mécanisme d'escalade afin d'améliorer la qualité de la segmentation des objets d'intérêt.

La figure 1.13 illustre les coupes extraites de la hiérarchie de l'image d'un trafic routier. On observe que la voiture n'est pas l'un des éléments principaux de la hiérarchie puisque ses structures n'apparaissent qu'à partir de la partition de 80 régions. La forte similarité radiométrique avec la route précipite la fusion des structures de la voiture avec cette dernière tandis que les lettres composant les panneaux demeurent persistantes malgré une occupation dans l'espace moins importante. Là encore, la forte dissimilarité radiométrique est à l'origine de cette persistance. Une connaissance *a priori* spatiale des zones d'intérêt peut là également rendre certaines régions plus pertinentes que d'autres. Evidemment, cette connaissance *a priori* est guidée par l'application.

#### 1.4.3.5 Discussion

Différentes questions se sont posées suite à ces expérimentations. La décision locale de fusion est elle l'unique source qui engendre des agrégations erronées ? ou est ce l'insuffisance d'information sur les caractéristiques des régions à construire quand ces dernières possèdent une variabilité et une complexité non négligeable ? Peut-on se reposer complètement sur des attributs bas niveau pour construire une représentation multi-échelle qui face à des images d'une grande variabilité telles que les images médicales, produit des hiérarchies très hétérogènes ? Peut-on mieux guider le processus de fusion afin de s'assurer que les objets d'intérêt seront présents dans la hiérarchie et que ces mêmes objets soient plus persistants afin qu'ils puissent se retrouver dans une même partition ?

Il est clair que toutes ces questions nous amènent à bifurquer complètement de l'analyse ensemble-échelle telle qu'elle a été conçue au départ, à savoir une

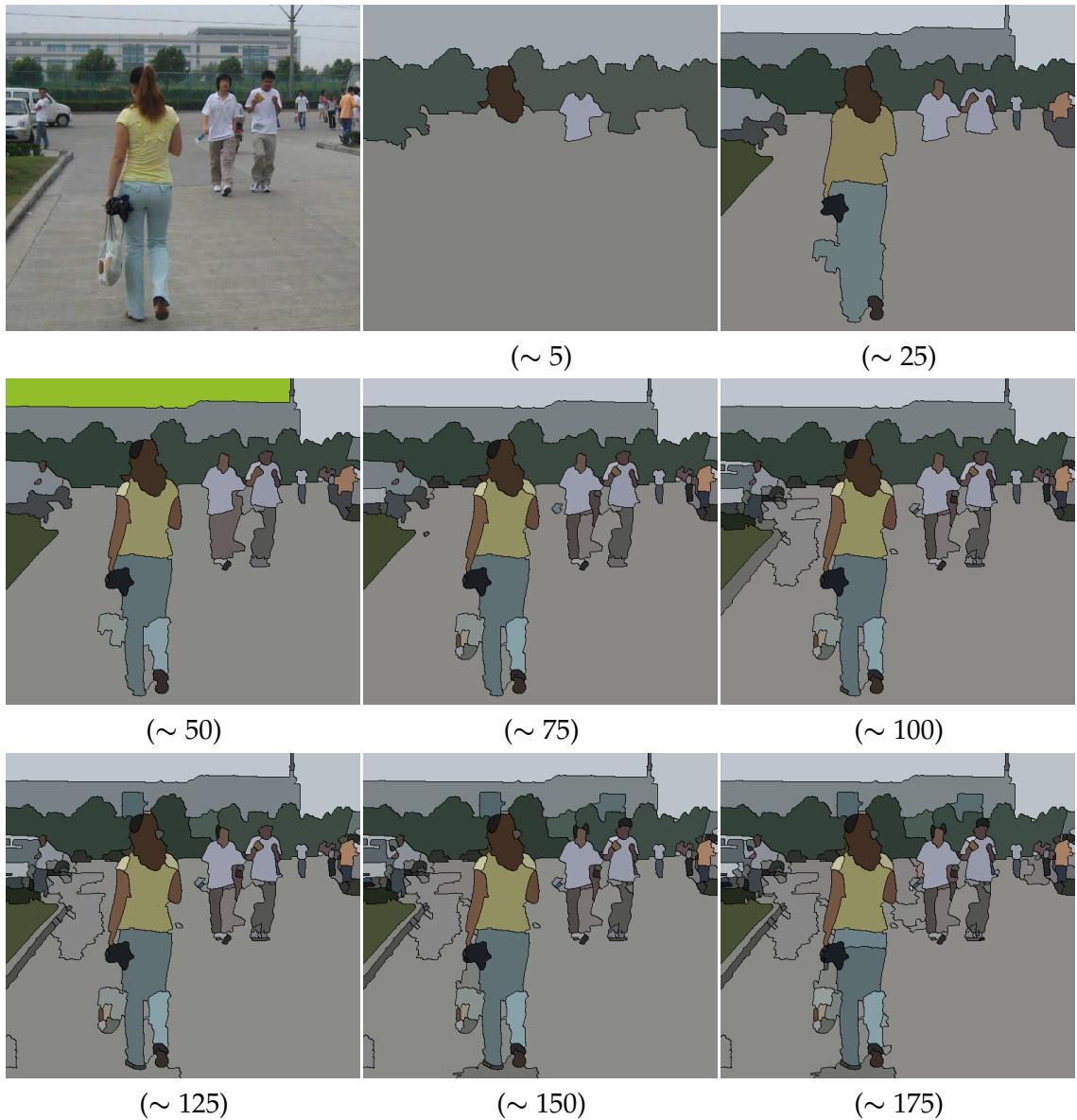


FIGURE 1.12 – Ensemble-échelle obtenu avec l'énergie de Mumford-Shah.

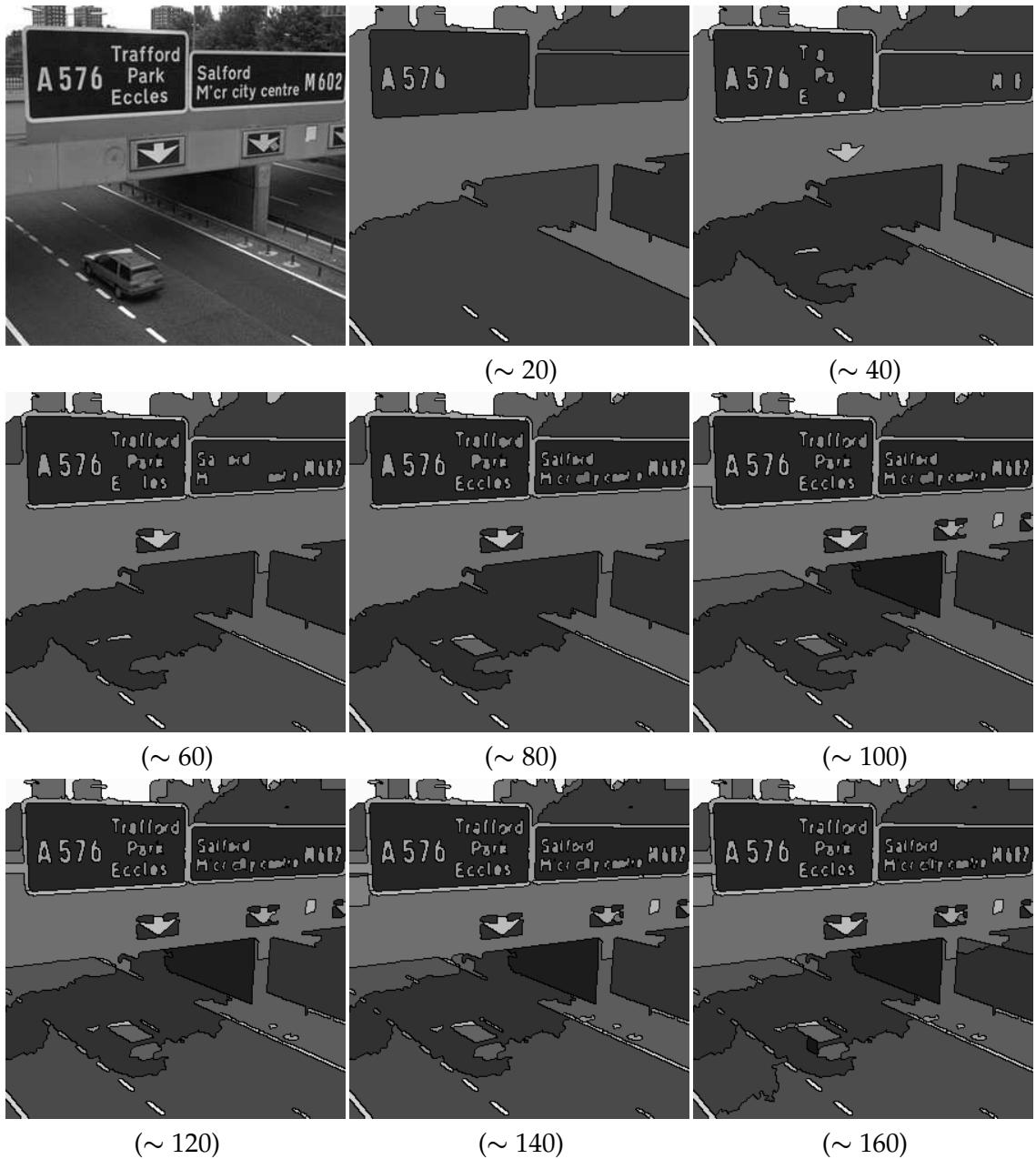


FIGURE 1.13 – Ensemble-échelle obtenu avec l'énergie de Mumford-Shah.

analyse multi-échelle, *bas-niveau*, sans connaissance *a priori* sur le contenu de la scène. L'introduction de connaissances peut se faire par la définition de classes sémantiques si l'on a des *a priori* sur le contenu des images traitées. De même, l'*a priori* peut intervenir dans la caractérisation des points d'intérêt, se manifester sous la forme de blobs ou représenter un modèle de forme d'un objet.

La démarche consistant à injecter des connaissances *a priori* dans le processus de segmentation n'est pas nouvelle puisque bon nombre d'approches de segmentations bas-niveau souffrent des mêmes limitations. Ce problème a amené de nombreux chercheurs à revisiter des méthodes de segmentation bas-niveau afin d'y intégrer des informations haut-niveau. Les méthodes issues de la théorie des ensembles de niveaux ont été adaptées afin de prendre en compte la variabilité et la complexité géométrique des objets que l'on cherche à segmenter. Ainsi, les travaux de Chen et al. [2001; 2002], Rousson et Paragios [2002] et plus récemment de Rousson et Paragios [2007] se sont attachés à incorporer un modèle sur la forme des objets dans la formulation énergétique qui guide la courbe déformable. Ce modèle est construit de différentes manières (Rousson et Paragios [2007]) mais toujours à partir d'un apprentissage de la forme de l'objet sur un ensemble de données.

## CONCLUSION

L'objectif principal de ce chapitre était d'aborder les différentes approches de segmentation multi-échelle et multi-niveau, notamment celles utilisant un processus de fusion. Ce chapitre a apporté une attention toute particulière à l'approche ensemble-échelle car elle représente le point de départ de nos investigations.

L'analyse bas-niveau, orientée-région, de l'image est conçue comme un problème de partitionnement du support de l'image. Elle est séparée de l'application visée. L'approche ensemble-échelle respecte également ce schéma. Elle propose une description de l'image sous-forme d'une hiérarchie de régions. Les auteurs proposent ainsi une analyse en échelle continue de l'image sans aucun réglage préalable. Néanmoins, on a vu que face à des images complexes contenant des zones d'intérêt d'une certaine variabilité, le modèle énergétique ne permet pas de guider le processus de fusion de manière satisfaisante. Face à des applications spécifiques, où la classe d'images et/ou la classe des objets à analyser est contrainte, l'injection d'informations radiométrique et/ou géométrique, liée à l'application apparaît comme une possible solution pourachever cette tâche de segmentation. Par ailleurs, en présence d'applications où certaines zones d'intérêt de l'image sont plus importantes que d'autres, l'information spatiale, si elle est disponible, peut s'avérer pertinente durant le processus de fusion.

Nous allons maintenant présenter dans les chapitres suivants, les différentes contributions apportées par cette thèse. Nous allons détailler les différentes manières d'introduire des connaissances *a priori* dans le processus de fusion.



# CONSTRUCTION SUPERVISÉE D'UNE HIÉRARCHIE DE RÉGIONS

2

## SOMMAIRE

4.1	CONCLUSIONS	165
4.1.1	Construction hiérarchique supervisée	166
4.1.2	Construction d'un ensemble-échelle via un <i>a priori</i> spatial	166
4.2	PERSPECTIVES	167
4.2.1	Injection d' <i>a priori</i> radiométrique et/ou géométrique	167
4.2.2	Injection d' <i>a priori</i> spatial	167
4.2.3	Perspectives générales	168

Ce chapitre présente une première ébauche d'une *construction supervisée d'une hiérarchie de régions* d'une image. L'approche proposée peut être considérée comme une nouvelle méthode appartenant à la famille des techniques de segmentation par fusions de régions. Elle dérive directement de l'approche ensemble-échelle, décrite dans le chapitre précédent. Cependant, elle repose sur la combinaison de trois critères d'agrégation dans le processus de fusion et s'écarte du formalisme énergétique sur lequel se base l'approche ensemble-échelle.

Il s'agit d'injecter des connaissances *a priori* sur le contenu de l'image, dans le processus d'agrégation afin de mieux guider le processus de fusion dans sa construction des régions d'intérêt. Nous verrons que cette approche permet i) d'améliorer la qualité de la décision de fusion, ii) de mieux prendre en compte la complexité des objets à reconstruire et iii) de réduire le nombre de fusions aberrantes entre régions appartenant à des classes sémantiques différentes.

Notre approche décrit l'image sous la forme d'une hiérarchie de régions. Cependant, cette description ne peut plus être qualifiée d'analyse multi-échelle de l'image dans le sens strict du terme<sup>1</sup>. Néanmoins, la hiérarchie construite par

<sup>1</sup>. La définition du terme « analyse ensemble-échelle » proposée par L. Guigues est lié à la définition d'une énergie affine multi-échelle :  $E_\lambda = (C, D)$ . La condition de sous-additivité de l'énergie de régularisation  $C$ , garantit que les coupes optimales de l'énergie affine forment une famille monotone

cette nouvelle approche, comme toute autre hiérarchie obtenue par d'autres méthodes, garde une certaine notion d'échelle. L'exploration en profondeur d'une hiérarchie permet de découvrir des détails de plus en plus fin de l'image au fur et à mesure de l'exploration. Une région se trouvant à un niveau « élevé » de la hiérarchie ne doit sa présence qu'à sa singularité vis à vis de son voisinage exprimée par des attributs (tels que le contraste, l'étendue spatiale, la longueur des frontières, etc.) sur lesquels ont été construits les critères d'agrégation. Par contre, une coupe horizontale, à une échelle donnée, obtenue par l'analyse ensemble-échelle représente la partition optimale de l'image à cette échelle ce qui n'est pas non plus le cas dans une hiérarchie de régions issue d'un formalisme qu'on appellera « agrégatif ». En outre, l'échelle d'une région dans une hiérarchie issue de l'analyse ensemble-échelle représente l'échelle à laquelle elle devient optimale par rapport à ses sous-structures ce qui n'est pas le cas dans une approche agrégative puisque on n'a pas de notion d'optimalité au sens d'une énergie mais plutôt une notion de meilleur coût de fusion binaire entre toutes les fusions binaires possibles.

Néanmoins, on mettra en évidence que le processus de fusion est mieux guidé à travers les connaissances *a priori* injectées. Nous verrons ainsi que la qualité des coupes obtenues est nettement supérieure à celles construites par la méthode ensemble-échelle, et que les structures d'intérêt sont mieux préservées et plus persistantes dans une hiérarchie construite via la supervision.

## 2.1 SÉMANTIQUE ET DÉCISION LOCALE

La **décision locale** intervient dans toutes les approches de fusions de régions quelque soit le formalisme dont elles découlent et ce pour des contraintes de temps de calcul et d'exécution. Elle est tout simplement l'opération qui consiste à fusionner une région avec une région frontalière (i.e, le degré de voisinage est égal à 1) sans se préoccuper des régions qui ont un degré de voisinage plus élevé ou de la fusion avec plusieurs régions frontalières. D'un point de vue algorithmique, les fusions testées se présentent toujours sous forme de paires de régions adjacentes. La décision de fusion dans ces approches repose en général sur la combinaison d'un ou de deux termes, l'un mesurant la (dis)similarité radiométrique et l'autre quantifiant la complexité ou la géométrie de la fusion. Toutes ces approches, utilisant la fusion, construisent une suite de partitions monotone. Elles imposent un critère d'arrêt (implicite ou explicite) afin que la dernière partition soit optimale au sens d'un certain critère ou d'une certaine énergie. Si l'on s'affranchit de ce critère d'arrêt, ces méthodes exhibent toutes une suite de partitions qui peut être présentée sous forme d'une hiérarchie de régions mais ne revêtant pas forcément un sens multi-échelle, on parlera alors plutôt d'une **description multi-niveau** de l'image.

Nous avons souligné dans le précédent chapitre que l'approche « ensemble-échelle » n'avait pas de vision globale sur les régions d'intérêt à construire. Ceci est vrai pour toutes les méthodes de fusion puisque le processus de segmentation est orienté de bas en haut et démarre à partir d'une sur-segmentation de l'image en s'appuyant sur des décisions de fusions locales. Il ne s'agit évidemment pas

---

décroissante (au sens de la relation d'inclusion entre partitions) quand l'échelle  $\lambda$  augmente. Il en résulte une description multi-échelle de l'image au sens où les coupes horizontales de la hiérarchie exhibent des détails de plus en plus fin au fur et à mesure que l'on explore, la hiérarchie obtenue, en profondeur. Notons ici que la finesse et/ou la pertinence d'une région de l'image englobe plusieurs notions à savoir l'étendue spatiale, le contraste et la complexité de la région.

d'un défaut, puisque ces approches visent à réaliser une analyse bas-niveau d'une scène. Par conséquent, elles reposent sur le principe qu'aucune information *a priori* n'est disponible sur les structures composant l'image.

En outre, nous avons avancé l'hypothèse que la description d'une région uniquement par deux attributs bas-niveau (énergies radiométrique et géométrique dans le cas de l'algorithme d'escalade) n'est pas suffisante pour guider correctement le processus de fusion. Spécialement, quand les objets d'intérêt se trouvent à des échelles élevées, là où les régions deviennent de plus en plus complexes. On peut décrire chaque région par plusieurs attributs bas-niveau, le problème est alors de choisir les attributs qui seront pertinents. Ce choix revient à analyser sémantiquement l'image et s'adapter à son contenu afin de sélectionner les caractéristiques qui représenteront le mieux le contenu de l'image. Cette option implique que nous disposons d'informations *a priori* sur le contenu sémantique de l'image. Demeure alors la question de la manière d'incorporer ces connaissances *a priori* dans le processus de fusion. Cela pourra-t-il compenser les carences liées à une prise de décision locale ?

Dans la section suivante nous ferons le lien entre la formulation énergétique de l'approche ensemble-échelle et la formulation par critère d'agrégation et nous mettrons surtout en évidence la différence fondamentale qui existe entre l'approche ensemble-échelle et la formulation par agrégation qui en découle.

## 2.2 D'UN FORMALISME ÉNERGÉTIQUE VERS UN FORMALISME AGRÉGATIF

Nous allons nous placer dans les mêmes conditions que l'approche ensemble-échelle. On se munit donc de deux énergies sur l'ensemble des parties de l'image  $I$ . Soit donc :

$$D : \mathbb{P}(\mathcal{D}) \longrightarrow \mathbb{R},$$

une énergie radiométrique sur-additive et

$$C : \mathbb{P}(\mathcal{D}) \longrightarrow \mathbb{R},$$

une énergie de régularisation sous-additive.

Nous supposerons au départ que ces deux énergies  $D$  et  $C$  sont séparables.

Soit  $P_1$  une partition de l'image et  $P_2$  une autre partition issue de  $P_1$  en fusionnant deux régions quelconques, qu'on notera  $R_k$  et  $R_l$ . Leurs énergies respectives s'écrivent :

$$E_\lambda(P_1) = \lambda \cdot \sum_{R \in P_1} C(R) + \sum_{R \in P_1} D(R),$$

$$E_\lambda(P_2) = \lambda \cdot \sum_{R \in P_2} C(R) + \sum_{R \in P_2} D(R).$$

La différence entre les énergies de ces deux partitions s'écrit :

$$\begin{aligned} \Delta E_\lambda(P_1, P_2) &= E_\lambda(P_2) - E_\lambda(P_1) \\ &= \lambda \cdot [C(R_k \cup R_l) - (C(R_k) + C(R_l))] \\ &\quad + [D(R_k \cup R_l) - (D(R_k) + D(R_l))], \end{aligned} \tag{2.1}$$

Cette équation peut se réécrire d'une autre manière :

$$\Delta E_\lambda(P_1, P_2) = \lambda \cdot \Delta C(R_k, R_l) + \Delta D(R_k, R_l), \quad (2.2)$$

tel que :

$$\Delta C(R_k, R_l) = C(R_k \cup R_l) - [C(R_k) + C(R_l)], \quad (2.3)$$

$$\Delta D(R_k, R_l) = D(R_k \cup R_l) - [D(R_k) + D(R_l)]. \quad (2.4)$$

On notera que  $\Delta C$  représente le coût géométrique de la fusion et  $\Delta D$  le coût radiométrique de la fusion.

Il est aisément de voir que le coût global de l'énergie  $\Delta E_\lambda$  est décroissant en fonction du paramètre  $\lambda$ , à cause de la sous-additivité de l'énergie de régularisation<sup>2</sup>.

Puisque

$$\begin{cases} \Delta E_\lambda \rightarrow +\infty & qd \quad \lambda \rightarrow -\infty \\ \Delta E_\lambda \rightarrow -\infty & qd \quad \lambda \rightarrow +\infty \end{cases} \Rightarrow \exists! \lambda_0, \quad \forall \lambda \geq \lambda_0 \quad \Delta E_\lambda \leq 0.$$

La partition  $P_1$  demeure optimale jusqu'à une certaine échelle, échelle à laquelle la partition  $P_2$  devient à son tour optimale. Cette valeur spécifique provoque alors la fusion des deux régions  $R_k$  et  $R_l$  et ainsi l'apparition de la région  $R_k \cup R_l$  à l'échelle :

$$\lambda^+(R_k \cup R_l) = -\frac{\Delta D(R_k, R_l)}{\Delta C(R_k, R_l)}. \quad (2.5)$$

Quand au signe de l'échelle d'apparition, il sera toujours positif en raison de la sur-additivité de  $D$  et de la sous-additivité de  $C$ .

On peut donc proposer un autre processus de fusion qui teste toutes les fusions binaires et choisit celle qui possède la plus petite échelle d'apparition. Ainsi, à chaque pas de fusion l'énergie de la partition est optimale par rapport à celle dont elle est issue. On peut aussi envisager ce processus comme une stratégie de minimisation d'un critère d'agrégation, se présentant sous la forme d'un ratio entre deux coûts agrégatifs, l'un radiométrique ( $\Delta D$ ) et l'autre géométrique ( $\Delta C$ ). Le comportement de cet algorithme est similaire à celui d'un algorithme de fusion classique. Cependant, un algorithme de fusion classique n'inclue que rarement un critère géométrique. Ici,  $\Delta D(R_k, R_l)$  - qui mesure la dissimilarité radiométrique - est pondéré par un coût géométrique de fusion. Ce dernier encouragera la construction de régions moins complexes.

La différence « fondamentale » avec l'algorithme d'escalade réside principalement dans le processus de minimisation de l'énergie. À travers la relation dynamique fonctionnelle (1.4.1) qui lie l'énergie d'une région de l'image à celles de ses sous-régions, l'approche de L. Guigues a l'avantage de mémoriser l'historique énergétique complet des fusions antérieures composant une région. L'énergie de la fusion de deux régions  $E_\lambda(R_k \cup R_l)$ , qui se présente sous forme d'une droite, sera comparée aux énergies de ses sous-structures optimales qui se présentent elles sous formes de fonctions affines par morceaux pour chacune des deux régions  $R_k$  et  $R_l$ <sup>3</sup>. L'approche agrégative proposée ici, ne s'appuie que sur l'énergie de la précédente partition pour optimiser le choix de la partition suivante. La méthode par agrégation qui dérive de l'approche ensemble-échelle, ne garantit que l'optimalité

2. il suffit de remarquer que la dérivée de  $\Delta E_\lambda$ , i.e  $\Delta C$ , est négative.

3. Schématiquement, on cherche l'intersection entre une droite et la somme de deux lignes brisées. Chaque ligne brisée représente l'énergie d'une région et se compose d'une suite de segments dont les points de liaisons indiquent une fusion de deux sous-structures de la région concernée.

d'une partition par rapport à la précédente. C'est un procédé sans mémoire qui ne tient pas compte de toutes les énergies des partitions antérieures. On s'expose alors à plus de phénomènes d'inversion lors de l'étape de l'élagage.

Si l'on s'affranchit de la modélisation énergétique et l'on considère que le processus de fusion repose sur l'évaluation d'un ratio entre deux critères d'agrégation alors la condition de séparabilité devient inutile. Seule la sous-additivité et la sur-additivité des critères radiométrique et géométrique seront nécessaires.

Examinons maintenant le comportement interne du mécanisme de fusion. La fusion de deux régions est favorisée lorsque le coût radiométrique est minimal, i.e quand  $\Delta D(R_k, R_l) \rightarrow 0$ , ce qui correspond au cas où les régions sont radiométriquement proches (au sens d'un certain attribut de niveau de gris). La fusion est pénalisée quand elles sont radiométriquement distantes. Par ailleurs, quand le coût géométrique des deux régions est élevé, i.e  $|\Delta C(R_k, R_l) \rightarrow +\infty|$ , indiquant une simplification significative du modèle géométrique de la région (au sens d'une certaine caractéristique), la fusion est encouragée. Cette fusion est pénalisée lorsque ce même coût ne varie que très sensiblement, i.e  $|\Delta C(R_k, R_l)| \rightarrow 0^+$ . Cette situation indique que la fusion des deux régions n'engendre pas une régularisation ou une simplification du modèle de la région obtenue<sup>4</sup>. Ces deux cas particuliers et extrêmes correspondent bien à l'interprétation visuelle intuitive que l'on peut se faire de la fusion de deux régions. Cependant, le processus de décision de fusion ne se cantonne pas à ces deux cas et il est en général composé de cas intermédiaires car n'oublions pas que la fusion repose sur la minimisation du ratio entre les deux critères agrégatifs. Les décisions de fusion sont prises sans aucun contrôle sur la signification et la sémantique des fusions réalisées. En outre, sachant que la variabilité en terme géométrique et radiométrique des zones d'intérêt composant une image peut être souvent très élevée, décrire une régions par deux simples attributs radiométrique et géométrique peut ne pas être suffisant pour caractériser et quantifier cette complexité.

Nous proposons donc de modifier l'approche initiale afin d'améliorer la qualité et la fiabilité de la décision de fusion. Nous souhaitons introduire un troisième terme qui reflétera les connaissances *a priori* sur le contenu de l'image. Il intégrera indirectement d'autres caractéristiques de régions afin de mieux décrire la complexité du contenu de l'image. Cependant, cette approche nous contraint à ne plus considérer la formulation énergétique de l'approche ensemble-échelle mais l'approche par critère d'agrégation que nous avons dérivée de cette dernière.

## 2.3 COÛT D'AGRÉGATION CONTEXTUEL

### 2.3.1 Modélisation du contenu de l'image

Précisons dans un premier temps le cadre de travail dans lequel on se placera. On suppose que l'on dispose d'une collection d'images ayant un même contenu sémantique. C'est souvent le cas en imagerie médicale, ou des bases d'images de différentes parties du corps sont stockées : Images I.R.M du cerveau, Images Tomographiques de l'abdomen, etc. On notera par  $\mathcal{F}$  cette famille d'images. Chaque

---

4. Un exemple concret représentant cette situation est celui de deux régions frontalières ne s'intersectant qu'en un seul point. Il en résulte une frontière partagée d'une longueur correspondant à un pixel. Ainsi la fusion des deux régions n'engendre pas un gain significatif en terme de longueur de frontière.

image de la base, notée  $I_s; s = 1,.., N$ , possède un contenu sémantique identique à n'importe quel image de la base.

Qu'entendons nous alors par contenu sémantique d'une image ? Partons du principe que toutes les régions de l'image peuvent être étiquetées par un nom ayant une signification dans le domaine d'application. Le contenu sémantique est l'ensemble de ces étiquettes. Chaque étiquette correspond à une classe dans laquelle les régions de l'image appartenant à cette classe partagent un ensemble de caractéristiques. Par conséquent, une classe permet de représenter la variabilité et la complexité des objets. Pour prendre un exemple, supposons que l'on dispose d'une base d'images tomographiques de l'abdomen. Son contenu sémantique peut être décrit par les classes suivantes : Rein, Rate, Foie, Vésicule biliaire, etc.

Faisons abstraction du domaine d'application et supposons que  $K$  représente le nombre de classes décrivant l'image. Une image  $I_s$  peut contenir plusieurs objets ou un seul objet appartenant à une même classe. Les  $K$  classes peuvent être toutes présentes dans l'image mais une image peut ne contenir qu'une partie de ces classes. En résumé, le nombre de régions d'intérêt par classe et le nombre de classes présentes dans l'image sont variables d'une image à une autre.

Soit  $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_K\}$  l'ensemble des  $K$  classes qui compose une image.  $K$  représente le nombre maximum de classes qui nous intéressent. Supposons aussi que chaque région  $R$  du domaine de l'image soit caractérisée par  $m$  attributs, regroupés sous la forme d'un vecteur  $x(R) = (x^1(R), \dots, x^m(R))$ . Ces attributs peuvent désigner des caractéristiques de texture, des mesures statistiques ou des propriétés géométriques de la région.

Une moyen de déterminer si un objet appartient à une certaine classe est d'utiliser les outils de la Reconnaissance de Formes (RdF), notamment en construisant un classifieur sur l'ensemble des classes d'intérêt, à partir d'une base d'apprentissage.

Dans ce cas, on supposera que le vecteur des attributs d'une région  $x = (x^1, \dots, x^m)$  est une réalisation d'un vecteur aléatoire qu'on notera  $X = (X^1, \dots, X^m)$ .

La classe d'une région est également supposée être une variable aléatoire qu'on désignera par  $G$  et prenant ses valeurs dans  $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_K\}$ .

Ici, nous avons opté pour la construction d'un **classifieur bayésien**<sup>5</sup>, pour la simplicité de sa mise en oeuvre et sa faible complexité.

La théorie de décision bayésienne suppose que chaque couple observé  $(x, g)$  est une réalisation d'un couple de variables aléatoires  $(X, G)$ . Les deux variables aléatoires sont supposées indépendantes.

On décrit alors la loi jointe du couple à travers la définition de :

1. La loi conditionnelle de  $X$  sachant  $G = \mathcal{G}_k$  ( $k = 1,.., K$ ). On notera par  $f_k(x)$  la densité conditionnelle de  $X$  dans la classe  $\mathcal{G}_k$ .
2. La loi marginale, appelée loi *a priori* de la variable  $G$ , définie par :  $P(G = \mathcal{G}_k) = \pi_k$ . Elle représente les proportions respectives des classes dans la population de référence.

La règle de Bayes permet alors de calculer la probabilité conditionnelle de  $G$  sachant  $X = x$ , qu'on appelle loi a posteriori de  $G$  :

---

5. Le classifieur bayésien appartient à la famille des classificateurs linéaires. Il modélise les probabilités conditionnelles. D'autres classificateurs sont possibles tels que la régression logistique, le perceptron, le séparateur à vaste marge, etc.

$$P(G = \mathcal{G}_k | X = x) = P(\mathcal{G}_k | x) = \frac{f_k(x) \cdot \pi_k}{f(x)},$$

tel que  $f(x)$  représente la densité de la variable  $X$  qui peut s'écrire :

$$f(x) = \sum_{k=1}^K \pi_k \cdot f_k(x).$$

Pour que la loi jointe soit complètement définie, il faut spécifier le modèle  $(f_k, \pi_k)$ , pour  $k = 1, \dots, K$ . Le choix du modèle gouvernant les lois conditionnelles du vecteur des attributs sachant la classe peut se faire en optant pour une famille paramétrées de densités conditionnelles,  $f_k(x, \theta), \theta \in \Theta \in \mathbb{R}^p$ . La détermination du modèle se résume alors à l'estimation du vecteur de paramètres  $\theta$ , à partir d'une base d'apprentissage étiquetée. Cette famille paramétrée permet de décrire les différentes formes des classes dans l'espace des attributs.

La règle de décision, pour une nouvelle observation, se base sur la définition d'une fonction de coût qui mesure le risque encouru si l'on prend une mauvaise décision. Une fois la fonction de coût fixée, on doit déterminer la fonction de décision optimale, celle qui minimise le risque moyen. Sans expliciter en détail les calculs, la règle de décision optimale, notée  $\mathbf{D}^*$  dans le cas où la fonction de coût 0-1<sup>6</sup> s'écrit de la manière suivante :

$$\mathbf{D}^*(x) = \underset{\mathcal{G}_k \in \mathcal{G}}{\operatorname{argmax}} P(G = \mathcal{G}_k | X = x), \quad (2.6)$$

tel que  $\mathbf{D}^* : \mathbb{R}^m \mapsto \mathcal{G}$  avec  $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_K\}$  représente l'ensemble des classes.

Dans ce cas précis, la règle de décision préconise de choisir, pour toute nouvelle observation, la classe qui maximise la probabilité *a posteriori* d'appartenir à cette classe sachant l'observation.

La mise en oeuvre du classifieur bayésien nécessite d'une part la construction d'une base d'apprentissage afin d'estimer le(s) paramètre(s) des lois conditionnelles  $f_k(x, \theta)$ . D'autre part, il faudra choisir les proportions des classes composant la population des individus, en d'autres termes les proportions des classes décrivant l'image.

### 2.3.2 Injection d'un *a priori*

Un classifieur bayésien est une règle de décision, il permet de prédire la classe d'une nouvelle observation, dans notre cas il s'agit d'une région, avec une certaine probabilité d'erreur, à partir des probabilités *a posteriori* calculées pour chaque classe.

Cependant, pour le problème nous concernant, à savoir la fusion ou non de deux régions adjacentes, **nous proposons non pas de statuer sur les classes de ces régions, mais plutôt de mesurer le degré d'appartenance de ces deux régions à une même classe**. Ainsi, chaque région  $R$  de l'image se verra assignée son vecteur de probabilité *a posteriori* d'appartenance aux différentes classes de l'image, i.e :

$$P_R = \left\{ P(R \in \mathcal{G}_k | x(R)) \right\}_{k=1, \dots, K},$$

---

6. Cette fonction de coût assigne un coût unité pour une mauvaise décision et nul pour une bonne décision

tel que  $x(R) = (x^1(R), \dots, x^K(R))$  représente le vecteur d'attributs de la région  $R$

L'idée intuitive sur laquelle nous nous basons est exprimée par le postulat suivant :

*Si deux régions adjacentes ont une forte probabilité d'appartenir à deux classes différentes, alors leur fusion doit être pénalisée ou retardée.*

Pour mesurer le degré de dissimilarité entre deux régions en terme d'appartenance ou non à la même classe, nous proposons de mesurer la dissimilarité entre leurs deux vecteurs de probabilité *a posteriori*. Il faudra alors se doter d'une mesure de dissimilarité entre distributions de probabilités :

$$d : \mathbb{R}^K \times \mathbb{R}^K \longrightarrow \mathbb{R}^+$$

tel que

$$d(P_{R_k}, P_{R_l}) = \begin{cases} \sim 0 & \text{si } R_k \text{ et } R_l \text{ appartiennent à la même classe,} \\ >> 0 & \text{si } R_k \text{ et } R_l \text{ sont de classes différentes.} \end{cases}$$

Cette mesure de dissimilarité devra vérifier les propriétés suivantes :

1. Séparation :  $\forall u, v \in \mathbb{R}^K, d(u, v) = 0 \iff u = v,$
2. Symétrie :  $\forall u, v \in \mathbb{R}^K, d(u, v) = d(v, u).$

Nous pouvons à partir de cette mesure de dissimilarité définir le coût d'agrégation contextuel :

**Définition 2.1** Soit  $d$  une mesure de dissimilarité entre distributions de probabilité. On appelle coût d'agrégation contextuel toute fonction réelle positive, notée  $L : \mathcal{P}(\mathcal{D}) \times \mathcal{P}(\mathcal{D}) \longrightarrow \mathbb{R}^+$ , définie à partir d'une mesure de dissimilarité  $d$  :

$$\forall R_k, R_l \in \mathcal{P}(\mathcal{D}) \quad \text{tel que} \quad R_k \cap R_l = \emptyset, \quad L(R_k, R_l) = d(P_{R_k}, P_{R_l})$$

Ainsi pour deux régions adjacentes  $R$  et  $R'$  munies de leur distribution *a posteriori*,  $P_R$  et  $P_{R'}$ , d'appartenance aux différentes classes, le coût contextuel  $L(R, R')$  sera élevé quand les deux régions ont une forte probabilité d'appartenir à des classes différentes. La fusion se trouvera dans ce cas là pénalisée. Inversement, la fusion des deux régions sera encouragée quand le coût  $L(R, R')$  sera faible.

A partir du critère d'agrégation (2.5) cité dans la précédente section, nous proposons un nouveau critère d'agrégation qui prend en compte la mesure de dissimilarité :

$$\lambda^+(R_k, R_l) = \frac{\mathbf{D}(R_k, R_l)}{\mathbf{C}(R_k, R_l)} \times (1 + \alpha \mathbf{L}(R_k, R_l)) \quad (2.7)$$

dans lequel  $\mathbf{D}(R_k, R_l)$  représente le coût radiométrique de la fusion,  $\mathbf{C}(R_k, R_l)$  le coût géométrique de la fusion et  $\mathbf{L}(R_k, R_l) = d(P_{R_k}, P_{R_l})$  le coût contextuel.  $\alpha$  joue le rôle d'un paramètre de pondération qui règle l'importance du coût contextuel.

### 2.3.3 Distances entre distributions de probabilités

Le nouveau critère d'agrégation 2.7 dépend donc de trois fonctions de coûts. Nous avons déjà abordé dans la section 1.3.2.1 les différents coûts radiométriques et géométriques utilisables dans un processus de fusion ( Ils dérouleront donc

soit des énergies d'attache aux données et de complexité définies dans l'approche ensemble-échelle, soit des critères de dissimilarité classiques utilisés dans les algorithmes de fusion de régions).

Il reste donc à définir les différents coûts contextuels envisageables dans notre approche. Comme évoqué, le coût contextuel repose sur la définition d'une mesure de dissimilarité entre distributions de probabilité. Différentes mesures permettent la comparaison entre deux distributions discrètes existent. La plus utilisée est sans doute la divergence de Kullback et Leibler [1951]<sup>7</sup>, dont on rappelle la définition ci-dessous, dans le cas discret :

**Définition 2.2** Pour des distributions  $p$  et  $q$  de variables aléatoires discrètes, à support dans  $[1;K]$ , la divergence Kullback-Leibler de  $q$  à  $p$  est :

$$KL(p, q) = \sum_{i=1}^K p_i \log\left(\frac{p_i}{q_i}\right),$$

Cette mesure présente l'inconvénient de ne pas être symétrique. Plusieurs mesures de dissimilarité symétrique entre distributions dérivant de la divergence de Kullback-Leibler ont été proposées. On peut citer :

- La forme symétrique la plus simple de la divergence de Kullback-Leibler, appelée *symmetric Kullback-Leibler distance* est définie par

$$KL_s(p, q) = \frac{KL(p, q) + KL(q, p)}{2}, \quad (2.8)$$

- La *Resistor Average distance* Johnson et Sinanovic [2001] définie par

$$RAD(p, q) = \frac{KL(p, q).KL(q, p)}{KL(p, q) + KL(q, p)}, \quad (2.9)$$

- La distance dite *distance de Jensen-Shannon* Cover et Thomas [1991]

$$KL_{aver}(p, q) = \sum_{i=1}^K p_i \log\left(\frac{p_i}{\frac{p_i+q_i}{2}}\right) + q_i \log\left(\frac{q_i}{\frac{p_i+q_i}{2}}\right), \quad (2.10)$$

- La *métrique de la moyenne harmonique*

$$d_{MH}(p, q) = \sum_{i=1}^K \frac{2.p_i.q_i}{p_i + q_i}, \quad (2.11)$$

- La divergence appelée *skew divergence* Lillian [2001] mais cette dernière n'est pas symétrique

$$KL_{sk}(p, q) = KL(p; \alpha q + (1 - \alpha)p) \quad (2.12)$$

On pourrait ajouter à ces mesures, d'autres mesures qui ne prennent pas en compte la caractéristique de distribution que représente chaque vecteur  $P_R$ , mais mesure juste la distance entre deux vecteurs de dimension  $K$  :

<sup>7</sup>. La divergence de Kullback-Leibler est une mesure de dissimilarité entre une « vraie » distribution de probabilités  $p$  et une distribution  $q$  arbitraire. Typiquement,  $p$  représente les données, observations, ou une distribution de probabilités calculée avec précision. La distribution  $q$  représente quand à elle, une théorie, un modèle, une description ou une approximation de  $p$ .

- La distance du cosinus

$$\cos(p, q) = \frac{\sum_{i=1}^K p_i \cdot q_i}{\sqrt{\sum_{i=1}^K (p_i^2 + q_i^2)}} \quad (2.13)$$

- La mesure dite *Hellinger distance estimation* ou aussi *Bhattacharyya distance* [Basu, Harris, et Basu [1997]]

$$d_h(p, q) = \sqrt{\sum_{i=1}^K p_i \cdot q_i} \quad (2.14)$$

### 2.3.4 Discussion

En posant  $\alpha = 0$  dans la nouvelle formule d'agrégation 2.7, on revient au cas étudié exposé précédemment (voir formule 2.5). Deux régions seront fusionnées soit parce qu'elles possèdent une forte similarité radiométrique, soit parce qu'elles composent la paire de régions qui simplifie le plus le modèle géométrique de la partition, après fusion. Hormis ces deux cas extrêmes, le processus choisira la paire qui offrira le meilleur compromis entre les deux coûts (celle qui minimise leur ratio).

Le processus de fusion étant initialisé à partir d'une sur-segmentation de l'image, les micro-régions la composant devraient être très homogènes. La similarité radiométrique entre micro-régions adjacentes devrait par conséquent être forte. Il en résulte un coût radiométrique faible favorisant la fusion de ces régions sans porter atteinte à la sémantique de la fusion. Cette phase nous semble correspondre à notre propre vision de la fusion. Quand le coût géométrique est élevé ( $|\Delta C| \rightarrow +\infty$ ), la fusion est également encouragée afin de favoriser l'émergence de régions simples et régulières. Cependant, plus l'échelle augmente et plus la complexité des régions à reconstruire augmente. Le processus de fusion ressemble alors de plus en plus à une compétition numérique où la paire de régions qui offre le meilleur compromis entre les deux coûts est favorisée sans aucun contrôle sur la sémantique de la fusion.

Dans le nouveau critère d'agrégation 2.7, avec  $\alpha \neq 0$ , le rôle du coût contextuel est de retarder la fusion entre régions appartenant à des classes distinctes et ce malgré le fait qu'elle puisse présenter un meilleur compromis radiométrie/complexité. Par conséquent, le coût contextuel devrait implicitement favoriser les fusions entre régions appartenant à la même classe et diminuer ainsi le nombre de fusions aberrantes, i.e celles entre régions issues de classes différentes. Par ailleurs, il devrait également favoriser la construction de régions complexes puisqu'il intègre des informations dérivées d'un ensemble d'attributs adapté à la complexité des classes.

Dans le cas où les régions sont vraisemblablement de la même classe, le coût contextuel n'influencera pas les priorités de fusions puisque :

$$\mathbf{L}(R_k, R_l) \rightarrow 0, \lambda^+(R_k \cup R_l) \rightarrow \frac{\mathbf{D}(R_k, R_l)}{\mathbf{C}(R_k, R_l)},$$

ce qui revient à approcher la formulation par agrégation initiale 2.5.

Pour résumer : “*Cette approche tend à fusionner les régions appartenant à la même classe avant celles appartenant à différentes classes, dernière étape pourachever la construction de la hiérarchie.*”. Ce comportement s'avère plus naturel et souhaitable.

## 2.4 RÉSULTATS EXPÉRIMENTAUX

Nous allons appliquer l'approche proposée dans ce chapitre à la base d'image ostéo-articulaire des métacarpes. Nous disposons d'un échantillon de 219 images 2-D acquises avec la technique d'Imagerie par Résonance Magnétique (I.R.M). Chaque image est de taille 512x512 avec une résolution de 8bits/pixel (i.e l'intensité varie dans l'intervalle [0 ;255]). L'échantillon comporte des coupes du métacarpe issues de différents patients et extraites à différents niveaux du volume. On étudiera dans un premier temps le cas de deux classes, à savoir les « *régions osseuses* » et les « *régions non-osseuses* ». On analysera ensuite les résultats obtenus en ajoutant une troisième classe représentant le fond de l'image.

### 2.4.1 Cas de 2 classes

Nous allons maintenant montrer et analyser les résultats obtenus par cette approche quand le nombre de classes est  $K = 2$ .

#### 2.4.1.1 Pré-traitements des images

Afin de se concentrer uniquement sur les deux classes, les images de la base ont toutes subies une chaîne de pré-traitements pour éliminer le fond de l'image entourant les deux mains du patient. Un simple seuillage (20%) a été appliqué pour permettre l'extraction et l'ébauche d'un premier masque générique. La deuxième étape est une simple opération d'érosion et de dilatation en vue de raffiner le masque obtenu. Un rectangle englobant est ensuite calculé afin de réduire l'influence du fond sur le contenu réel de l'image. La nouvelle image qui résulte de ce masque est ensuite redimensionnée à la taille 256x256 (voir les Figures 2.1, 2.2, 2.3).

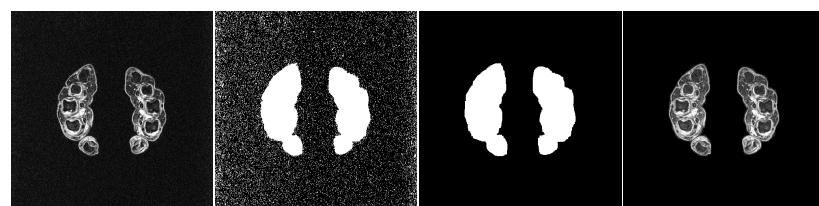


FIGURE 2.1 – Séquences de pré-traitements d'une coupe du métacarpe : seuillage à 20%, érosion puis dilatation du masque.

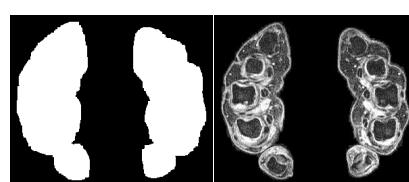


FIGURE 2.2 – Séquence de pré-traitements d'une coupe du métacarpe : calcul et application du rectangle englobant.



FIGURE 2.3 – Image du métacarpe après l'opération de redimension à la taille 256x256.

#### 2.4.1.2 Choix du modèle paramétrique

Nous nous placerons dans un cadre gaussien. On supposera donc que les densités conditionnelles  $f_k(x); k = 1, \dots, K$  suivent toutes une loi normale multidimensionnelle, i.e :

$$f_k(x) \sim \mathcal{N}(\mu_k, \Sigma_k), \forall k = 1, 2,$$

tel que :

$$f_k(x) = \frac{1}{(2\pi)^{m/2} |\Sigma_k|^{1/2}} \exp \left[ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right],$$

avec  $\mu_k \in \mathbb{R}^m$  la moyenne de la classe  $G_k$  et  $\Sigma_k \in \mathcal{M}_{K,K}$  sa matrice de covariance. Dans le cas gaussien, trois situations peuvent se présenter :

1. cas linéaire 1 : la variance est la même pour tout attribut quelque soit la classe considérée,  $\sigma^2$ . les attributs sont indépendants. Par conséquent, la matrice de covariance est une matrice diagonale, identique pour chaque classe :

$$\Sigma_k = \Sigma = \sigma^2 I_m,$$

2. cas linéaire 2 : les attributs sont indépendants de la nature de la classe. Par conséquent, la matrice de covariance sera la même pour chaque classe et s'écrira

$$\Sigma_k = \Sigma,$$

où  $\Sigma$  est la matrice de covariance des attributs quelque soit la classe.

3. cas quadratique : c'est le cas général où chaque classe a sa propre matrice de covariance  $\Sigma_k$ . Aucune hypothèse d'indépendance entre attributs n'est formulée.

Dans nos expérimentations, nous n'avons pris en compte que les deux derniers cas, considérant que la première hypothèse n'était pas réaliste.

On supposera que les probabilités *a priori* des classes sont uniformément distribuées, i.e :

$$\pi_k = 1/K, \quad \forall k = 1, \dots, K.$$

Afin de définir complètement le classifieur, il faut d'une part, préciser les attributs qui vont caractériser chaque individu (région) de la base et d'autre part, estimer les paramètres du modèle, à savoir :  $(\hat{\mu}_k, \hat{\Sigma}_k)_{k=1, \dots, K}$ , à partir d'une base d'apprentissage.

### 2.4.1.3 Phase d'apprentissage

Nous avons constitué un échantillon d'apprentissage comportant 1000 régions étiquetées, 500 par classe. Ces régions ont été extraites à partir d'un échantillon de 50 images de la base initiale. Pour caractériser chaque région de la base, nous avons choisi différents attributs géométriques, statistiques et texturaux. On peut citer pour exemple les 7 moments de Hu [1962], les neufs énergies de texture issues des masques de Laws [1980b], les moments d'inerties, la compacité, etc. En tout, plus d'une quarantaine d'attributs ont été affectés pour la description d'une région. L'annexe A présente les définitions et les propriétés de la plupart de ces attributs.

Cependant, la taille du vecteur des attributs est assez élevée ( $\sim 40$ ) et risque d'alourdir le processus de fusion au cours duquel devront être calculés les attributs des régions issues de chaque nouvelle fusion. En particulier, s'il n'existe pas de relation de récurrence entre les attributs de la nouvelle région et ceux issus des régions la composant. Par conséquent, afin de sélectionner les attributs qui discriminent le mieux les deux classes et afin d'améliorer le taux d'erreur de classification, une procédure de « *Validation Croisée* » (Geisser [1975]), et une procédure de sélection d'attributs « *Forward Selection* » ont été intégrées dans le processus d'apprentissage. La technique de sélection des attributs par *Forward Selection* est un processus itératif qui vise à sélectionner les meilleurs attributs qui minimisent le taux d'erreur. Le processus commence par tester chaque attribut individuellement puis il sélectionne l'attribut qui produit le plus faible taux d'erreur. L'attribut sélectionné est retiré de la liste initiale des attributs disponibles et une nouvelle phase de sélection est appliquée sur cette liste en combinant l'attribut choisi avec chaque attribut de la liste. Le procédé est itéré jusqu'à ce que le taux d'erreur cesse de décroître<sup>8</sup><sup>9</sup>. L'étape de validation croisée est aussi un processus itératif qui consiste à diviser l'échantillon d'apprentissage en  $M$  parties, estimer les paramètres du classifieur sur les  $M - 1$  parties (ensemble d'apprentissage), et utiliser le dernier échantillon pour estimer l'erreur (ensemble de test). Il s'agit alors d'itérer ce processus en permutant à chaque fois l'échantillon destiné à évaluer l'erreur. L'erreur moyenne issue de toutes les permutations représentera alors le taux d'erreur d'une configuration d'attributs.

### 2.4.1.4 Sélection des fonctions de coûts

Afin de permettre une comparaison objective entre notre approche et l'approche ensemble-échelle, les coûts radiométrique et géométrique ont été dérivés des énergies radiométrique et géométrique utilisées dans l'algorithme d'escalade1.4.2. L'énergie affine utilisée dans nos expérimentations est issue de la fonctionnelle constante par morceaux de Mumford-Shah :

$$\forall R \in \mathcal{P}(\mathcal{D}), \quad C(R) = |\delta R| \quad \text{et} \quad D(R) = \sum_{x \in R} ||I(x) - \bar{I}_R||^2.$$

où  $C$  représente le périmètre de la région  $R$  et  $D$  la somme des écarts quadratiques de la valeur moyenne de la région,  $\bar{I}_R$ .

8. du moins significativement, puisque ce taux d'erreur peut continuer à décroître très lentement et ainsi continuer à sélectionner de nouveaux attributs et agrandir la taille du vecteur d'attributs.

9. Cette approche permet de minimiser le taux d'erreur localement.

Ainsi, le critère d'agrégation radiométrique s'écrira :

$$\mathbf{D}(R_k, R_l) = \Delta D(R_k, R_l) = \frac{|R_k| \cdot |R_l|}{|R_k| + |R_l|} (\bar{I}_{R_k} - \bar{I}_{R_l})^2,$$

noter que ce critère est équivalent à la formule de Ward [1963], dont la propriété de récurrence permettra d'accélérer la mise à jour du coût radiométrique de la nouvelle région à partir du coût des régions fusionnées. Le critère d'agrégation géométrique s'écrira :

$$\mathbf{C}(R_k, R_l) = -\Delta C(R_k, R_l) = -2 \cdot |\delta(R_k \cap R_l)|,$$

où  $\delta$  représente la longueur de la frontière.

Pour terminer, on utilisera la distance de Kullback-Leibler symétrisée afin de construire le coût contextuel :

$$\mathbf{L}(R_k, R_l) = KL_s(P_{R_k}, P_{R_l})$$

où  $P_{R_k}$  et  $P_{R_l}$  représentent les distributions *a posteriori* d'appartenance aux différentes classes.

#### 2.4.1.5 Rehaussement du contenu sémantique de la hiérarchie

Nous avons opté pour le modèle quadratique où la matrice de covariance est spécifique à chaque classe car il ne pose aucune hypothèse d'indépendance entre attributs.

L'étape d'apprentissage a permis de dégager une famille d'attributs composées des énergies de texture issues des masques de Law ainsi qu'un attribut mesurant la compacité de la région. La probabilité d'erreur estimée à l'issue de cette phase était de l'ordre de 14%.

Les attributs sélectionnés vérifient tous une relation de récurrence qui permet d'accélérer leurs mises à jour et par conséquent le processus de fusion.

L'algorithme de segmentation se comporte comme un algorithme classique de fusion par minimisation d'un critère d'agrégation. Cependant, pour chaque paire de régions adjacentes un superviseur externe est utilisé pour calculer le coût contextuel de la fusion. Ce coût contextuel exprime dans une certaine mesure la cohérence de la fusion entre ces régions. L'algorithme est initialisé avec une partition obtenue par l'algorithme de lignes de partage des eaux sur le module du gradient de l'image (voir l'annexe B).

La figure 2.4 présente les résultats obtenus sur l'image de la figure 1.9-(a). Il s'agit des dernières coupes du métacarpe, extraites des hiérarchies obtenues en utilisant :

- le critère d'agrégation de Ward illustrées dans la colonne (a),
- le critère d'agrégation de Mean-Ward illustrées dans la colonne (b),
- le critère d'agrégation 2.5 représentées dans la colonne (c),
- l'algorithme d'escalade à partir de la fonctionnelle de Mumford-Shah dans figurant la colonne (d),
- et le nouveau critère d'agrégation 2.7 pour  $\alpha = 1$  représentées dans la colonne (e).

Regardons dans un premier temps le rôle fondamental que joue l'information géométrique dans la construction d'une hiérarchie de régions. Les partitions des

colonnes (a) et (b) ont été obtenues en utilisant un critère d'agrégation radiométrique. En l'absence d'un terme géométrique, les frontières des régions composant ces partitions sont plus sinueuses. Les coupes des lignes (8) et (9) de ces deux colonnes illustrent clairement cet inconvénient. Le processus de fusion préfère agréger des régions de radiométrie similaire au détriment de la régularité des contours des régions. La région représentant la main droite n'est pas reconstituée en tant qu'objet global dominant le reste de ses sous-structures et le principe hiérarchique des objets d'intérêt contenus dans l'image est transgressé. La colonne (c) présente les coupes extraites d'une hiérarchie construite avec un critère d'agrégation composé du critère de Ward pondéré par le critère géométrique portant sur la longueur de la frontière (2.5). On constate que le compromis similarité radiométrique/régularité géométrique permet de gommer les défauts cités ci-dessus. Il en est de même des coupes de la colonnes (d) obtenues avec la méthode d'escalade binaire en s'appuyant sur la fonctionnelle de Mumford-Shah.

On constate néanmoins que les fusions sémantiques, entre régions osseuses et régions non-osseuses se produisent dans toutes ces colonnes ((a)-(d)) et ce malgré le fait qu'il subsiste des régions non-osseuses adjacentes. La colonne 2.4-(e) illustre parfaitement le rôle positif que joue le coût contextuel dans le processus de fusion, à savoir mieux guider les décisions de fusions. La comparaison visuelle entre cette séquence et la séquence obtenue par la formulation énergétique (colonne (d)), permet de constater une réelle amélioration dans le contenu sémantique des partitions. Il semble que le troisième terme permet d'éviter, dans la majorité des cas, les fusions précoces entre régions issues de classes différentes. Si l'on se focalise sur la main gauche du patient, on remarquera que là où les régions osseuses sont successivement absorbées par les régions non osseuses dans les coupes de la colonnes (d), l'inverse se produit dans les coupes de la colonne (e). Le coût contextuel pénalise ce type de fusion inter-classes et encourage par conséquent les fusions entre régions de même classe. Cependant, on relèvera, dans cette même colonne (e), que l'os du milieu de la main droite a été fusionnée avec une région non-osseuse (2.4-(3)-(e)). La même remarque est à noter au sujet de l'os se trouvant en haut dans la main droite. Il n'apparaît pas dans cette séquence puisque ce dernier a été fusionné précédemment avec une région non-osseuse. Il se pose alors le problème de la pondération du paramètre  $\alpha$  entre le critère d'agrégation 2.5 et le coût contextuel.

La figure 2.5 illustre les résultats, obtenus sur le métacarpe de l'image 1.9-(c). La colonne 2.5-(e) montre clairement l'amélioration qualitative des partitions obtenues par rapport à celles issues de l'approche ensemble-échelle exposées dans la colonne 2.5-(c). En observant uniquement la première ligne de cette figure, on s'aperçoit réellement de l'impact du coût contextuel sur les décisions de fusion. La figure 2.5-(1)-(e) répond parfaitement à nos attentes en terme de description simple et concise des objets d'intérêt présents dans l'image sans sur-segmentation ou sous segmentation de ces objets. Les hiérarchies obtenues par les autres méthodes ne contiennent pas une coupe similaire à celle exhibée dans la figure 2.5-(1)-(e).

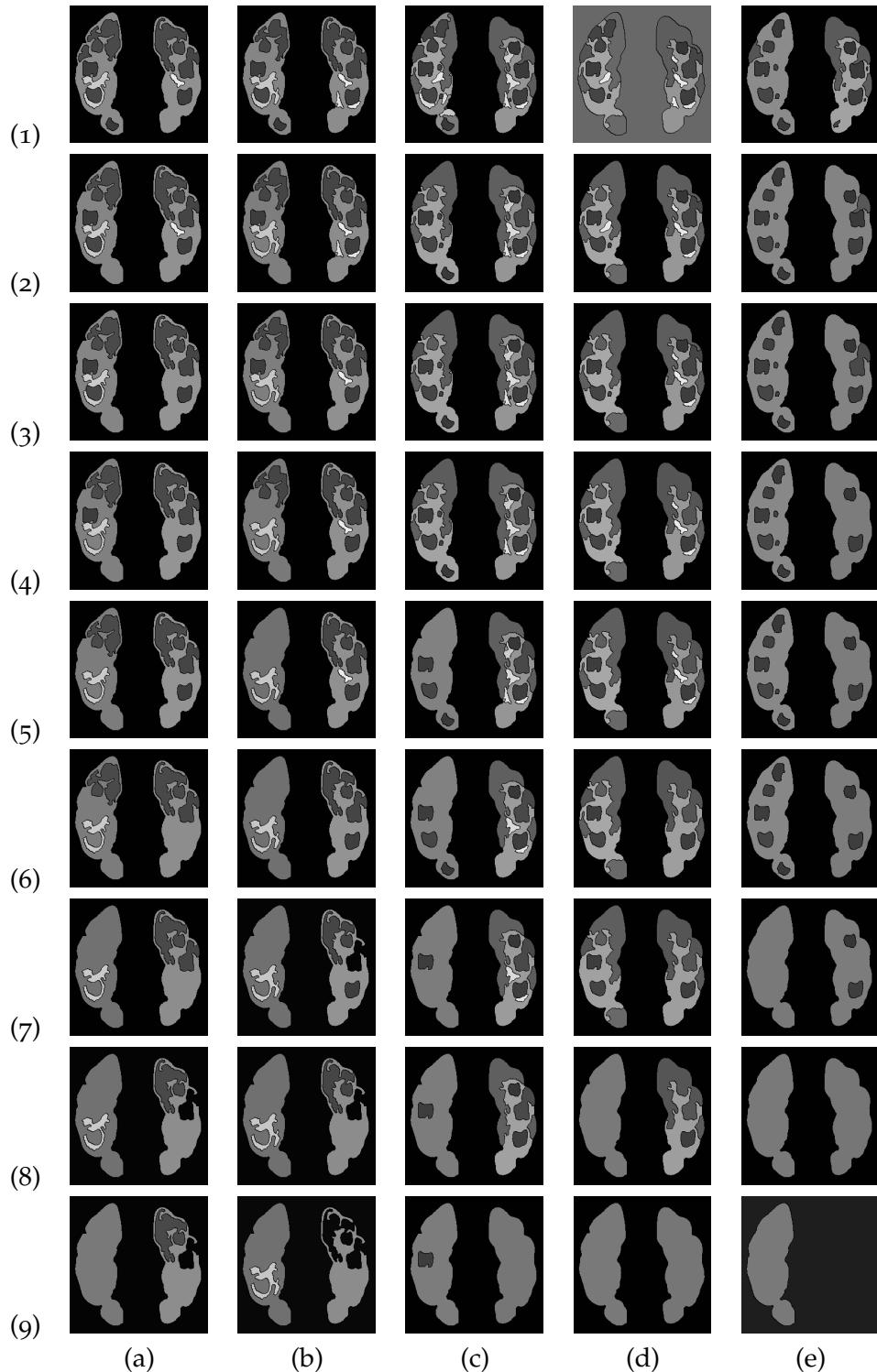


FIGURE 2.4 – La colonne (a) présente la séquence des dernières fusions de la figure 1.9-(a) avec le critère de Ward, la colonne (b) celle obtenue avec le coût Mean-Ward, la colonne (c) celle construite avec le critère d'agrégation dérivé de la formulation énergétique, la colonne (d) celle obtenue avec l'approche ensemble-échelle et la colonne (e) représente la séquence construite par le nouveau critère d'agrégation qui intègre des connaissances a priori pour  $\alpha = 1$ .

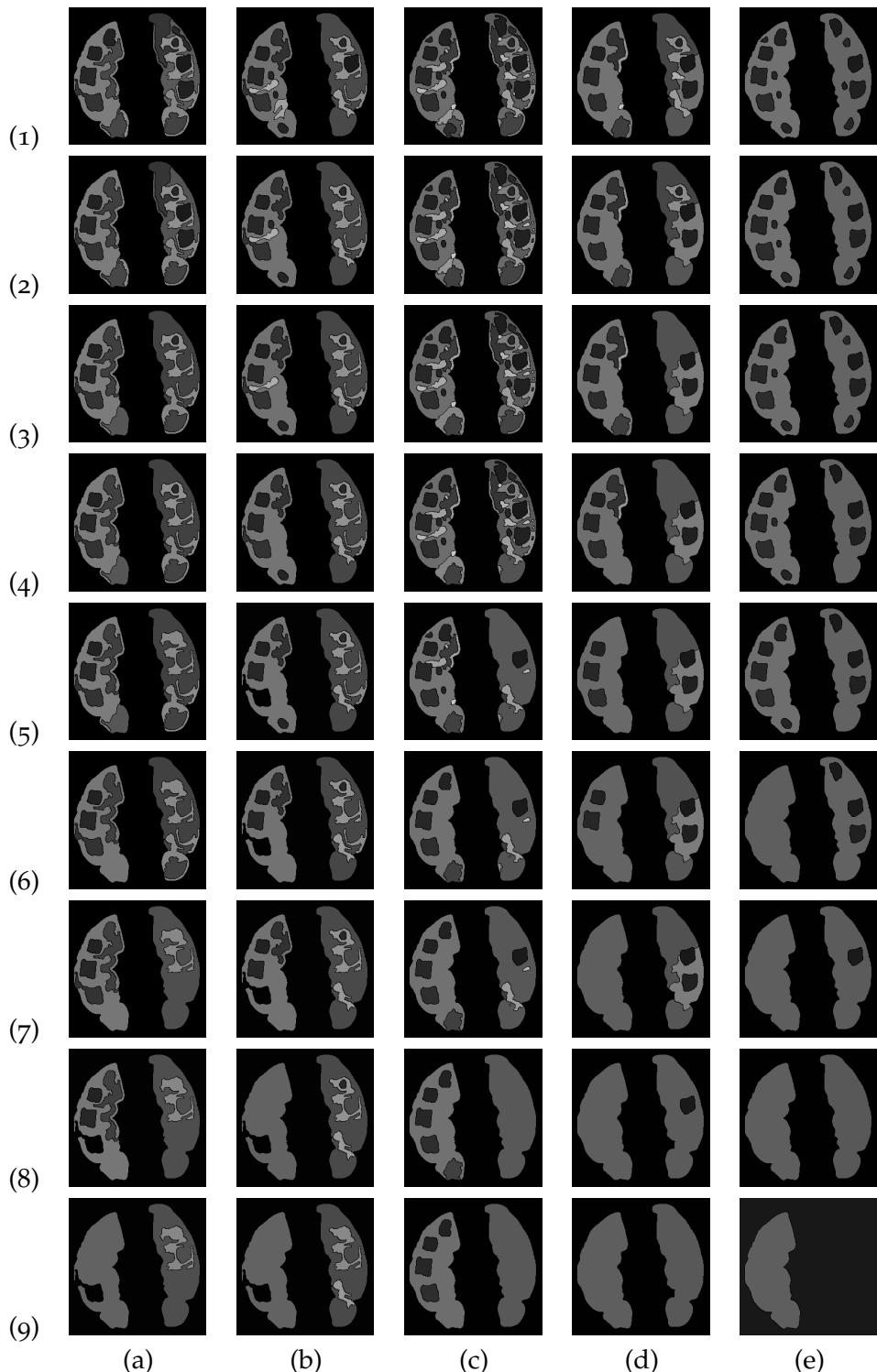


FIGURE 2.5 – La colonne (a) présente la séquence des dernières fusions de la figure 1.9-(c) avec le critère de Ward, la colonne (b) celle obtenue avec le coût Mean-Ward, la colonne (c) celle construite avec le critère d'agrégation dérivé de la formulation énergétique, la colonne (d) celle obtenue avec l'approche ensemble-échelle et la colonne (e) représente la séquence construite par le nouveau critère d'agrégation qui intègre des connaissances a priori.

Nous allons maintenant analyser les résultats obtenus sur les deux précédentes images du métacarpe quand le paramètre de pondération est porté à 3 ( $\alpha = 3$ ). L'objectif est de donner une plus grande influence au coût contextuel et de pénaliser ainsi fortement les fusions inter-classes.

La figure 2.6 montre les séquences de partitions obtenues sur l'image 1.9-(a). La colonne (a) représente les coupes obtenues avec l'algorithme d'escalade binaire, la colonne (b) représente les coupes extraites de la hiérarchie construite avec le critère d'agrégation 2.7 pour  $\alpha = 1$  et la colonne (c) celles obtenues en utilisant ce même critère pour  $\alpha = 3$ . Une comparaison visuelle entre les colonnes (b) et (c) montre clairement l'amélioration significative des décisions de fusion lorsque  $\alpha = 3$ . Les coupes (1) à (5) de la colonne (c) exhibent toutes des partitions qui séparent clairement les régions osseuses des régions non-osseuses. En augmentant l'influence du coût contextuel dans la formule d'agrégation 2.7, le processus de fusion est supervisé de manière plus efficace. On notera également que les fusions inter-classes qui ont persisté dans le cas  $\alpha = 1$  sont corrigées pour  $\alpha = 3$ . Les os du milieu et du haut de la main droite ne sont plus fusionnés avec des régions non-osseuses. Par conséquent, leur persistance dans la hiérarchie est rehaussée. Ainsi, en évitant les fusions entre régions inter-classes, la persistance des os dans la hiérarchie s'en trouve augmentée. Ce type de comportement nous semble naturel et approprié pour toute région d'intérêt. On notera, tout de même, que l'os du bas dans la main gauche n'apparaît pas dans toutes les séquences exposées dans les précédentes figures. Ceci est du simplement au fait qu'il n'y a pas d'os à cet endroit mais une zone correspondant à l'articulation du doigt. Par ailleurs, le classifieur estime une probabilité d'appartenance allant dans ce sens.

#### 2.4.1.6 Persistance des régions

Afin de quantifier l'amélioration du contenu sémantique des hiérarchies construites avec cette approche et analyser le comportement global du nouveau critère d'agrégation durant le processus de fusion, nous avons appliqué cette méthode, qu'on notera « CHS » (pour Construction Supervisée Hiérarchique par critère d'agrégation), et l'approche ensemble-échelle, qu'on notera « ASC » sur une base d'images de 30 métacarpes dont un échantillon est représenté dans la figure 2.8. Chaque image représente un coupe du métacarpe d'une seule main. On dispose donc de 150 os potentiels.

Dans un cadre idéal, il est souhaitable que les régions d'intérêt soient plus persistantes dans la hiérarchie par rapport à des structures moins pertinentes. Afin de vérifier si notre approche a une influence quelconque sur la persistance des régions osseuses par rapport à la méthode ensemble-échelle, nous avons extrait manuellement les régions osseuses dans les hiérarchies de la base ainsi que leurs échelles d'apparition et de disparition respectives.

La figure 2.9-(a) représente les boîtes à moustache de la persistance des os obtenue par notre approche (figure de gauche) et par l'approche ensemble-échelle (figure de droite). On voit clairement que la persistance des os est plus élevée que celle obtenue par l'algorithme d'escalade. Les figures 2.9-(b) et 2.9-(c) illustrent les apparitions et les disparitions des os représentées sous forme de boîtes à moustache. On remarque que les échelles d'apparitions et de disparitions des os sont globalement ultérieures à celles obtenues avec l'approche ensemble-échelle. Par conséquent, les régions osseuses auront une forte probabilité d'être présentes dans les hautes échelles des hiérarchies ce qui est plus naturel et cohérent. On a

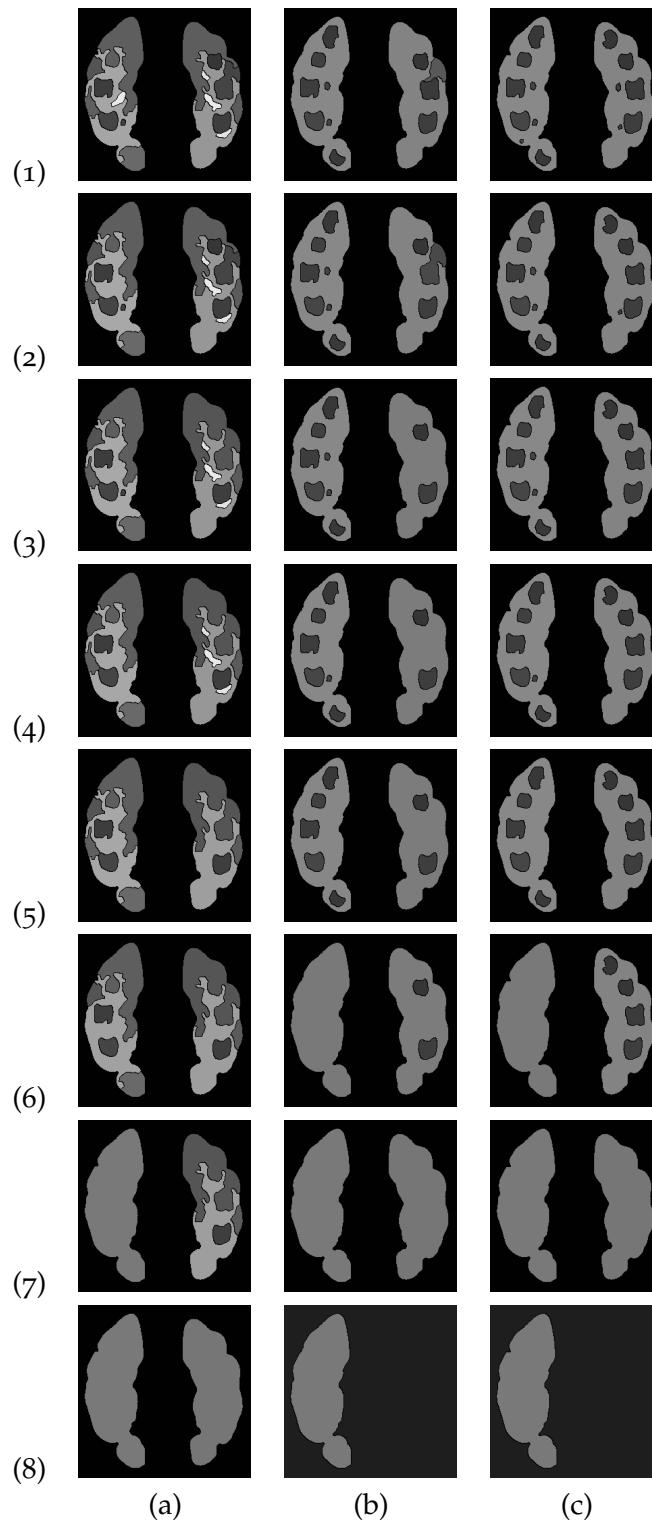


FIGURE 2.6 – Colonnes (a)-(c) : séquences des dernières fusions de la figure 1.9-(a) obtenue par l'algorithme d'escalade, par notre approche avec  $\alpha = 1$  et avec  $\alpha = 3$

relevé que 80% des régions osseuses construites par notre approche sont plus persistantes que celles obtenues par la formulation énergétique. Ceci illustre le rôle positif joué par le coût contextuel durant le processus de fusion. Le tableau 2.1 résume les mesures statistiques calculées sur la persistance, l'échelle d'apparition

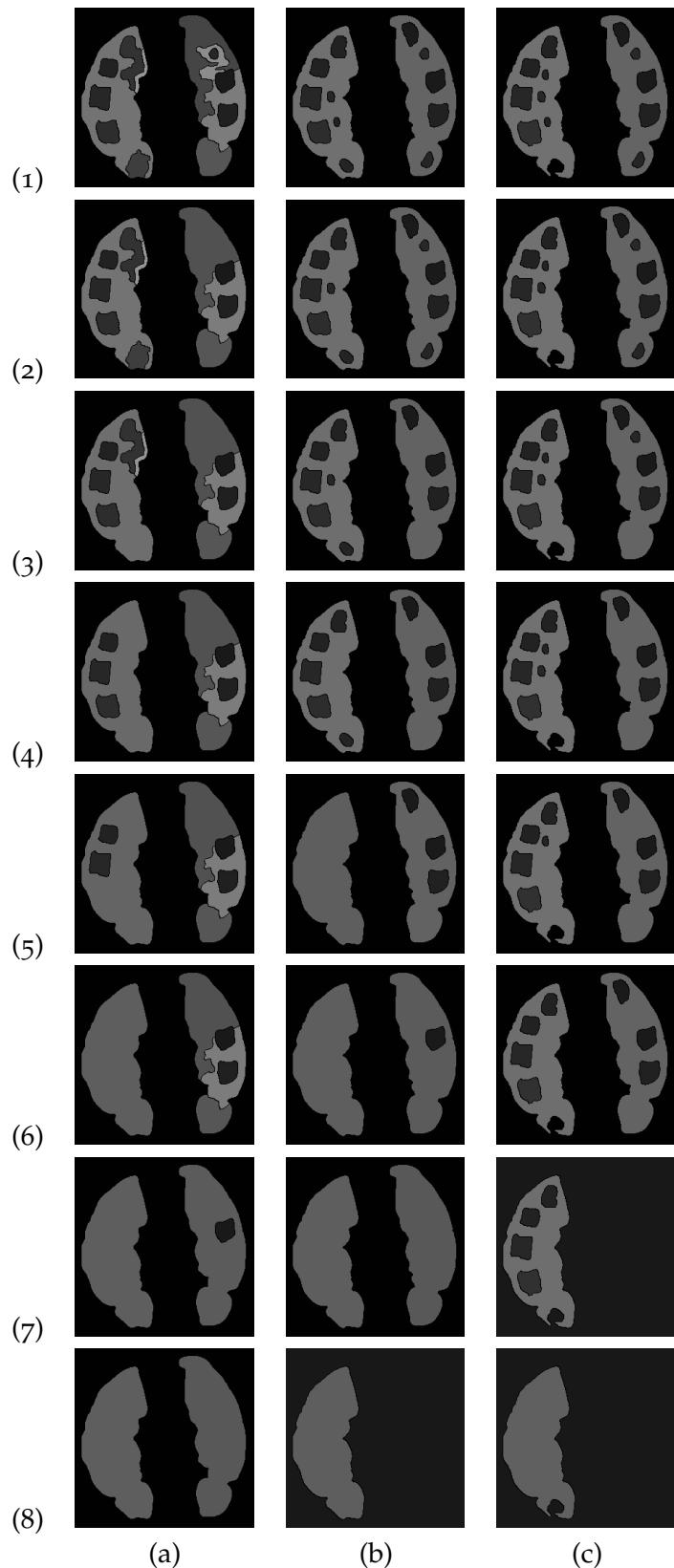


FIGURE 2.7 – Colonnes (a)-(c) : séquences des dernières fusions de la figure 1.9-(c) obtenue par l'algorithme d'escalade, par notre approche avec  $\alpha = 1$  et avec  $\alpha = 3$ .

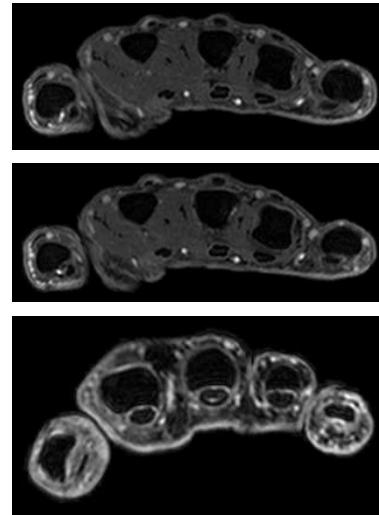


FIGURE 2.8 – Exemples de métacarpes extraits de la base d'image.

et de disparition des os. Elles confirment les propriétés positives, soulignées ci-dessus, de la construction supervisée d'une hiérarchie. Cependant, on relève que les variances empiriques des échelles d'apparition et de disparition et de la persistance sont plus élevées que celles issues des hiérarchies construites par l'algorithme d'escalade binaire.

	<i>Energie</i>	<i>agrégation-contextuelle</i> ( $\alpha = 1$ )	<i>agrégation-contextuelle</i> ( $\alpha = 3$ )
$\hat{\mu}(\Delta\lambda)$	0.139	0.297	0.511
$\hat{\sigma}(\Delta\lambda)$	0.123	0.298	0.601
$med(\Delta\lambda)$	0.125	0.235	0.228
$\hat{\mu}(\lambda^+)$	0.005	0.011	0.013
$\hat{\sigma}(\lambda^+)$	0.009	0.024	0.025
$med(\lambda^+)$	0.001	0.002	0.003
$\hat{\mu}(\lambda^-)$	0.144	0.308	0.524
$\hat{\sigma}(\lambda^-)$	0.123	0.305	0.602
$med(\lambda^-)$	0.127	0.256	0.253

TABLE 2.1 – Tableau de mesures statistiques pour la persistance, l'échelle d'apparition et de disparition des régions osseuses. La première ligne représente les moyennes et les écarts type des persistances, échelle d'apparition et échelle de disparition des os obtenus avec l'approche énergétique. La deuxième et la troisième ligne concernent les os obtenus par notre approche pour  $\alpha = 1$  et pour  $\alpha = 3$ .

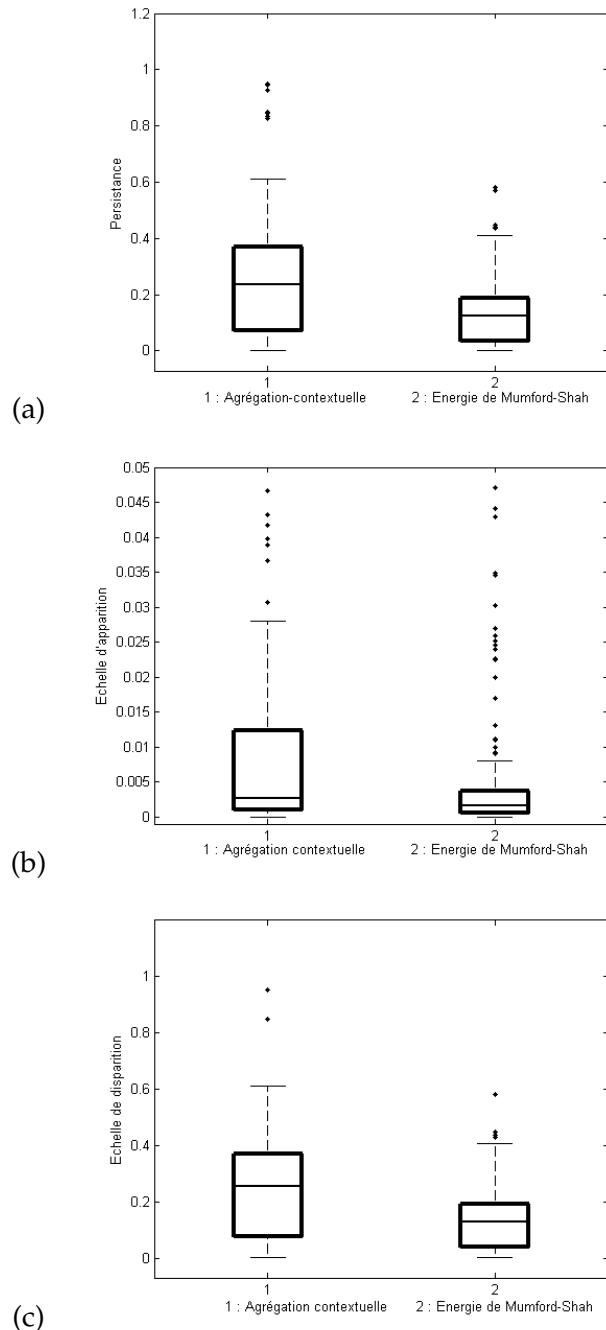


FIGURE 2.9 – Boîtes à moustache pour la persistance, l'échelle d'apparition et l'échelle de disparition des régions osseuses : (a) persistance des os obtenue par l'algorithme d'escalade et par notre approche pour  $\alpha = 1$ . On observe clairement que les régions osseuses, obtenues par notre méthode, sont globalement plus persistantes. (b) et (c) échelles d'apparition et de disparition des régions osseuses.

#### 2.4.1.7 Temps d'exécution

La complexité de l'algorithme d'escalade binaire dépend de plusieurs éléments. L'algorithme se divise en trois étapes. Il comprend une étape d'initialisation lors de laquelle intervient le calcul d'une partition initiale de l'image (ex : Ligne de Partage de Eaux), l'initialisation des énergies sur les régions composant cette partition et le tri de la queue des priorités des fusions binaires. Cette dernière per-

met de trier toutes les fusions binaires possibles selon un certain ordre de priorité. Ici, cet ordre est régit par l'échelle d'apparition. La deuxième étape s'occupe de la construction hiérarchique, elle comprend la suppression de la frontière séparant les deux régions à fusionner, la mise à jour du graphe, la mise à jour des énergies de la nouvelle région ainsi que la mise à jour de la queue de priorité. L'algorithme se termine par une procédure (réursive) d'élagage de la hiérarchie afin d'éliminer les régions non persistantes. L'algorithme dépend également de la nature de l'énergie utilisée à savoir agrégative ou non. Si l'énergie de la fusion peut se déduire des énergies des régions composant la fusion alors l'énergie est dite *agrégative*. L'énergie de Mumford-Shah est un exemple d'une énergie agrégative. La complexité de l'algorithme d'escalade étudiée dans Guigues [2003]<sup>10</sup>, est au pire de  $N^2 \log N$  pour une image  $I$  de taille  $N$ , en partant de la sur-segmentation absolue et en utilisant une énergie agrégative. Cependant, dans la pratique les auteurs soulignent que cette complexité est quasi-linéaire.

L'algorithme par agrégation contextuelle se divise également en plusieurs étapes. L'étape d'initialisation commence aussi par le calcul d'une segmentation initiale de l'image. Il s'en suit le calcul des coûts d'agrégation radiométrique, géométrique et contextuel pour chaque paire de régions adjacentes. Le calcul du coût contextuel est précédé par celui du vecteur d'attributs pour chaque région. La complexité de cette phase dépend du nombre d'attributs et de leurs nature. Ainsi, le calcul des attributs de texture de Law, sélectionnés lors de la phase d'apprentissage, exige une ou plusieurs étapes de filtrage de l'image par les masques de Law correspondant. L'opération de filtrage d'une image est rapide mais le cumul de plusieurs opérations de filtrage par différents masques ralentira le temps d'exécution de l'algorithme. Le calcul d'attributs reposant sur l'histogramme de la région tel que l'entropie, est par contre plus coûteux que le calcul d'énergie, de moyenne ou de variance de régions filtrées. Après avoir calculé les attributs de chaque région, il faut ensuite estimer les vecteurs de probabilités pour chaque région. Cette opération a un coût constant et fait appel à un simple calcul matriciel utilisant les moyennes et les matrices de variances empiriques estimées pour chaque classe. L'étape d'initialisation se termine par la construction d'une queue de priorité des fusions binaires. Ainsi, cette étape d'initialisation est plus lourde que celle de l'escalade binaire. La deuxième étape de l'algorithme est quasiment identique à celle de l'algorithme d'escalade à la différence près qu'elle requiert le calcul des attributs de la nouvelle région. L'utilisation d'attributs possédant une relation de récurrence entre la région et ses sous-structures permet d'accélérer la mise à jour. Des attributs tels que l'entropie, les moments de Hu ne possèdent pas de relation de récurrence, à l'inverse des attributs telles que les moyennes, les variances et les énergies (des régions filtrées) sont dotés de cette relation. Cette relation de récurrence est également utile dans le calcul des critères d'agrégation radiométrique et géométrique. La dernière étape est une procédure d'élagage des régions persistantes. Si la taille du vecteur des attributs n'est pas très élevée et si la relation de récurrence est présente partout où elle est nécessaire alors on devrait aboutir à un algorithme relativement rapide.

Pour résumer, le coût de l'algorithme par agrégation contextuel dépend du nombre d'attributs utilisés, de la nature de ces attributs et de la présence ou non de relations de récurrence.

Le tableau 2.2 montre, pour 3 images de taille  $256 \times 256$ , les temps d'exécution

---

<sup>10</sup>. Nous renvoyons le lecteur intéressé à l'annexe C du rapport de thèse de L. Guigues pour une étude plus complète.

obtenus en utilisant l'algorithme d'escalade binaire « **AEB** » et l'algorithme par agrégation contextuel « **AAC** ». Le temps de calcul a été réalisé avec un processeur Pentium M 740 cadencé à 1.73 GHz. Le tableau indique pour chaque image la taille de la partition de départ (colonne 3) obtenue en utilisant l'algorithme de ligne de partage des eaux. Le temps d'exécution est composée de trois parties : la durée de l'initialisation décrite ci-dessus, la durée de la fusion totale de l'image et celle de la finition de la hiérarchie (phase d'élagage). La dernière colonne indique le temps d'exécution total pour chaque algorithme. L'algorithme d'agrégation contextuelle utilise un vecteur d'attributs de taille 7 pour décrire chaque région de l'image :  $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_5, \epsilon_7, \epsilon_9$  et la compacité de la région. L'attribut  $\epsilon_i$  désigne l'énergie<sup>11</sup> de la région après filtrage de l'image avec le masque de Law correspondant. Il faudra donc opérer 6 opérations de filtrage lors de la phase d'initialisation correspondant chacune à un masque différent. Il faudra également mettre à jour ces mêmes attributs à chaque étape de fusion.

Le temps de finition est quasiment constant pour chaque algorithme et ne représente qu'une partie infime dans le temps d'exécution total. Le temps d'initialisation et le temps de fusion occupe donc une partie importante dans la durée totale. Pour l'algorithme d'escalade le temps d'initialisation est presque identique (au dixième près) et augmente en fonction de la taille de la partition initiale. Ce temps d'initialisation est plus important dans l'algorithme par agrégation contextuelle. Il s'explique par l'étape de filtrage et de calcul des attributs de chaque région. La durée de fusion pour l'algorithme d'escalade est également inférieur à celui par agrégation-contextuel. Ce dernier requiert pour chaque fusion la mise à jour des attributs de la nouvelle région ainsi que ses coût radiométrique et géométrique. Bénéficiant de relations de récurrence partout où cela est nécessaire, il n'en demeure pas moins que l'étape de fusion dans l'approche énergétique ne requiert que la mise à jour de deux attributs, à savoir les énergies radiométrique et géométrique de la nouvelle structure. Le coût d'une fusion dans l'approche agrégative sera de toute façon plus élevé puisque l'on doit effectuer un nombre de mises à jour plus élevé, correspondant au nombre des attributs sélectionnés à l'issu de la phase d'apprentissage.

La durée totale de l'exécution de l'algorithme d'escalade avoisine la demi seconde pour les trois images, tandis que dans l'approche agrégative elle approche la seconde. Malgré la taille du vecteur des attributs ainsi que le nombre d'opérations de filtrage effectuées, la durée d'exécution est assez rapide par rapport à d'autres algorithmes de segmentation tels que les algorithmes de *Graph-Cut*<sup>12</sup>.

La figure 2.10 présente les temps d'exécutions des deux algorithmes en fonction de la taille de la partition initiale, sur un échantillon de 50 images de taille  $256 \times 256$ . Elle montre clairement que le temps d'exécution de l'approche agrégative est supérieur à celui de l'approche énergétique.

11. Voir la définition des masques de Law ainsi que celle de l'énergie dans l'annexe.

12. Le temps d'exécution de l'algorithme de la coupe normalisée (*Normalized Graph Cut*) de Shi et Malik [2000] est en moyenne de 10 secondes sur les mêmes images pour extraire des partitions de 20 régions. Néanmoins, la comparaison est biaisée puisque fondamentalement le problème n'est pas le même. La fusion supprime une information (contour) alors que la division doit créer cette information ce qui engendre un coût beaucoup plus élevé. Le programme utilisé est disponible à l'adresse <http://www.cis.upenn.edu/jshi/software/>.

	Algo.	Base	tps d'init. (sec.)	tps fusion (sec.)	tps finit. (sec.)	tps. total (sec.)
(a)	AEC	1164	0.328	0.218	0.016	0.562
	AAC	1164	0.547	0.453	0.015	1.015
(b)	AEC	1251	0.36	0.218	0	0.578
	AAC	1251	0.468	0.407	0.015	0.89
(c)	AEC	1074	0.328	0.188	0	0.516
	AAC	1074	0.437	0.390	0.016	0.859

TABLE 2.2 – Comparaison des temps d'exécution entre l'algorithme d'escalade binaire et l'algorithme par agrégation contextuelle, pour 3 images du métacarpe de même taille  $256 \times 256$ .

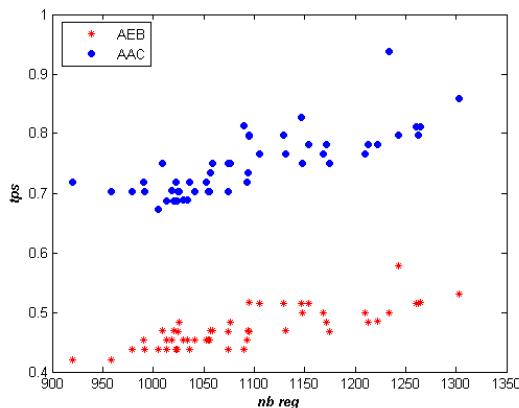


FIGURE 2.10 – Comparaison du temps d'exécution entre l'algorithme d'escalade binaire et l'algorithme d'agrégation contextuelle : cas 2 classes .

## 2.4.2 Cas de 3 classes

Nous allons tester notre approche dans le cas où le nombre de classes est de 3. En plus des deux classes précédentes, nous prendrons en compte la classe représentant le fond de l'image, ce qui signifie que l'étape de pré-traitements consistant à extraire le masque de chaque image sera ignorée. La phase d'apprentissage est en tout point similaire à celle décrite ci-dessus, hormis l'ajout d'un échantillon d'apprentissage pour la troisième classe. Les critères d'agrégation sont également les mêmes que ceux utilisés dans le cas 2 classes. La phase d'apprentissage a permis d'extraire un vecteur d'attributs comportant des mesures de texture, la compacité et l'entropie. Ce dernier ne possède pas de relation de récurrence reliant la région fusionnée aux régions composant la fusion. La procédure de fusion s'en trouvera forcément ralenti.

### 2.4.2.1 Exploration visuelle de la hiérarchie

La figure 2.11 illustre les dernières coupes extraites de deux hiérarchies. La colonne (a) représente celles issues de l'approche énergétique et la colonne (b) celle issues de l'approche par agrégation contextuelle. La colonne (b) montre que les régions osseuses construites avec l'approche agrégative résistent plus longtemps à la fusion avec des régions non-osseuses contrairement à celles construites avec l'approche énergétique. Cependant, on relève que l'os se trouvant au coin inférieur droit du cadre de l'image a survécu moins longtemps que les autres os de

l'image. La raison réside dans le fait que le vecteur des probabilités *a posteriori* de cet os présente cette région comme non-osseuse alors que ce n'est pas le cas. Par conséquent, le coût contextuel a incité sa fusion avec les régions non-osseuses adjacentes. Néanmoins, le coût radiométrique entre cette région et les régions avoisinantes étant très élevé, cela a permis à cette dernière de résister longuement à la fusion. Des erreurs de classification peuvent naturellement se produire et sont inévitables et ce quelque soit le classifieur utilisé. Dans notre contexte, il s'agit d'estimer des probabilités *a posteriori* d'appartenance aux différentes classes et non de décider si telle région appartient à l'une des classes étudiées. Cependant, l'estimation du vecteur des probabilités peut être non-conforme à la réalité et produire un vecteur dont la composante la plus élevée désigne une classe qui n'est pas celle de la région. Dans ce cas, le coût contextuel peut favoriser la fusion entre deux régions en considérant qu'elles appartiennent à la même classe alors que l'une des deux régions appartient à une classe différente. Dans cette situation, les coûts radiométrique et géométrique peuvent éviter des fusions précoces entre régions de classes différentes.

La figure 2.12 montre également que la persistance des régions osseuses est plus élevées dans l'approche agrégative utilisant le coût contextuel. On remarque que l'os du milieu de la main droite n'apparaît dans aucune des deux séquences (colonnes (a) et (b)). En réalité l'os n'a pas été correctement segmenté par les deux méthodes. Cette région osseuse possède une frontière moins saillante avec ses régions voisines et un contraste très faible ce qui n'a pas permis au processus de fusion d'emprunter le bon chemin pour délimiter la zone osseuse.

#### 2.4.2.2 Visualisation par carte de contours

Nous allons maintenant présenter les résultats obtenus sur les deux précédentes coupes du métacarpe sous un autre angle. La représentation hiérarchique est une représentation orientée région mais grâce à la relation étroite qui lie une hiérarchie indicée à une distance ultramétrique (voir l'équation 1.4.1) nous pouvons représenter les résultats sous forme de cartes de contours.

Soient  $R$  et  $R'$  deux régions appartenant à la base de la hiérarchie (elles appartiennent donc à la sur-segmentation qui a permis d'initialiser le processus de fusion). Les régions composant la base de la hiérarchie représentent les points de l'espace. L'ultramétrique est définie par la relation suivante :

$$\delta R, R' = \lambda^+(R \vee R') \quad (2.15)$$

On remarque donc que la distance ultramétrique entre deux régions de la base de la hiérarchie correspond à l'échelle d'apparition de la région *supremum* de ces deux régions dans la hiérarchie, autrement dit à l'échelle de disparition de la frontière qu'elles partagent. Par conséquent, une frontière peut être caractérisée par cette distance ultramétrique. Plus la fusion entre deux régions est tardive et plus la distance ultramétrique qui caractérise la frontière interne de la fusion est élevée. On peut donc présenter une représentation hiérarchique d'une image en régions connexes par une *carte de contours*, dans laquelle l'intensité d'un contour correspond à son échelle de disparition.

Soit  $I$  une image quelconque et soit  $(H, \lambda^+)$  sa hiérarchie indicée construite à partir de la partition initiale  $P_0$ . On notera par  $\lambda^+(I)$  l'échelle à laquelle toute l'image a été fusionnée en une seule région.

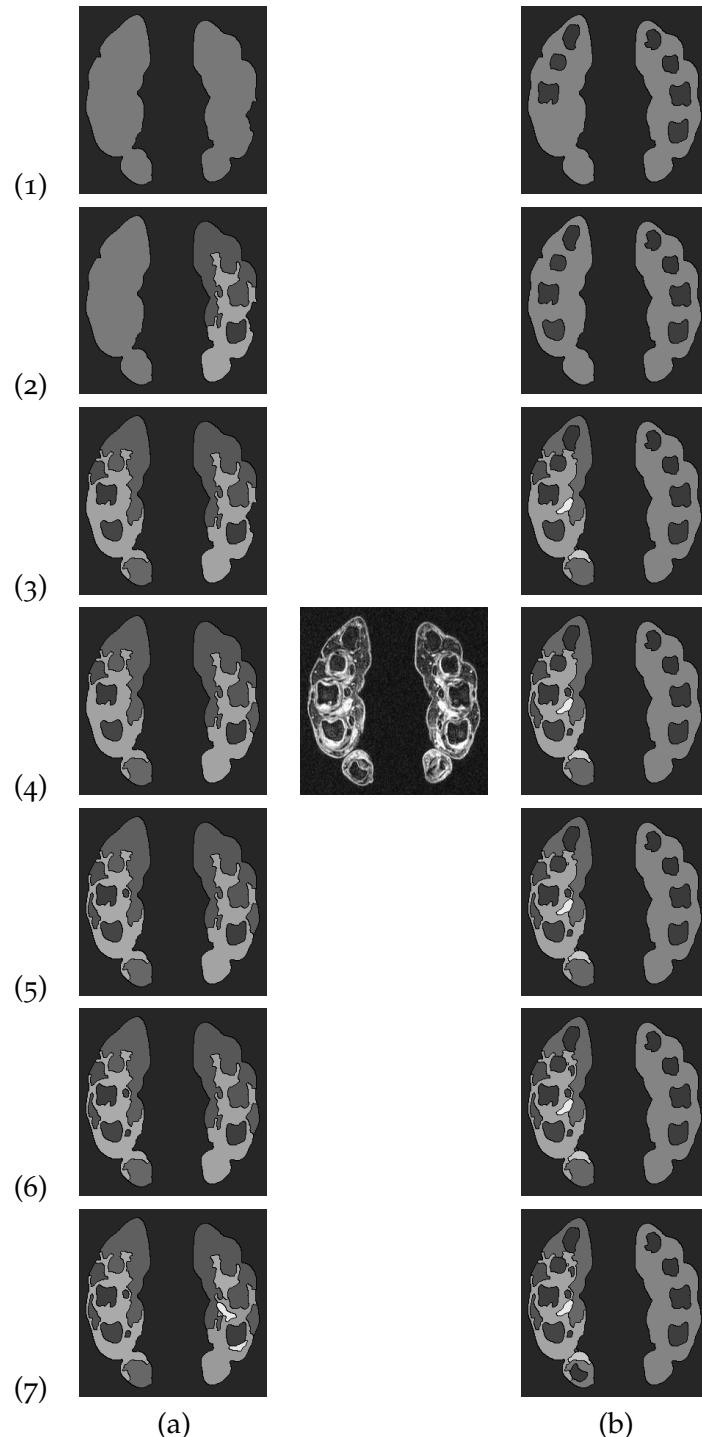


FIGURE 2.11 – Colonnes (a) et (b) : séquences des dernières fusions obtenues par l’algorithme d’escalade et par l’approche agrégative pour  $\alpha = 1$ .

Soit  $f$  une frontière appartenant à l’ensemble des frontières composant la partition initiale  $P_0$ . Son échelle de disparition est définie par

$$\lambda^-(f) = \lambda^+(R \vee R')$$

tel que  $R$  et  $R'$  sont les deux régions de la partition  $P_0$  partageant la frontière  $f$ .

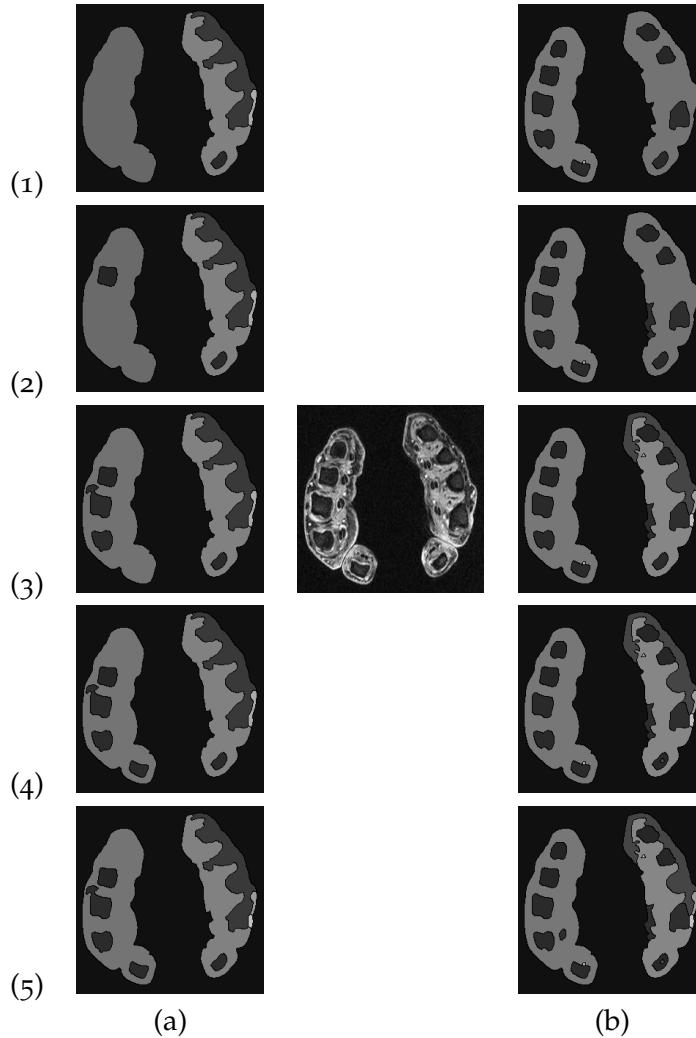


FIGURE 2.12 – Colonnes (a) et (b) : séquences des dernières fusions obtenues par l'algorithme d'escalade et par l'approche agrégative avec  $\alpha = 1$ .

On affecte une intensité au contour  $f$  proportionnelle à  $\log\left(\frac{\lambda^-(f)}{\lambda_0}\right)$ , où  $\lambda_0$  est une constante qui joue le rôle d'un seuil en deçà duquel les frontières sont ignorées.

On pose

$$\lambda_0 = \frac{\lambda^+(\mathcal{I})}{\gamma_0}$$

où  $\gamma_0$  est un réel supérieur à 1.

Ainsi, une frontière ayant une échelle de disparition inférieure à  $\frac{\lambda^+(\mathcal{I})}{\gamma_0}$  aura une intensité négative et on choisira de ne pas la représenter dans la carte.

L'intensité normalisée d'une frontière  $f$  sera définie par

$$\Gamma(f) \triangleq \begin{cases} 0 & \text{si } \lambda^-(f) \leq \lambda_0, \\ \frac{\log(\lambda^-(f))}{\lambda_0} / \log\left(\frac{\lambda^+(\mathcal{I})}{\gamma_0}\right) & \text{sinon.} \end{cases} \quad (2.16)$$

et que l'on peut réécrire par

$$\Gamma(f) \triangleq \begin{cases} 0 & \text{si } \lambda^-(f) \leq \frac{\lambda^+(\mathcal{I})}{\gamma_0}, \\ 1 - \frac{\log(\frac{\lambda^+(\mathcal{I})}{\lambda^-(f)})}{\log(\gamma_0)} & \text{sinon.} \end{cases} \quad (2.17)$$

On voit que  $\gamma_0$  règle le niveau du détail visible sur une carte.

Afin de réaliser une **exploration dichotomique** de la carte de contour, on pose  $\gamma_0 = 2^k$ , tel que  $k = 1, 2, \dots$  (ce schéma est également celui utilisé par Guigues [2003]). On obtient une suite de cartes de contours :

$$\Gamma_k(f) \triangleq \begin{cases} 0 & \text{si } \lambda^-(f) \leq \frac{\lambda^+(\mathcal{I})}{2^k}, \\ 1 - \frac{\log(\frac{\lambda^+(\mathcal{I})}{\lambda^-(f)})}{k \log(2)} & \text{sinon.} \end{cases} \quad (2.18)$$

On notera que lorsque  $k$  augmente, le niveau de détail des contours visibles augmente.

La figure 2.13 illustre les cartes de contours obtenues pour  $k = 1, \dots, 5$ . Lorsque  $k$  décroît, les frontières les plus persistantes demeurent et le niveau de détail diminue. La comparaison visuelle entre les deux premières cartes, correspondant à  $k = 1$ , montrent clairement que les frontières des régions osseuses obtenues avec l'approche par agrégation contextuelle sont plus persistantes que celles obtenues avec l'approche énergétique. Les os sont absents de la première carte issue de l'approche énergétique et l'intensité des frontières osseuses est relativement faible dans les cartes (2) et (3) tandis que dans les cartes (1), (2) et (3) elles sont saillantes. En outre, on remarquera que certains contours osseux sont incomplets dans les cartes (2) et (3) de l'approche ensemble-échelle mais également que les intensités d'un contour osseux peuvent être différentes selon les morceaux des frontières le composant. Ces deux derniers points s'expliquent par le fait que les régions osseuses fusionnent avec d'autres régions non-osseuses formant des régions dont une partie du contour appartient au pourtour réel de l'os. Ainsi, certaines parties du contour de l'os seront plus persistantes que d'autres et auront une intensité en niveau de gris plus élevée. A l'inverse, l'approche par agrégation contextuelle montrent que l'intégrité des frontières osseuses est préservée dans les sept os sur les neuf contenus dans l'image (le dixième os - au coin bas gauche - représente un cartilage). Les intensités des frontières composant les pourtours des os sont quasiment similaires ce qui permet de percevoir ces os comme des régions à part entière puisqu'il n'y a pas de discontinuité visuelle dans l'intensité. Les deux dernières régions osseuses sont moins persistantes que les autres, néanmoins leurs pourtours ont des intensités similaires le long des morceaux de frontières les composant ce qui n'est pas le cas dans l'approche énergétique.

La figure 2.14 représente également des cartes de contours (exploration dichotomique) issues de l'image du métacarpe de la figure 2.12. Ce deuxième exemple vient confirmer toutes les observations faites ci-dessus.

#### 2.4.2.3 Persistance des régions

Afin de mesurer et matérialiser l'effet de l'introduction du coût contextuel, nous avons extrait les persistance, les échelles d'apparition et de disparition des régions osseuses sur la même base utilisée dans la section 2.4.1.6. Les échelles des régions osseuses (qu'elles soient d'apparition ou de disparition) ont été normalisées relativement aux échelles maximales des hiérarchies de leurs images respectives.

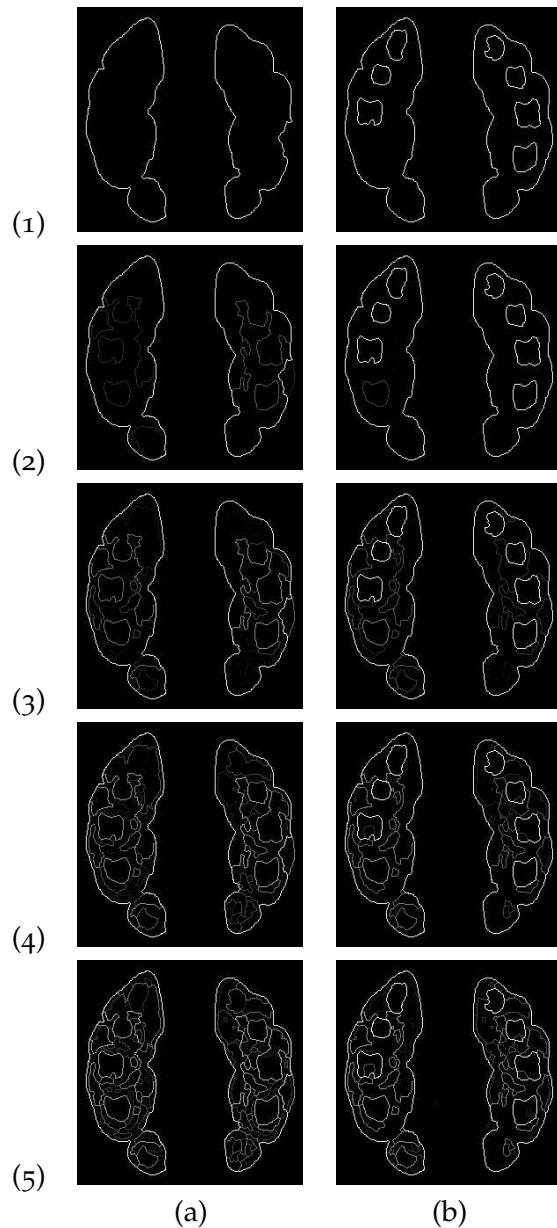


FIGURE 2.13 – Colonnes (a)-(b) : séquences des cartes de contours (exploration dichotomique) obtenue par l'algorithme d'escalade et par notre approche avec  $\alpha = 1$ .

La figure 2.15-(a) représente les boîtes à moustache de la persistance des os obtenue avec l'approche par agrégation-contextuelle (figure de gauche) et avec l'approche ensemble-échelle (figure de droite). On remarque clairement que la persistance des os est plus élevée dans l'approche agrégative que celle obtenue par l'algorithme d'escalade. La figure 2.15-(b) illustrent les apparitions des os représentées sous forme d'une boîte à moustache. On remarque que les échelles d'apparitions dans notre approche sont globalement identiques à celles obtenues avec l'approche ensemble-échelle avec une dispersion plus faible. Les échelles de disparition, quand à elles sont globalement largement supérieures dans l'approche agrégative à celles issues de l'approche énergétique, comme l'illustre les boîtes à moustache de la figure 2.15-(c).

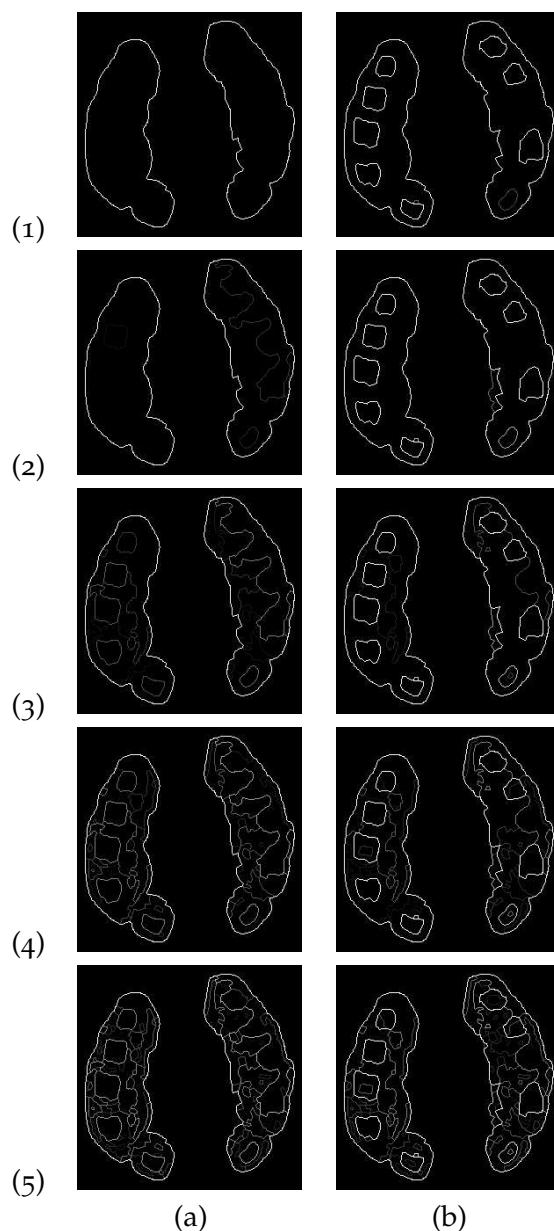


FIGURE 2.14 – Colonnes (a)-(b) : séquences des cartes de contours dichotomiques obtenue par l'algorithme d'escalade et par notre approche avec  $\alpha = 1$ .

Ainsi, comme cela a été souligné dans le cas 2 classes, les régions osseuses auront une forte probabilité d'être présentes dans les hautes échelles des hiérarchies. On a également relevé que 85% des régions osseuses construites par notre approche sont plus persistantes que celles obtenues par la formulation énergétique. Ceci illustre le rôle positif joué par le coût contextuel durant le processus de fusion.

Le tableau 2.3 résume lui aussi les mesures statistiques calculées sur la persistance, l'échelle d'apparition et de disparition des os. Elles confirment les observations, soulignées ci-dessus, dans le cas 2 classes. On notera également que les dispersions de l'échelle de disparition et de la persistance sont plus élevées que

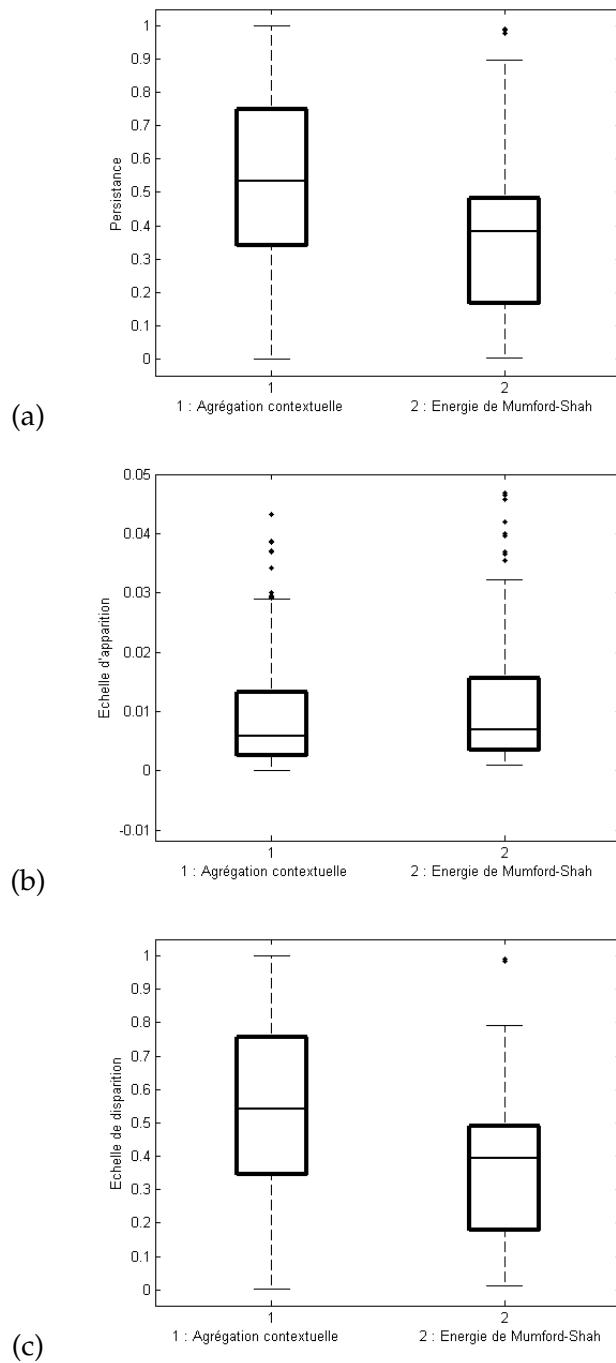


FIGURE 2.15 – Boîtes à moustache pour la persistance, l'échelle d'apparition et l'échelle de disparition des régions osseuses : (a) persistance des os obtenue par l'algorithme d'escalade et par notre approche pour  $\alpha = 1$ . On observe clairement que les régions osseuses, obtenues par notre méthode, sont globalement plus persistantes. (b) et (c) échelles d'apparition de disparition des régions osseuses.

celles issues des hiérarchies construites par l'algorithme d'escalade binaire, alors que la dispersion dans l'échelle d'apparition est quasiment identique.

	<i>Energie</i>	<i>agrégation-contextuelle (<math>\alpha = 1</math>)</i>
$\hat{\mu}(\Delta\lambda)$	0.3428	0.5218
$\hat{\sigma}(\Delta\lambda)$	0.1952	0.2992
$med(\Delta\lambda)$	0.3846	0.5346
$\hat{\mu}(\lambda^+)$	0.015	0.0114
$\hat{\sigma}(\lambda^+)$	0.0208	0.0211
$med(\lambda^+)$	0.0069	0.0058
$\hat{\mu}(\lambda^-)$	0.3578	0.5332
$\hat{\sigma}(\lambda^-)$	0.1970	0.3043
$med(\lambda^-)$	0.3946	0.5404

TABLE 2.3 – Tableau de mesures statistique pour la persistance, l'échelle d'apparition et de disparition des régions osseuses. La première ligne représente les moyennes et les écarts type des persistances, échelle d'apparition et échelle de disparition des os obtenus avec l'approche énergétique. La deuxième et la troisième ligne concernent les os obtenus par notre approche pour  $\alpha = 1$  et pour  $\alpha = 3$ .

#### 2.4.2.4 Temps d'exécution

Le tableau 2.4 montre, pour les mêmes images utilisées dans le tableau 2.2, les temps d'exécutions obtenus en utilisant l'algorithme d'escalade binaire « **AEB** » et l'algorithme par agrégation contextuel « **AAC** ». On notera que la taille de la partition initiale (colonne 3) est plus élevée que dans le cas deux classes puisque le fond de l'image n'a pas été supprimé. Le temps d'exécution est toujours représenté par la somme de la durée de l'initialisation, la durée de la fusion totale et celle de l'élagage). L'algorithme d'agrégation contextuelle utilise un vecteur d'attributs de taille 8 pour décrire chaque région de l'image :  $\epsilon_5, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_5, \hat{\mu}_8, \epsilon$  (l'entropie) et la compacité de la région. L'attribut  $\hat{\mu}_i$  désigne la moyenne de la région après filtrage de l'image avec le masque de Law correspondant. L'étape d'initialisation comprend entre autre donc 6 opérations de filtrage. Lors de la mise à jour des attributs, une relation de récurrence est disponible pour chaque attribut sauf pour l'entropie.

Comme relevé dans le cas 2 classes, le temps de finition est quasiment constant pour chaque algorithme et ne représente qu'une partie infime dans le temps d'exécution total. Il est sensiblement plus élevé dans l'approche agrégative (au centième près). Expérimentalement, on a constaté que le nombre de noeuds non persistants après l'étape d'élagage, dans une approche agrégative avec ou sans coût contextuel, est toujours supérieur à celui produit par l'approche énergétique. Le temps d'initialisation dans l'approche par agrégation contextuel est sensiblement supérieur à celui de l'approche énergétique alors que le temps de fusion a plus que doublé dans notre approche. C'était également le cas lorsque l'on a étudié le cas de 2 classes (voir tableau 2.1). Les raisons de cette augmentation sont les mêmes que celles qu'évoquées dans le cas 2classes.

La durée totale de l'exécution de l'algorithme par agrégation contextuelle a doublée par rapport à celle de l'approche énergétique. C'est à l'étape de fusion où les écarts sont significatifs. Les mises à jours des différents attributs et l'absence de relation de récurrence pour l'attribut de l'entropie ont contribuer à accroître la durée de la fusion.

Les temps d'exécutions des deux algorithmes en fonction de la taille de la partition initiale, sont illustrés dans la figure 2.16 (échantillon de 50 images de

	Algo.	Base	tps d'init. (sec.)	tps fusion (sec.)	tps finit. (sec.)	tps. total (sec.)
(a)	AEC	3672	0.609	1.219	0.031	1.859
	AAC	3672	0.890	2.875	0.047	3.812
(b)	AEC	3251	0.547	1.047	0.031	1.625
	AAC	3251	0.828	2.515	0.047	3.390
(c)	AEC	3269	0.562	1.188	0.031	1.781
	AAC	3269	0.860	2.546	0.047	3.453

TABLE 2.4 – Comparaison des temps d'exécution entre l'algorithme d'escalade binaire et l'algorithme par agrégation contextuelle, pour 3 images du métacarpe de même taille  $256 \times 256$ .

taille  $256 \times 256$ ). Ils sont en moyenne deux fois supérieurs à ceux de l'approche énergétique.

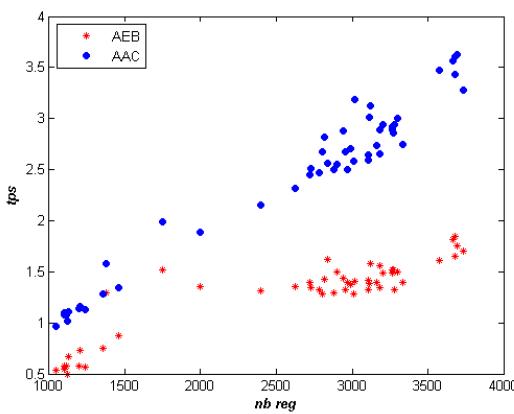


FIGURE 2.16 – Comparaison du temps d'exécution entre l'algorithme d'escalade binaire et l'algorithme d'agrégation contextuelle : cas 2 classes .

## 2.5 DISCUSSION

La nouvelle approche par fusion de régions présentée dans ce chapitre intègre à la fois des informations radiométrique, géométrique et contextuelle sous forme d'un critère d'agrégation. Le premier terme représente la qualité de la régression, le second est un terme de régularisation alors que le troisième terme représente les connaissances *a priori* sur les classes composant l'image. Ce dernier pénalise les fusions entre régions ayant une forte probabilité d'appartenir à des classes différentes et par conséquent retarde la décision de fusions de ces régions. À travers ce comportement, notre approche encourage implicitement les fusions entre régions possédant une forte probabilité d'appartenir à la même classe. Les premiers résultats montrent une réelle amélioration dans la sémantique des coupes extraites des hiérarchies. Ils montrent aussi un accroissement dans la persistance des régions pertinentes. Il semble clair que le processus de fusion est mieux guidé quand le *coût contextuel* est inclus. Un autre avantage de cette approche est son caractère générique. Elle peut être facilement adaptée aux spécificités de l'application en terme des régions d'intérêt sur lesquelles elle doit se focaliser. Cette nouvelle approche conduit à une représentation hiérarchique de l'image avec un contenu sémantique

plus cohérent et conforme à la réalité contrairement à celle obtenue par l'approche ensemble-échelle. Il faut toutefois noter que cette approche dépend d'un paramètre  $\alpha$  qui pondère l'influence du coût contextuel par rapport aux autres coûts. Les expérimentations que nous avons mené n'ont pas permis de dégager un compromis quand au choix de la valeur de ce paramètre. Il dépend fortement de la qualité de l'image en terme de contraste des régions et de saillance des frontières.

Nous avons appliqué cette approche dans le cas 2 et 3 classes. Les résultats obtenus montrent que l'ajout d'un coût contextuel en contre partie de la perte du formalisme énergétique permet de mieux segmenter les régions d'intérêt. Les résultats exposés utilisent tous la distance Kullback-Leibler symétrisée comme coût contextuel. Néanmoins, nos expériences avec d'autres distances telles que la distance de Jensen-Shannon (2.3.3) ou celle de Johnson et Sinanovic (2.3.3) n'ont pas montré une différence significative avec les résultats obtenus avec la distance de Kullback-Leibler symétrisée.

Le risque d'erreur d'estimation n'est pas négligeable et peut conduire à des aberrations, notamment quand le nombre de classes augmentent et que la séparabilité entre les classes n'est pas assez forte. L'estimation peut produire des vecteurs de probabilités (d'appartenance aux différentes classes) non conformes à la nature réelle des régions et dans certains cas un vecteur de probabilités peut être quasi-uniforme ce qui traduit une certaine confusion quand à l'appartenance de la région à une certaine classe. Visiblement, pour que le coût contextuel puisse se prononcer de manière franche sur l'appartenance de deux régions à une même classe ou non, il faudra que les vecteurs de probabilités possèdent une composante plus élevée que le reste des composantes. Dans ce cas, il en résulte soit un faible coût contextuel (ici la distance de Kullback-Leibler symétrisée) révélant que les deux régions sont *a priori* de même classe, soit un coût élevé indiquant que les deux régions sont probablement de classes différentes.

Nous illustrons dans ce qui suit le comportement global de la distance de Kullback-Leibler symétrisée. Le premier tableau 2.5 montre un exemple des distances entre 3 vecteurs de probabilités empiriques où chacun possède une composante élevée qui désigne *a priori* la classe à laquelle appartient la région. L'indice de cette composante est différent pour chaque vecteur, désignant chacun une classe différente. On remarque que la distance (la pénalisation) est élevée entre les trois vecteurs.

$R / R'$	$p_1 = (0.8, 0.1, 0.1)$	$p_2 = (0.2, 0.8, 0)$	$p_3 = (0, 0.2, 0.8)$
$p_1 = (0.8, 0.1, 0.1)$	0	1.98	4.95
$p_2 = (0.2, 0.8, 0)$		0	5.22
$p_3 = (0, 0.2, 0.8)$			0

TABLE 2.5 – Illustration du comportement de la distance de Kullback-Leibler symétrisée. Les trois vecteurs de probabilités possèdent tous un seul mode désignant une classe différente. La pénalisation est élevée entre ces vecteurs de probabilités.

Le deuxième tableau 2.6 montre les distances obtenues entre vecteurs désignant *a priori* la même classe. Les deux distributions les plus proches sont  $p_1 = (0.8, 0.1, 0.1)$  et  $p_3 = (0.7, 0.2, 0.1)$ . Les vecteurs  $p_1$  et  $p_2$  ont une plus forte probabilité d'appartenir à la même classe que le vecteur  $p_3$ . Cependant, la distance de Kullback-Leibler symétrisée ne désigne pas les vecteurs  $p_1$  et  $p_2$  comme étant les plus proches. La pénalisation entre  $p_1$  et  $p_2$  est plus forte qu'entre les vecteurs  $p_1$  et  $p_3$  (en raison de la valeur nulle de la deuxième composante de  $p_2$ ). La distance

de Kullback-Leibler mesure la similarité entre l'allure de deux distributions et ne se focalise pas sur une composante en particulier d'une distribution empirique.

$R / R'$	$p_1 = (0.8, 0.1, 0.1)$	$p_2 = (0.9, 0, 0.1)$	$p_3 = (0.7, 0.2, 0.1)$
$p_1 = (0.8, 0.1, 0.1)$	0	0.34	0.09
$p_2 = (0.9, 0, 0.1)$		0	0.80
$p_3 = (0.7, 0.2, 0.1)$			0

TABLE 2.6 – Illustration du comportement de la distance de Kullback-Leibler symétrisée. Les trois vecteurs de probabilités possèdent tous le même mode désignant la première classe. Le vecteur  $p_3$  est plus proche que le vecteur  $p_2$  du vecteur  $p_1$  alors qu'a priori ces deux derniers ont une plus forte probabilité d'appartenir à la même classe que le vecteur  $p_3$ .

Le dernier tableau 2.7 montre qu'une distribution empirique relativement uniforme,  $p_3 = (0.4, 0.3, 0.3)$  peut avoir une distance assez faible avec des distributions qui ont une composante relativement élevée dans une classe. En outre, on remarquera que la distance entre  $p_1 = (0.7, 0.2, 0.1)$  et  $p_3 = (0.4, 0.3, 0.3)$  est plus faible que celle entre  $p_1 = (0.8, 0.1, 0.1)$  et  $p_2 = (0.9, 0, 0.1)$  dans le tableau 2.6, malgré que ces deux dernières reflètent de notre point de vue une plus forte similarité mais en terme d'appartenance à une même classe.

$R / R'$	$p_1 = (0.7, 0.2, 0.1)$	$p_2 = (0.6, 0.3, 0.1)$	$p_3 = (0.4, 0.3, 0.3)$
$p_1 = (0.7, 0.2, 0.1)$	0	0.04	0.29
$p_2 = (0.6, 0.3, 0.1)$		0	0.19
$p_3 = (0.4, 0.4, 0.3)$			0

TABLE 2.7 – Illustration du comportement de la distance de Kullback-Leibler symétrisée. Le vecteur  $p_3$  est proche d'une distribution uniforme. Néanmoins, sa distance avec des vecteurs ayant un mode ( $p_1$  et  $p_2$ ) est assez faible.

L'analyse du comportement de la « distance de Kullback-Leibler symétrisée » à travers ces quelques exemples numériques, montre qu'il est souhaitable d'avoir des vecteurs de probabilité tendant de manière relativement significative vers l'une des classes étudiées. Lorsqu'un le vecteur des probabilités *a posteriori* d'une région est relativement uniforme, la pénalisation avec des régions adjacentes, ayant des distributions *a posteriori* penchant vers une classe, peut être relativement faible. Cette contrainte paraît relativement surmontable dans le cas binaire càd le cas de 2 classes mais elle deviendra plus difficile à atteindre au fur et à mesure que le nombre de classes augmentera et particulièrement lorsque la séparabilité entre ces classes sera moins franche.

## CONCLUSION DU CHAPITRE

Au cours de ce chapitre, nous avons proposé une méthode de construction supervisée d'une hiérarchie de régions. Elle dérive directement du formalisme énergétique de l'approche ensemble-échelle. Cette approche repose sur la combinaison de trois critères d'agrégation. L'injection de connaissances *a priori* montre clairement que le processus de fusion est mieux guidé. La preuve de l'influence positive du coût contextuel dans la qualité de décision s'exprime à travers l'accroissement de la persistance des régions d'intérêt. Les résultats expérimentaux montrent que les partitions obtenues sont plus conformes au contenu de l'image.

La représentation par cartes de contours vient également appuyer nos observations, elles montrent clairement que les frontières composant le contour d'une région d'intérêt restent unies et ne disparaissent pas les unes après les autres en raison des fusions inter-classes.

Il est vrai que l'on perd la notion d'échelle introduite par le formalisme énergétique. Mais une représentation hiérarchique garde toujours une certaine notion d'échelle puisqu'elle permet d'explorer l'image du détail le grossier au détail le plus fin.

L'erreur de classification ou dans notre cas d'estimation des probabilités empiriques est inhérente à n'importe quelle méthode de supervision. Notre approche n'en réchappe pas et le fait que le coût contextuel soit combiné avec deux autres critères permet d'atténuer cet inconvénient. L'autre inconvénient est la mesure de dissimilarité entre distributions empiriques, utilisée pour calculer le coût contextuel. Son comportement peut parfois s'avérer paradoxal mais théoriquement prévisible, il n'en demeure pas moins que les mesures utilisant la distance de Kullback-Leibler s'attachent à comparer l'allure des distributions et ne se focalisent pas uniquement sur les modes de celles-ci (à savoir les pics de la distribution).

Expérimentalement, le temps d'exécution de l'algorithme d'agrégation contextuelle est deux fois plus élevé que celui de l'escalade binaire. Cependant, compte tenu du nombre d'attributs à calculer ainsi que des opérations de filtrage nécessaires, il reste relativement rapide. Néanmoins, ceci n'est valable qu'à condition que les attributs et les critères radiométrique et géométrique possèdent une relation de récurrence. L'approche nous contraint donc d'éviter d'utiliser plusieurs attributs non pourvu de relations de récurrence. Cette dernière remarque doit être atténuée car tout est une question d'intelligence dans le calcul et l'implémentation. Un attribut peut ne pas avoir une relation de récurrence directe mais l'on peut dans certains cas décomposer son calcul en plusieurs valeurs. Certaines valeurs peuvent alors contenir une relation de récurrence permettant une rapide mise à jour lors de la fusion.

L'approche par agrégation contextuelle a permis de montrer qu'une supervision de la décision de fusion améliore nettement la qualité de cette dernière et construit des hiérarchies plus conformes au contenu de l'image.



# ENSEMBLE-ÉCHELLE PAR INJECTION D'UN *a priori* SPATIAL

3

L'injection de connaissances *a priori*, qu'elles soient radiométriques et/ou géométriques, dans le processus de fusion, décrit dans le chapitre précédent a permis de mieux guider ce processus. La supervision permet d'améliorer la qualité de décision face à la variabilité et à la complexité des objets. Cependant, elle nécessite la mise à disponibilité d'une base d'images dont le contenu sémantique est identique. Cette approche est indiquée dans des problèmes tels que l'aide à l'interprétation d'images médicales. Cependant, quand l'objectif d'un expert est d'analyser une seule image composée d'objets complexes, cette méthode n'est d'aucune utilité.

Dans ce chapitre, nous proposons une autre approche complètement différente mais dont l'objectif est toujours d'adapter la construction hiérarchique à la complexité des objets composant l'image. L'objectif ici est d'intégrer un *a priori* spatial sous la forme d'une énergie sous-additive<sup>1</sup> afin de souligner l'importance de certains objets et structures dans l'image.

Dans la première partie on explicitera plus en détail notre démarche et nous justifierons l'importance d'injecter un *a priori* spatial dans l'image. La deuxième partie du chapitre sera consacrée au développement théorique de notre approche. On présentera les différentes façons d'aboutir à la construction d'une énergie spatiale sous-additive. Dans la troisième partie, on présentera des exemples concrets d'énergies spatiales sous-additives. La dernière partie sera consacrée aux différentes expérimentations menées sur un large panel d'images.

## 3.1 PRÉSENTATION DE LA DÉMARCHE

Nous avons vu qu'une représentation ensemble-échelle permet de décrire l'image sous la forme d'une séquence monotone de partitions, par fusions successives des régions. Cette construction s'appuie sur la minimisation d'une énergie qui suppose le plus souvent un modèle constant par morceaux. Néanmoins, une image contient des objets complexes pour lesquels le modèle constant par morceaux n'est pas adapté. Plus précisément, des objets tels qu'une voiture, un visage,

---

1. La justification de la condition de sous-additivité sera abordée dans la section 3.1.2

un piétons, se décomposent en objets homogènes dont la somme forme un objet cohérent<sup>2</sup>. Les structures composant de tels objets possèdent des caractéristiques radiométriques et géométriques différentes et les fusionner en s'appuyant uniquement sur ce modèle s'avère être une tâche ardue. De fait, les régions composant ces objets se trouvent absorbées par d'autres régions (dépend aussi du fond entourant l'objet d'intérêt). L'objet d'intérêt ainsi que ses structures n'apparaissent plus comme des zones remarquables dans la hiérarchie et la recherche d'une coupe contenant les structures de cet objet devient alors plus fastidieuse puisqu'elle nécessite une exploration plus profonde de la hiérarchie. Il en résulte des partitions comportant un nombre important de régions.

Néanmoins, les structures d'un objet, même si elles sont hétérogènes, partagent une propriété qui permet en partie au système visuel humain de faire leur regroupement. Elles sont spatialement contiguës, de ce fait il existe une certaine cohésion spatiale entre ces régions qui n'est pas partagée avec les régions voisines. Cette cohésion spatiale s'étend au domaine spatial qu'occupe l'objet d'intérêt. Ainsi, incorporer une information *a priori* spatiale en amont permet de souligner l'importance de certaines zones de l'image. L'introduction d'un tel *a priori* peut se faire par une localisation grossière des objets et se présenter sous la forme de blobs obtenus soit par un pré-traitement de l'image soit à l'aide d'un outil graphique interactif comme le montre les figures 3.1 et 3.2.

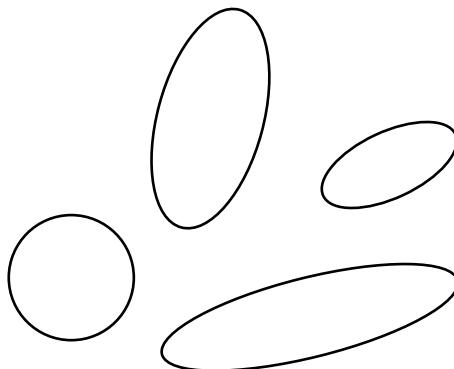


FIGURE 3.1 – Exemple d'un *a priori* spatial sous forme de blobs elliptiques.

Généralement, qualifier une partition de « bonne segmentation » dépend fortement de l'application visée. Or, celle-ci peut nous indiquer les zones de l'image qui sont pertinentes et par conséquent nous recommander de nous focaliser particulièrement sur ces zones. Dans l'absolu, une analyse multi-échelle sans *a priori* produit une description de l'image où l'échelle d'apparition ainsi que la persistance caractérisent la pertinence de chaque région de la hiérarchie. Comme nous l'avons souvent rappelé, l'approche se base sur des critères bas-niveau et dans ce cas les objets prédominants dans la hiérarchie le seront grâce à leur homogénéité, leur taille, et tout particulièrement la différence relative qui existe entre leurs propriétés et celles de leur voisinage. La hiérarchie obtenue n'est pas nécessairement pertinente pour l'application car des régions intéressantes ont pu être plus rapidement absorbées lors du processus de fusion. C'est souvent le cas quand la région d'intérêt est un objet composé de régions homogènes mais différentes entre elles

2. Notons que cette opération est difficilement réalisable par un algorithme de segmentation de bas-niveau. Elle relève du domaine de la reconnaissance des formes ou les connaissances *a priori* sont essentielles pour faire un regroupement d'objets homogènes sous la bannière d'un objet ayant une interprétation sémantique.

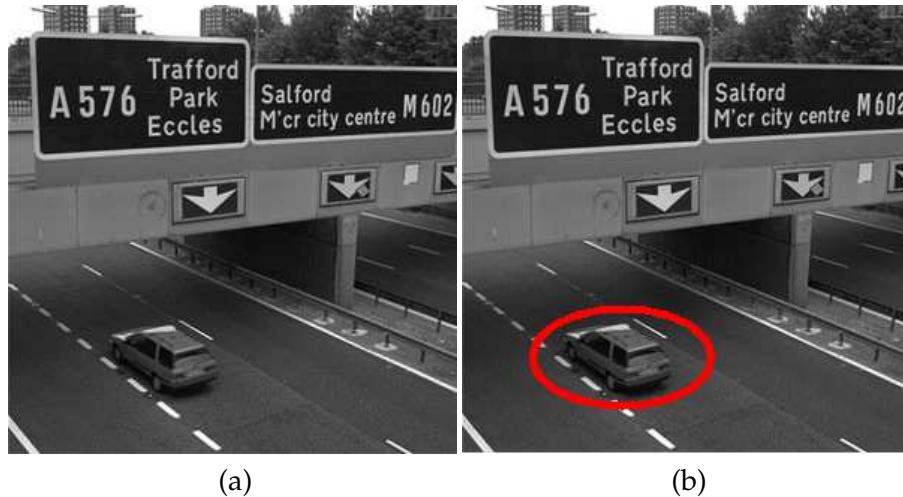


FIGURE 3.2 – (a) une image du périphérique ainsi que l'image des blobs illustrant deux zones d'intérêt (b).

(piéton, voiture, etc.). Dans ces conditions, si un expert estime que certaines zones de l'image sont pertinentes, il convient alors de les délimiter grossièrement afin de freiner les fusions des régions concernées et préserver ces zones d'intérêt. (voir la figure 3.3).

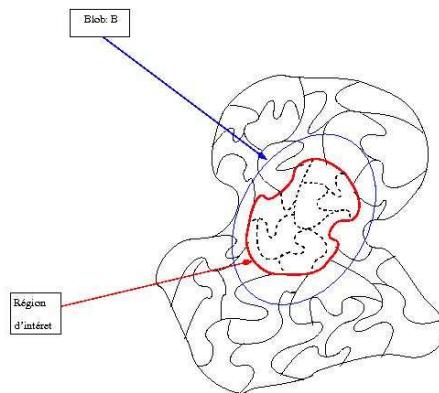


FIGURE 3.3 – Exemple d'un blob elliptique contenant un objet d'intérêt.

### 3.1.1 Principe d'un potentiel d'attraction

Soit  $I$  une image dont le support spatial est noté  $\Omega$ . Soit :

$$\mathcal{B} = \{B_1, \dots, B_M\}.$$

un ensemble de blobs disjoints appartenant au support de l'image  $\Omega$ .

On désignera par  $I_{\mathcal{B}}$  l'image-blob correspondant à l'image  $I$  comme le montre la figure 3.1.

Hormis la condition de séparation entre les blobs, aucune autre contrainte n'est imposée aux blobs (telle que la forme), cependant un blob est le plus souvent de forme elliptique. Notre modélisation ainsi que tout le formalisme théorique est indépendant de la forme du blob mais nos expérimentations se baseront sur des blobs elliptiques.



FIGURE 3.4 – *Objet complexe avec des structures relativement homogènes.*

Tout blob  $B_k$  sera caractérisé par son centre de gravité spatial qu'on notera  $c_k$ .

L'idée directrice de notre approche repose sur le principe de l'existence d'une forte attraction entre les structures à l'intérieur des blobs par rapport aux structures entourant les blobs. Nous allons pour cela, à la manière du principe d'attraction entre objets dans la théorie de gravitation, définir un potentiel caractérisant l'attraction qu'exerce un blob sur une région de l'image. Néanmoins, nous ne parlerons pas de forces d'attraction, qui impliquent une représentation vectorielle orientée comme c'est le cas pour la théorie de gravitation, mais plutôt de potentiel ou d'intensité de l'attraction entre régions ou entre régions et blobs.

Dans la théorie de la gravitation (ou celle de l'électrostatique), l'attraction entre deux objets est inversement proportionnelle à la distance les séparant. Dans notre modélisation, nous adopterons la position inverse. Le potentiel d'une région relativement à un blob sera croissant par rapport à la distance les séparant. Autrement dit, le potentiel entre deux régions sera proportionnel à la distance séparant deux régions<sup>3</sup>.

Face à un ensemble de blobs, chaque région sera par conséquent décrite par un vecteur de potentiels dont la dimension correspond au nombre de blobs contenu dans l'image. L'énergie d'une région s'écrit alors en fonction de ce vecteur.

### 3.1.2 Pour une énergie spatiale sous-additive

L'intégration d'une énergie spatiale dans le formalisme ensemble-échelle exige que celle-ci soit sur-additive ou sous-additive (cf. section 1.4). L'énergie affine comporte un terme de régularisation qui exprime un *a priori* sur le modèle inféré. Ainsi dans le modèle de Mumford-Shah, l'énergie de régularisation est le périmètre d'une région, elle exprime le souhait d'obtenir des partitions aux contours réguliers. On peut utiliser l'énergie de convexité en lieu et place du périmètre afin de privilégier la construction de régions convexes. L'énergie spatiale qu'on

<sup>3</sup>. En physique, l'attraction est inversement proportionnelle au carré de la distance séparant les deux objets :  $\frac{1}{d^2}$ .

souhaite introduire se base également sur un *a priori* et par conséquent il nous paraissait évident qu'elle devait vérifier la propriété de sous-additivité.

Nos recherches se sont naturellement orientées vers la construction d'une énergie spatiale sous-additive. Néanmoins, la construction d'une énergie spatiale sur-additive n'est pas à exclure mais nous n'avons pas exploré cette voie pour les raisons évoquées ci-dessus.

## 3.2 MODÉLISATION DE L'ÉNERGIE SPATIALE

Nous envisageons de définir trois types d'énergies spatiales comme le montre la figure 3.5. L'énergie mono-potentielle reflète l'intensité de l'attraction entre la région et un blob. Cette pseudo-atraction sera mesurée par un potentiel entre une région et un blob que l'on définira ultérieurement. A la différence de l'énergie mono-potentielle, l'énergie multi-potentielle tiendra en compte de l'énergie mono-potentielle de la région étudiée mais également des énergies des régions qui lui sont voisines. L'énergie bi-potentielle d'une régions prendra également en compte le voisinage de la région mais elle s'appuiera sur la définition d'une intensité d'attraction entre deux régions voisines relativement à un blob. Ainsi, on définira la notion de bi-potentiel entre deux régions relativement à un blob qui mesurera l'attraction entre deux régions.

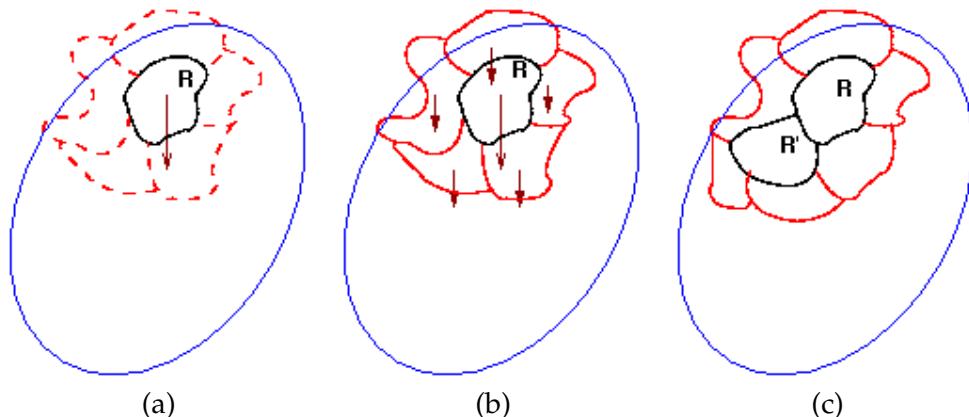


FIGURE 3.5 – (a) L'énergie mono-potentielle d'une région représente par une pseudo attraction de la région avec un blob. (b) l'énergie multi-potentielle de la région prend en compte la pseudo attraction de la région avec le blob mais intègre également les énergies des régions qui lui sont voisines. (c) Ici l'énergie bi-potentielle repose sur la définition d'une pseudo-attraction entre la deux régions voisines relativement à un blob

## 3.3 CONSTRUCTION D'UNE ÉNERGIE SPATIALE SOUS-ADDITIVE

Soit  $P = \{R_1, \dots, R_N\}$  une partition de l'image  $I$ . L'ensemble des parties sur  $\Omega$  sera désigné par  $\mathcal{P}(\Omega)$  tandis que l'ensemble des partitions sur le support de l'image  $\Omega$  sera noté  $\mathbb{P}(\Omega)$ .

La relation d'adjacence entre deux régions  $R_i$  et  $R_j$  sera notée :  $R_i \leftrightarrow R_j$ .

On notera  $V(R) = \{R' \in P - \{R\} / R' \leftrightarrow R\}$  l'ensemble des régions voisines de la région  $R$  dans la partition  $P$ .

### 3.3.1 Potentiel spatial relatif à un blob

**Définition 3.1** On appelle **potentiel sous-additif relatif à un blob  $\mathbf{B}$**  toute application  $\delta(., B) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  vérifiant les propriétés suivantes :

- $\forall R \in \mathcal{P}(\Omega), \delta(R, B) \geq 0,$
- $\forall R, R' \in \mathcal{P}(\Omega) \text{ t.q } R \leftrightarrow R', \delta(R \cup R', B) \leq \delta(R, B) + \delta(R', B).$

De même on peut définir les notions de **potentiels ultra-additif ou supra-additif relatif à un blob  $\mathbf{B}$**  :

**Définition 3.2** On appelle **potentiel ultra-additif (resp. supra-additif) relatif à un blob  $\mathbf{B}$**  toute application  $\delta(., B) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  vérifiant les propriétés suivantes :

- $\forall R \in \mathcal{P}(\Omega), \delta(R, B) \geq 0,$
- **Ultra-additivité** :  $\forall R, R' \in \mathcal{P}(\Omega) \text{ t.q } R \leftrightarrow R',$   

$$\delta(R \cup R', B) \leq \max(\delta(R, B), \delta(R', B)).$$
- **Supra-additivité** :  $\forall R, R' \in \mathcal{P}(\Omega) \text{ t.q } R \leftrightarrow R',$   

$$\delta(R \cup R', B) \leq \min(\delta(R, B), \delta(R', B)).$$

**Remarque** : Tout potentiel supra-additif est ultra-additif et tout potentiel ultra-additif est un potentiel sous-additif.

**Définition 3.3** On appelle **potentiel convexe relatif à un blob  $\mathbf{B}$**  toute application  $\delta(., B) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  vérifiant les propriétés suivantes :

- $\forall R \in \mathcal{P}(\Omega), \delta(R, B) \geq 0,$
- $\forall R_1, R_2 \in \mathcal{P}(\Omega) \text{ t.q } R_1 \leftrightarrow R_2, \exists \alpha_{1,2} \in ]0, 1[ \text{ tel que }$   

$$\delta(R_1 \cup R_2, B) \leq \alpha_{1,2} \delta(R_1, B) + (1 - \alpha_{1,2}) \delta(R_2, B).$$

**Remarque** : Tout potentiel convexe est ultra-additif et par conséquent sous-additif.

**Définition 3.4** Soit  $B$  un blob appartenant au support  $\Omega$  de l'image  $I$ . On appelle **un bi-potentiel sous-additif relativement au blob  $\mathbf{B}$**  toute application vérifiant les propriétés suivantes :

- **Domaine de définition** :

$$\begin{aligned} \delta_B : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) &\longrightarrow \mathbb{R}^+ \\ (R, R') &\longmapsto \delta_B(R, R'), \end{aligned}$$

- $\forall R, R' \in \mathcal{P}(\Omega), \delta_B(R, R') = 0 \text{ si } R \leftrightarrow R',$
- $\forall R, R' \in \mathcal{P}(\Omega), \delta_B(R, R') > 0 \text{ si } R \not\leftrightarrow R',$
- **Symétrie** :  $\forall R, R' \in \mathcal{P}(\Omega), \delta_B(R, R') = \delta_B(R', R),$
- **Sous-additivité** :  $\forall R, R_1, R_2 \in \mathcal{P}(\Omega) \text{ t.q } (R_1 \cup R_2) \leftrightarrow R,$

$$\delta_B(R_1 \cup R_2, R) \leq \delta_B(R_1, R) + \delta_B(R_2, R).$$

**Proposition 3.1** Soit  $\delta(., B)$  un potentiel sous-additif (resp. ultra-additif, supra-additif) relatif au blob  $B$  alors :

1.  $\delta_B(R, R') = \delta(R, B) + \delta(R', B)$  est sous-additive (resp. ultra-additive, supra-additive) relativement au blob  $B$ .
2.  $\delta_B(R, R') = \delta(R, B) \cdot \delta(R', B)$  est sous-additive (resp. ultra-additive, supra-additive) relativement au blob  $B$ .

**Preuve :** Nous nous contenterons de démontrer la stabilité de la sous-additivité par rapport à l'addition et la multiplication. La stabilité des deux autres formes d'additivité se démontre de la même manière.

Soit  $R_1$  et  $R_2$  deux régions adjacentes appartenant à une partition quelconque  $P$  de l'image  $I$ .

Par définition les deux bi-potentiels sont symétriques et positifs.

La sous-additivité du potentiel  $\delta(., B)$  permet d'écrire :

$$\delta(R_1 \cup R_2, B) \leq \delta(R_1, B) + \delta(R_2, B),$$

Soit  $R$  une région adjacente de la région  $R_1 \cup R_2$  on a :

$$\begin{aligned} 1) - \delta_B(R_1 \cup R_2, R) &= \delta(R_1 \cup R_2, B) + \delta(R, B) \\ &\leq \delta(R_1, B) + \delta(R_2, B) + \delta(R, B) \\ &\leq \delta(R_1, B) + \delta(R_2, B) + \delta(R, B) + \delta(R, B) \\ &\leq \delta_B(R_1, R) + \delta_B(R_2, R) \end{aligned}$$

$$\begin{aligned} 2) - \delta_B(R_1 \cup R_2, R) &= \delta(R_1 \cup R_2, B) \times \delta(R, B) \\ &\leq (\delta(R_1, B) + \delta(R_2, B)) \times \delta(R, B) \\ &\leq \delta(R_1, B) \times \delta(R, B) + \delta(R_2, B) \times \delta(R, B) \\ &\leq \delta_B(R_1, R) + \delta_B(R_2, R) \end{aligned}$$

**Proposition 3.2** Soit  $\delta(., B)$  un potentiel supra-additif relatif au blob  $B$  alors :

$$\delta_B^+(R, R') = \begin{cases} \delta(R, B) + \delta(R', B) & \text{si } R \leftrightarrow R', \\ 0 & \text{si } R \not\leftrightarrow R'. \end{cases}$$

et

$$\delta_B^\times(R, R') = \begin{cases} \delta(R, B) \times \delta(R', B) & \text{si } R \leftrightarrow R', \\ 0 & \text{si } R \not\leftrightarrow R'. \end{cases}$$

sont des bi-potentiels sous-additifs relativement au blob  $B$

**Preuve :** Par définition, les fonctions  $\delta_B^+$  et  $\delta_B^\times$  vérifient aisément toutes les propriétés que requièrent un bi-potentiel, à savoir la symétrie et la positivité. Il reste donc à vérifier la sous-additivité de ces deux applications. Soient  $R_1, R_2$  et  $R$  trois régions adjacentes.

i- Montrons la sous-additivité de  $\delta_B^+$  :

$$\delta_B^+(R_1 \cup R_2, R) = \begin{cases} \delta(R_1 \cup R_2, B) + \delta(R, B) & \text{si } R \leftrightarrow R_1 \cup R_2, \\ 0 & \text{sinon.} \end{cases}$$

or

$$R \leftrightarrow R_1 \cup R_2 \Leftrightarrow \begin{cases} R \leftrightarrow R_1 \quad \text{et} \quad R \leftrightarrow R_2 \dots (1) \\ R \leftrightarrow R_1 \quad \text{et} \quad R \leftrightarrow R_2 \dots (2) \\ R \leftrightarrow R_1 \quad \text{et} \quad R \leftrightarrow R_2 \dots (3) \end{cases}$$

et

$$R \Leftrightarrow R_1 \cup R_2 \Leftrightarrow R \Leftrightarrow R_1 \quad \text{et} \quad R \Leftrightarrow R_2 \dots \quad (4)$$

Nous avons trois cas à traiter puisque les cas (2) et (3) sont symétriques.

1. Si l'on est dans la situation (1) alors  $\delta_B^+(R_1, R) \neq 0$  et  $\delta_B^+(R_2, R) \neq 0$

$\Downarrow$

$$\begin{aligned} \delta_B^+(R_1, R) + \delta_B^+(R_2, R) &= \delta(R_1, B) + \delta(R, B) + \delta(R_2, B) + \delta(R, B) \\ &\geq \underbrace{\delta(R_1, B) + \delta(R_2, B)}_{\geq} + \delta(R, B) \\ &\geq \delta(R_1 \cup R_2, B) + \delta(R, B) \quad (\delta(\cdot, B) \text{ est aussi sous-additive}) \\ &\geq \delta_B^+(R_1 \cup R_2, R). \end{aligned}$$

2. Si l'on est dans la situation (3) alors

$$\begin{aligned} \delta_B^+(R_1, R) + \delta_B^+(R_2, R) &= \delta^+(R_1, R') \\ &= \underbrace{\delta(R_1, B)}_{\geq} + \delta(R, B) \\ &\geq \delta(R_1 \cup R_2, B) + \delta(R, B) \quad \text{du fait de la supra-add. de } \delta(\cdot, B) \\ &\geq \delta_B^+(R_1 \cup R_2, R). \end{aligned}$$

3. La situation (4) est triviale puisque dans ce cas :

$$\delta_B^+(R_1, R) = \delta_B^+(R_2, R) = \delta_B^+(R_1 \cup R_2, R) = 0.$$

Il en résulte que  $\delta_B^+$  est sous-additive.

**ii-** La sous-additivité de  $\delta_B^\times$  se démontre de la même manière que celle de la sous-additivité de  $\delta_B^+$ .

**Proposition 3.3** Soit  $\delta_s(\cdot, B)$  un potentiel supra-additif et soit  $\delta(\cdot, B)$  un potentiel sous-additif (resp. ultra-additif, convexe) alors l'application  $(\delta_s \times \delta)(\cdot, B)$  est également un potentiel sous-additif (resp. ultra-additif, convexe).

**Preuve :** Nous allons uniquement démontrer que le produit d'un potentiel convexe par un potentiel supra-additif est un potentiel convexe. Le reste des démonstrations se fait de la même manière.

Soient  $R_1$  et  $R_2$  deux régions adjacentes. On a :

$$\delta_s(R_1 \cup R_2, B) \leq \min(\delta_s(R_1, B), \delta_s(R_2, B)) \Rightarrow \begin{cases} \delta_s(R_1 \cup R_2, B) \leq \delta_s(R_1, B), \\ \delta_s(R_1 \cup R_2, B) \leq \delta_s(R_2, B), \end{cases}$$

d'autre part, la convexité de  $\delta(\cdot, B)$  permet d'écrire :

$$\delta(R_1 \cup R_2, B) \leq \alpha \delta(R_1, B) + (1 - \alpha) \delta(R_2, B),$$

tel que  $\alpha \in [0, 1]$ .

$$\begin{aligned}
(\delta_s \times \delta)(R_1 \cup R_2, B) &= \delta_s(R_1 \cup R_2, B) \times \delta(R_1 \cup R_2, B) \\
&\leq \delta_s(R_1 \cup R_2, B)(\alpha \delta(R_1, B) + (1 - \alpha) \delta(R_2, B)) \\
&\leq \underbrace{\alpha \delta_s(R_1 \cup R_2, B)}_{\leq} \delta(R_1, B) + \underbrace{(1 - \alpha) \delta_s(R_1 \cup R_2, B)}_{\leq} \delta(R_2, B) \\
&\leq \alpha \delta_s(R_1, B) \delta(R_1, B) + (1 - \alpha) \delta_s(R_2, B) \delta(R_2, B) \\
&\leq \alpha (\delta_s \times \delta)(R_1, B) + (1 - \alpha) (\delta_s \times \delta)(R_2, B)
\end{aligned}$$

ce qui montre que  $(\delta_s \times \delta)(., B)$  demeure convexe.

### 3.3.2 Energie spatiale relative à un blob

Après avoir défini les différents potentiels relatifs à un blob, nous allons maintenant introduire les définitions des différentes énergies spatiales relatives à un blob.

#### 3.3.2.1 Energie bi-potentielle relative à un blob :

Nous avons défini précédemment la notion de bi-potentiel sous-additif entre deux régions voisines, relativement à un blob  $B$ . L'énergie bi-potentielle d'une région est défini par la somme des bi-potentiels entre cette région et ses régions voisines comme le résume la définition ci-dessous :

**Définition 3.5** Soit  $B$  un blob appartenant au support  $\Omega$  de l'image  $I$  et  $P$  une partition quelconque. Soit  $\delta_B$  un bi-potentiel sous-additif relatif au blob  $B$ . On appelle **énergie bi-potentielle relativement au blob  $B$**  toute application  $D(., B) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  définie par la relation suivante :

$$\forall R \in \mathcal{P}(\Omega), \quad D(R, B) = \sum_{R' \in V(R)} \delta_B(R, R').$$

tel que  $V(R)$  représente le voisinage de la région  $R$ .

**Proposition 3.4** Soit  $P$  une partition de l'image  $I$  et  $\delta_B$  un bi-potentiel sous-additif relatif au blob  $B$ . Toute énergie bi-potentielle relative à un blob  $B$  est sous-additive, i.e :

$$\forall R_1, R_2 \in P \text{ t.q } R_1 \leftrightarrow R_2, \quad D(R_1 \cup R_2, B) \leq D(R_1, B) + D(R_2, B).$$

#### Preuve :

Soit  $R_1$  et  $R_2$  deux régions adjacentes appartenant à la partition  $P$ , munies respectivement de leurs ensembles de voisinages  $V(R_1)$  et  $V(R_2)$ .

On désignera par  $W(R_1, R_2) = V(R_1) \cap V(R_2)$  l'ensemble des régions voisines à la fois de la région  $R_1$  et de la région  $R_2$ . On notera aussi par :

$$W(R_1) = V(R_1) - \{W(R_1, R_2) \cup \{R_2\}\},$$

l'ensemble des régions exclusivement adjacentes à la région  $R_1$  et ne contenant pas la région  $R_2$ .

De même que pour la région  $R_2$  :

$$W(R_2) = V(R_2) - \{W(R_1, R_2) \cup \{R_1\}\}.$$

On peut alors écrire :

$$V(R_1) = W(R_1) \cup W(R_1, R_2) \cup \{R_2\},$$

$$V(R_2) = W(R_2) \cup W(R_1, R_2) \cup \{R_1\}.$$

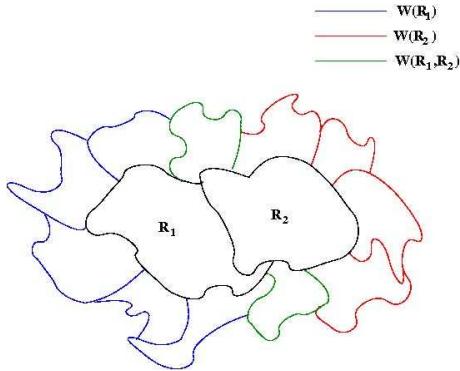


FIGURE 3.6 – Décomposition du voisinage de deux régions adjacentes.

On a alors :

$$\begin{aligned} D(R_1, B) + D(R_2, B) &= \sum_{R \in V(R_1)} \delta_B(R_1, R) + \sum_{R \in V(R_2)} \delta_B(R_2, R) \\ &= \sum_{R \in W(R_1)} \delta_B(R_1, R) + \sum_{R \in W(R_1, R_2)} \delta_B(R_1, R) + \delta_B(R_1, R_2) \\ &\quad + \sum_{R \in W(R_2)} \delta_B(R_2, R) + \sum_{R \in W(R_1, R_2)} \delta_B(R_2, R) + \delta_B(R_2, R_1) \\ &= \sum_{R \in W(R_1)} [\delta_B(R_1, R) + \delta_B(R_2, R)] + \sum_{R \in W(R_2)} [\delta_B(R_2, R) + \delta_B(R_1, R)] \\ &\quad + \sum_{R \in W(R_1, R_2)} [\delta_B(R_1, R) + \delta_B(R_2, R)] + 2 \cdot \delta_B(R_1, R_2) \\ &\quad - \sum_{R \in W(R_1)} \delta_B(R_2, R) - \sum_{R \in W(R_2)} \delta_B(R_1, R). \end{aligned}$$

Puisque  $\delta_B$  est un bi-potentiel sous-additif on a :

$$\delta_B(R_1, R) + \delta_B(R_2, R) \geq \delta_B(R_1 \cup R_2, R).$$

Ceci implique que :

$$\begin{aligned} D(R_1, B) + D(R_2, B) &\geq \sum_{R \in W(R_1)} \delta_B(R_1 \cup R_2, R) + \sum_{R \in W(R_2)} \delta_B(R_1 \cup R_2, R) \\ &\quad + \sum_{R \in W(R_1, R_2)} \delta_B(R_1 \cup R_2, R) + 2 \cdot \delta_B(R_1, R_2) \\ &\quad - \sum_{R \in W(R_1)} \delta_B(R_2, R) - \sum_{R \in W(R_2)} \delta_B(R_1, R). \\ &\geq \sum_{R \in W(R_1) \cup W(R_2) \cup W(R_1, R_2)} \delta_B(R_1 \cup R_2, R) + 2 \cdot \delta_B(R_1, R_2) \\ &\quad - \sum_{R \in W(R_1)} \delta_B(R_2, R) - \sum_{R \in W(R_2)} \delta_B(R_1, R). \end{aligned}$$

Or,  $V(R_1 \cup R_2) = W(R_1) \cup W(R_2) \cup W(R_1, R_2)$  donc :

$$\begin{aligned} D(R_1, B) + D(R_2, B) &\geq \sum_{R \in V(R_1 \cup R_2)} \delta_B(R_1 \cup R_2, R) + 2 \cdot \delta_B(R_1, R_2) \\ &\quad - \sum_{R \in W(R_1)} \delta_B(R_2, R) - \sum_{R \in W(R_2)} \delta_B(R_1, R). \end{aligned}$$

Par définition d'un bi-potentiel sous-additif on a :

$$\begin{aligned} \forall R, R' \text{ tel que } R \leftrightarrow R', \quad \delta_B(R, R') = 0, \\ \Rightarrow \quad \sum_{R \in W(R_1)} \delta_B(R_2, R) = 0 \quad \text{et} \quad \sum_{R \in W(R_2)} \delta_B(R_1, R) = 0. \end{aligned}$$

On obtient finalement :

$$\begin{aligned} D(R_1, B) + D(R_2, B) &\geq D(R_1 \cup R_2, B) + 2 \cdot \delta_B(R_1, R_2) \\ &\geq D(R_1 \cup R_2, B). \quad [\text{c.q.f.d}] \end{aligned}$$

Par conséquent l'énergie bi-potentielle  $D(., B)$  est sous-additive.

### 3.3.2.2 Energie multi/mono-potentielle relative à un blob :

Nous pouvons également définir une énergie spatiale relative à un blob  $B$  en utilisant simplement la notion de potentiel sous-additif. L'énergie sera qualifiée de *multi-potentielle* si elle prend en compte les potentiels des régions voisines. L'énergie de la région est dite *mono-potentielle* si le voisinage de celle-ci est exclu.

**Définition 3.6** Soit  $B$  un blob appartenant au support  $\Omega$  de l'image  $I$  et  $P$  une partition quelconque. Soit  $\delta(., R)$  un potentiel sous-additif relatif au blob  $B$ . On appelle **énergie multi-potentielle relativement au blob  $B$**  toute application  $D(., B) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  définie par la relation suivante :

$$\forall R \in \mathcal{P}(\Omega), \quad D(R, B) = \sum_{R' \in T(R)} \delta(R', B).$$

tel que  $T(R) = V(R) \cup \{R\}$  englobe le voisinage de la région  $R$  ainsi que la région  $R$  elle-même. Dans le cas où  $T(R) = \{R\}$  on dira que l'énergie est **mono-potentielle**.

**Proposition 3.5** Toute énergie multi-potentielle (ou mono-potentielle) relative à un blob  $B$  est sous-additive.

**Preuve :**

1- Si  $T(R) = \{R\}$  alors l'énergie mono-potentielle n'est autre que le potentiel sous-additif lui-même. Elle est par conséquent sous-additive.

$$\forall R \in \mathcal{P}(\Omega), \quad D(R, B) = \delta(R, B).$$

2- Soit  $D(., B)$  une énergie multi-potentielle relativement au blob  $B$ . Soit  $R$  une région quelconque de la partition  $P$  et soit  $T(R) = V(R) \cup \{R\}$ , on a alors par définition :

$$D(R, B) = \sum_{R' \in T(R)} \delta(R', B)$$

tel que  $\delta(., B)$  est un potentiel sous-additif.

Soit  $R_1, R_2$  deux régions adjacentes appartenant à la partition  $P$  :

$$\begin{aligned} D(R_1 \cup R_2, B) &= \sum_{R \in T(R_1 \cup R_2)} \delta(R, B) \\ &= \delta(R_1 \cup R_2, B) + \sum_{R \in V(R_1 \cup R_2)} \delta(R, B) \end{aligned}$$

Or

$$V(R_1 \cup R_2) \subseteq V(R_1) \cup V(R_2)$$

$$\begin{aligned} D(R_1 \cup R_2, B) &\leq \delta(R_1 \cup R_2, B) + \sum_{R \in V(R_1) \cup V(R_2)} \delta(R, B) \\ &\leq \delta(R_1 \cup R_2, B) + \sum_{R \in V(R_1)} \delta(R, B) + \sum_{R \in V(R_2)} \delta(R, B) \\ &\leq \delta(R_1, B) + \delta(R_2, B) + \sum_{R \in V(R_1)} \delta(R, B) + \sum_{R \in V(R_2)} \delta(R, B) \\ &\leq \sum_{R \in T(R_1)} \delta(R, B) + \sum_{R \in T(R_2)} \delta(R, B) \\ &\leq D(R_1, B) + D(R_2, B). \quad [\text{c.q.f.d}] \end{aligned}$$

### 3.3.2.3 Convexité d'une énergie spatiale relative à un blob :

**Définition 3.7** Soit  $D(., B)$  une énergie relative au blob  $B$ . On dit que  $D(., B)$  est une énergie convexe si et seulement si :

$$\forall R_1, R_2 \text{ t.q } R_1 \leftrightarrow R_2, \exists \alpha_{1,2} \in [0, 1], \quad D(R_1 \cup R_2, B) \leq \alpha_{1,2} D(R_1, B) + (1 - \alpha_{1,2}) D(R_2, B).$$

La constante  $\alpha_{1,2}$  dépend des régions  $R_1$  et  $R_2$ .

**Proposition 3.6** Toute énergie, relative à un blob  $B$ , convexe est **ultra-additive** et par conséquent **sous-additive**.

**Preuve :** L'ultra-additivité découle directement de la propriété de convexité de l'énergie. Sans expliciter en détail la démonstration, la fonction :

$$f(\alpha) = \alpha D(R_1, B) + (1 - \alpha) D(R_2, B),$$

est continue et bijective sur l'intervalle  $]0, 1[$ . Elle prend ses valeurs dans  $[\min(D(R_1, B), D(R_2, B)), \max(D(R_1, B), D(R_2, B))]$ . En outre, pour  $\alpha_{1,2} \in ]0, 1[$ , on a :  $f(\alpha_{1,2}) \leq \max(D(R_1, B), D(R_2, B))$  **[c.q.f.d.]**.

Quand à la sous-additivité, elle résulte de l'ultra-additivité de l'énergie, puisque toute fonction ultra-additive est sous-additive.

### 3.3.3 Energie d'*a priori* spatial

Nous allons maintenant définir les notions d'énergies spatiales bi/mono/multi-potentielle d'une région de l'image relativement à l'ensemble des blobs  $\mathcal{B}$ , i.e, relativement à l'image-blob  $I_{\mathcal{B}}$ . Pour cela, nous considérons le vecteur  $\overrightarrow{D_{\mathcal{B}}(R, .)}$  caractérisant les différentes interactions entre la région  $R$  et les blobs.

**Définition 3.8** Soit  $\mathcal{B} = \{B_1, \dots, B_M\}$  l'ensemble des blobs issus de l'image  $I$  et soient  $D(\cdot, B_k)$ ,  $k = 1, \dots, M$ , les énergies bi-potentielles sous-additives relatives aux blobs  $\{B_k\}_{k=1, \dots, M}$ . Soit  $\|\cdot\|$  la norme issue du produit scalaire  $\langle \cdot, \cdot \rangle$ , défini sur l'espace  $\mathbb{R}^M$ . On appelle **énergie spatiale bi-potentielle** toute application  $D_{\mathcal{B}} : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  définie par :

$$\forall R \in \mathcal{P}(\Omega), \quad D_{\mathcal{B}}(R) = \|\overrightarrow{D_{\mathcal{B}}(R, \cdot)}\| = \sqrt{\langle \overrightarrow{D_{\mathcal{B}}(R, \cdot)}, \overrightarrow{D_{\mathcal{B}}(R, \cdot)} \rangle},$$

tel que  $\overrightarrow{D_{\mathcal{B}}(R, \cdot)} = \{D(R, B_k)\}_{k=1, \dots, M}$ .

**Définition 3.9** Soit  $\mathcal{B} = \{B_1, \dots, B_M\}$  l'ensemble des blobs issus de l'image  $I$  et soient  $D(\cdot, B_k)$ ,  $k = 1, \dots, M$ , les énergies multi/mono-potentielles sous-additives relatives aux blobs  $\{B_k\}_{k=1, \dots, M}$ . Soit  $\|\cdot\|$  la norme issue du produit scalaire  $\langle \cdot, \cdot \rangle$ , défini sur l'espace  $\mathbb{R}^M$ . On appelle **énergie spatiale multi/mono-potentielle** toute application  $D_{\mathcal{B}} : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  définie par :

$$\forall R \in \mathcal{P}(\Omega), \quad D_{\mathcal{B}}(R) = \|\overrightarrow{D_{\mathcal{B}}(R, \cdot)}\| = \sqrt{\langle \overrightarrow{D_{\mathcal{B}}(R, \cdot)}, \overrightarrow{D_{\mathcal{B}}(R, \cdot)} \rangle},$$

tel que  $\overrightarrow{D_{\mathcal{B}}(R, \cdot)} = \{D(R, B_k)\}_{k=1, \dots, M}$ .

Ainsi, l'énergie d'une région relativement à l'ensemble des blobs représente l'intensité du vecteur potentiel de la région relativement aux blobs  $B_1, \dots, B_M$ . Cependant, la sous-additivité des énergies  $D(R, B_k)$ ,  $k = 1, \dots, M$  (qu'ils soient bi, mono ou multi potentielles) entraîne-t-elle pour autant la sous-additivité de la norme du vecteur  $\overrightarrow{D_{\mathcal{B}}(R, \cdot)}$  ?

Les deux propositions étendent la définition de l'énergie spatiale  $D_{\mathcal{B}}$  à la norme  $\mathcal{L}^p$ <sup>4</sup>. Elles posent les conditions supplémentaires pour que l'énergie  $D_{\mathcal{B}}$  soit sous-additive :

**Proposition 3.7** Si  $D(\cdot, B)$  est une énergie bi/mono/multi-potentielle convexe alors l'énergie spatiale «  $D_{\mathcal{B}}$  » issue de cette énergie et définie par

$$D_{\mathcal{B}}(R) = \|\overrightarrow{D_{\mathcal{B}}(R, \cdot)}\|_p^p$$

est convexe pour  $p \geq 1$  et par conséquent sous-additive.

**Preuve :** Soit  $\mathcal{B}$  l'ensemble des blobs de l'image  $I$ . Soit  $R_1$  et  $R_2$  deux régions adjacentes.

$$\forall k = 1, \dots, M, \exists \alpha_{1,2} \in ]0, 1[, D(R_1 \cup R_2, B_k) \leq \alpha_{1,2} D(R_1, B_k) + (1 - \alpha_{1,2}) D(R_2, B_k),$$

car  $D(\cdot, B_k)$  est sous-additive pour tout  $k$  du fait de la convexité de l'énergie loco-spatiale relative à chaque blob.

<sup>4</sup>

$\|\vec{x}\|_p^p = \sum_{k=1, \dots, M} x_k^p$ , pour  $x \in \mathbb{R}^M$  et  $p \geq 1$ .

On a alors :

$$\begin{aligned}
 D_{\mathcal{B}}(R_1 \cup R_2) &= \|\overrightarrow{D_{\mathcal{B}}(R_1 \cup R_2, .)}\|_p^p \\
 &\leq \alpha_{1,2} \|\overrightarrow{D_{\mathcal{B}}(R_1, .)}\|_p^p + (1 - \alpha_{1,2}) \|\overrightarrow{D_{\mathcal{B}}(R_2, .)}\|_p^p (*) \\
 &\leq \|\overrightarrow{D_{\mathcal{B}}(R_1, .)}\|_p^p + \|\overrightarrow{D_{\mathcal{B}}(R_2, .)}\|_p^p \\
 &= D_{\mathcal{B}}(R_1) + D_{\mathcal{B}}(R_2). \quad [\text{c.q.f.d}]
 \end{aligned}$$

\* : provient de la convexité de la fonction  $x^p$ . Ainsi la convexité de l'énergie relative à un blob, plus contraignante que la sous-additivité, permet d'obtenir une énergie spatiale  $D_{\mathcal{B}}$  sous-additive et ce pour tout  $p \geq 1$ .

**Proposition 3.8** Si  $D(., B)$  est une énergie bi/mono/multi-potentielle sous-additive alors l'énergie spatiale «  $D_{\mathcal{B}}$  » issue de cette énergie et définie par

$$D_{\mathcal{B}}(R) = \|\overrightarrow{D_{\mathcal{B}}(R, .)}\|_p$$

est sous-additive pour  $p = 1, +\infty$ .

**Preuve :** La démonstration est aisée et nous ne nous y attarderons pas.

La précédente proposition indique que l'énergie spatiale  $D_{\mathcal{B}}$ , issue d'une énergie spatiale relative à un blob sous-additive, est sous-additive uniquement pour  $p = 1$  ou  $p = +\infty$ .

Nous disposant maintenant d'une énergie spatiale  $D_{\mathcal{B}}$  sous-additive, il devient alors aisés de définir une énergie spatiale sous-additive sur **l'espace des partitions**,  $\mathbb{P}(\Omega)$ , de l'image  $I$ .

**Définition 3.10** Soit  $\mathcal{B} = \{B_1, \dots, B_M\}$  l'ensemble des blobs issus de l'image  $I$  et soit  $D_{\mathcal{B}}$  une énergie spatiale (mono, multi ou bi-potentielle). Alors l'énergie  $D_{\mathcal{B}} : \mathbb{P}(\Omega) \longrightarrow \mathbb{R}^+$  définie par :

$$D_{\mathcal{B}}(P) = \sum_{R \in P} D_{\mathcal{B}}(R),$$

est une énergie séparable et sous-additive par construction.

### 3.4 CONSTRUCTION DE POTENTIELS SOUS, ULTRA ET/OU SUPRA-ADDITIONS

Nous allons maintenant nous intéresser aux différentes énergies spatiales sous-additives que l'on peut construire à partir de la définition d'un potentiel sous/ultra/supra-additif ou convexe relatif à un blob  $B$ .

Comme nous l'avons vu précédemment les différentes énergies spatiales définies ci-dessous (mono, multi ou bi-potentielle) reposent toutes sur la définition d'un potentiel sous, ultra ou supra-additif ou la définition d'un potentiel convexe.

#### 3.4.1 Potentiel supra-additif

La condition de supra-additivité est la plus contraignante au regard des autres, à savoir la sous-additivité, la convexité et l'ultra-additivité. Cependant, elle est une

condition nécessaire pour la construction d'une énergie bi-potentielle relativement à un blob  $B$ .

Nous commencerons par énoncer une condition suffisante pour qu'un potentiel soit supra-additif.

**Proposition 3.9** *Tout potentiel  $\delta(., B)$  décroissant au sens de l'ordre d'inclusion est un potentiel supra-additif.*

**Preuve :** Soient  $R_1$  et  $R_2$  deux régions adjacentes et disjointes d'une partition  $P$  quelconque. Si  $\delta(., B)$  est un potentiel décroissant par rapport à l'ordre d'inclusion alors :

$$\begin{cases} \delta(R_1 \cup R_2, B) \leq \delta(R_1, B) \\ \delta(R_1 \cup R_2, B) \leq \delta(R_2, B) \end{cases} \Rightarrow \delta(R_1 \cup R_2, B) \leq \min(\delta(R_1, B), \delta(R_2, B)).$$

ce qui montre la supra-additivité de  $\delta(., B)$  [c.q.f.d].

Nous allons dans un premier temps définir un potentiel supra-additif assez trivial, qui n'inclut aucune notion de distance spatiale entre régions et blobs.

$$\delta_s(R, B) = \frac{\beta}{\gamma_0 + \gamma_1 |R \cap B|},$$

tel que  $\beta$ ,  $\gamma_0$  et  $\gamma_1$  sont des constantes positives.

La décroissance de cette fonction assure sa supra-additivité. Par ailleurs ce potentiel est borné, i.e. :

$$0 < \frac{\beta}{\gamma_0 + \gamma_1 \min(|R|, |B|)} \leq \delta_s(R, B) \leq \frac{\beta}{\gamma_0}.$$

Nous allons maintenant définir deux distances spatiales entre régions et blobs directement inspirées des distances entre clusters et couramment utilisées dans la classification hiérarchique ascendante.

**Définition 3.11** *Soit  $d$  une distance métrique définie sur le support  $\Omega$  de l'image  $I$  et  $c_B$  le centre de gravité d'un blob  $B$ .*

*On appellera distance de lien minimum relativement au blob  $B$  la distance définie par la relation suivante :*

$$d_{min}(R, B) = \min_{x \in R} \{d(x, c_B)\},$$

*On appellera distance de lien complet relativement au blob  $B$  la distance définie par la relation suivante :*

$$d_{max}(R, B) = \max_{x \in R} \{d(x, c_B)\}.$$

**Proposition 3.10** *Pour  $R_1$  et  $R_2$  deux régions adjacentes et disjointes, on a les propriétés suivantes :*

- $d_{min}(R_1 \cup R_2, B) = \min(d_{min}(R_1, B), d_{min}(R_2, B))$ ,
- $d_{max}(R_1 \cup R_2, B) = \max(d_{max}(R_1, B), d_{max}(R_2, B))$ .

**Preuve :** Soient  $R_1$  et  $R_2$  deux régions adjacentes et disjointes. Soit  $B$  un blob quelconque de l'image  $I$ .

$$\begin{aligned} d_{min}(R_1 \cup R_2, B) &= \min_{x \in R_1 \cup R_2} \{d(x, c_B)\} \\ &= \min(\min_{x \in R_1} \{d(x, c_B)\}, \min_{x \in R_2} \{d(x, c_B)\}) \\ &= \min(d_{min}(R_1, B), d_{min}(R_2, B)) \end{aligned}$$

Pour la deuxième propriété, on a :

$$\begin{aligned} d_{\max}(R_1 \cup R_2, B) &= \max_{x \in R_1 \cup R_2} \{d(x, c_B)\} \\ &= \max(\max_{x \in R_1} \{d(x, c_B)\}, \max_{x \in R_2} \{d(x, c_B)\}) \\ &= \max(d_{\max}(R_1, B), d_{\max}(R_2, B)) \end{aligned}$$

**Proposition 3.11** Soit  $\delta_s(., B)$  un potentiel supra-additif relativement au blob  $B$  et soient  $d_{\min}$  et  $d_{\max}$  respectivement les distances du lien minimum et du lien maximum relatives au blob  $B$  et définies à partir d'une distance métrique  $d$ , alors :

- $\delta_{\min}(., B) = \delta_s(., B)d_{\min}(., B)$  est un potentiel supra-additif dit de lien minimum et
- $\delta_{\max}(., B) = \frac{\delta_s(., B)}{d_{\max}(., B)}$  est un potentiel supra-additif dit de lien complet.

**Preuve** : Soient donc  $R_1$  et  $R_2$  deux régions adjacentes et disjointes. Soit  $B$  un blob quelconque de l'image  $I$ .

$$\begin{aligned} \delta_{\min}(R_1 \cup R_2, B) &= \delta_s(R_1 \cup R_2, B)d_{\min}(R_1 \cup R_2, B) \\ &= \delta_s(R_1 \cup R_2, B) \min(d_{\min}(R_1, B), d_{\min}(R_2, B)) \\ &= \min(\delta_s(R_1 \cup R_2, B)d_{\min}(R_1, B), \delta_s(R_1 \cup R_2, B)d_{\min}(R_2, B)) \end{aligned}$$

Or, la supra-additivité de  $\delta_s(., B)$  implique :

$$\delta_s(R_1 \cup R_2, B) \leq \delta_s(R_1, B) \text{ et } \delta_s(R_1 \cup R_2, B) \leq \delta_s(R_2, B)$$

Il en résulte que :

$$\begin{aligned} \delta_{\min}(R_1 \cup R_2, B) &\leq \min(\delta_s(R_1, B)d_{\min}(R_1, B), \delta_s(R_2, B)d_{\min}(R_2, B)) \\ &\leq \min(\delta_{\min}(R_1, B), \delta_{\min}(R_2, B)). \end{aligned}$$

De la même manière, on démontre la propriété de supra-additivité du potentiel de lien complet.

$$\begin{aligned} \delta_{\max}(R_1 \cup R_2, B) &= \frac{\delta_s(R_1 \cup R_2, B)}{d_{\max}(R_1 \cup R_2, B)} \\ &= \frac{\delta_s(R_1 \cup R_2, B)}{\max(d_{\max}(R_1, B), d_{\max}(R_2, B))} \\ &= \delta_s(R_1 \cup R_2, B) \min\left(\frac{1}{d_{\max}(R_1, B)}, \frac{1}{d_{\max}(R_2, B)}\right) \\ &= \min\left(\delta_s(R_1 \cup R_2, B) \frac{1}{d_{\max}(R_1, B)}, \delta_s(R_1 \cup R_2, B) \frac{1}{d_{\max}(R_2, B)}\right) \\ &= \min\left(\frac{\delta_s(R_1 \cup R_2, B)}{d_{\max}(R_1, B)}, \frac{\delta_s(R_1 \cup R_2, B)}{d_{\max}(R_2, B)}\right) \\ &\leq \min(\delta_{\max}(R_1, B), \delta_{\max}(R_2, B)). \end{aligned}$$

[c.q.f.d].

Ainsi, nous pouvons combiner un quelconque potentiel supra-additif  $\delta_s(., B)$  avec la distance spatiale (celle du lien minimum ou celle du lien complet) entre une région et un blob  $B$ . Plus particulièrement, le potentiel supra-additif cité ci-dessus pourrait être combiné avec la distance de lien minimum ou celle du lien maximum pour donner naissance à deux potentiels supra-additifs :

- $\delta_s^{min}(R, B) = \frac{\beta}{\gamma_0 + \gamma_1 |R \cap B|} d_{min}(R, B),$
- $\delta_s^{max}(R, B) = \frac{\beta}{(\gamma_0 + \gamma_1 |R \cap B|) d_{max}(R, B)}.$

D'autres potentiels supra-additifs peuvent être construits en se basant sur des fonctions autres que  $\frac{\beta}{\gamma_0 + \gamma_1 |R \cap B|}$ .

Nous donnons ici une liste qui n'est pas exhaustive :

- $\delta_s(R, B) = \frac{1}{|R||B|},$
- $\delta_s(R, B) = \frac{|B|}{|R|},$
- $\delta_s(R, B) = e^{-K \cdot \frac{|R|}{|B|}},$
- $\delta_s(R, B) = e^{-K \frac{|R \cap B|}{|B|}},$
- $\delta_s(R, B) = e^{K \cdot |B| - |R|}.$

L'existence de potentiels supra-additifs permet alors de garantir que les bi-potentiels  $\delta_B^+$  et  $\delta_B^\times$ , définis à travers les potentiels cités ci-dessus, seront sous-additifs (cf. la proposition 3.2). Ainsi, par construction, les énergies bi-potentielles spatiales seront par conséquent sous-additives (voir la proposition 3.3).

### 3.4.2 La convexité

Nous avons vu que la convexité d'un potentiel entraîne son ultra-additivité et par conséquent sa sous-additivité. Nous allons dans un premier temps nous intéresser aux potentiels dits « convexes » qui permettent aisément la construction d'énergies mono et/ou multi potentielles sous-additives.

Soit  $P$  une partition de l'image  $I$  et  $\mathcal{B}$  l'ensemble des blobs de cette image. Le premier potentiel trivial que l'on peut construire est celui qui représente la proportion d'une région dans un blob  $B$ .

$$\forall R \in P, \quad \delta_1(R, B) = \frac{|R \cap B|}{|R|}.$$

**Propriétés :**

1.  $\forall R \in P, \quad 0 \leq \delta_1(R, B) \leq 1,$
2. Si  $R \subseteq B$  alors  $\delta_1(R, B) = 1,$
3. Si  $R \cap B = \emptyset$  alors  $\delta_1(R, B) = 0,$
4.  $\forall R_i, R_j \in P \quad t.q \quad R_i \leftrightarrow R_j,$

$$\delta_1(R_i \cup R_j, B) = \frac{|R_i|}{|R_i| + |R_j|} \delta_1(R_i, B) + \frac{|R_j|}{|R_i| + |R_j|} \delta_1(R_j, B).$$

**Preuve :** triviale et nous ne nous y attarderons pas.

Il apparaît donc que ce potentiel est sous-additif du fait de sa convexité ( $\alpha = |R_i| / (|R_i| + |R_j|)$ ). On peut alors construire une énergie spatiale mono ou multi potentielle sous-additive.

Cette énergie ne comportera aucune information spatiale en termes de distance. Elle prendra en compte uniquement les intersections entre les régions et les blobs.

Nous allons maintenant définir un potentiel sous-additif qui prend en compte la masse de la région (exprimée par sa surface) et la distance métrique entre la région et le blob.

**Définition 3.12** Soit  $d$  une distance métrique sur  $\Omega \subseteq \mathbb{R}^2$  issue d'un produit scalaire  $\langle \cdot, \cdot \rangle$ . On appelle **potentiel attractif relativement au blob**  $B$  la fonction définie par :

$$\forall R \in \mathcal{P}(\Omega), \quad \delta_2(R, B) = \beta \frac{d^2(R, B)}{|R| \cdot |B|},$$

où  $\beta$  est une constante qui dépend du blob  $B$  et  $d(R, B) = d(c_R, c_B)$  représente la distance entre les centres de gravité de la région  $R$  et du blob  $B$ .

**Proposition 3.12** Le potentiel attractif relatif au blob  $B$  est convexe et par conséquent sous-additif.

**Lemme 3.1** Soit  $d$  une distance métrique sur  $\Omega \subseteq \mathbb{R}^2$  issue d'un produit scalaire  $\langle \cdot, \cdot \rangle$  alors :

$$\forall R_i, R_j \quad t.q \quad R_i \leftrightarrow R_j,$$

$$d^2(R_i \cup R_j, B) = \alpha d^2(R_i, B) + (1 - \alpha) d^2(R_j, B) - \alpha(1 - \alpha) d^2(R_i, R_j),$$

$$\text{avec } \alpha = \frac{|R_i|}{|R_i| + |R_j|}.$$

**Preuve du lemme 3.1 :**

Désignons par  $c_{R_i}$ ,  $c_{R_j}$ ,  $c_{R_i \cup R_j}$  et  $c_B$  les centres de gravité respectifs des régions  $R_i$ ,  $R_j$ ,  $R_i \cup R_j$  et du blob  $B$ . Notons aussi par

$$\alpha = \frac{|R_i|}{|R_i| + |R_j|} \Rightarrow 0 \leq \alpha \leq 1.$$

La relation entre le barycentre de l'union de deux régions disjointes et les barycentres des régions avant l'union est :

$$c_{R_i \cup R_j} = \alpha \cdot c_{R_i} + (1 - \alpha) \cdot c_{R_j}, \quad (1)$$

On sait que :

$$\begin{aligned} d^2(R_i, B) &= d^2(c_{R_i}, c_B) = \|c_{R_i} - c_B\|^2 = \langle c_{R_i} - c_B, c_{R_i} - c_B \rangle \\ &= \langle c_{R_i}, c_{R_i} \rangle + \langle c_B, c_B \rangle - 2 \langle c_{R_i}, c_B \rangle \\ &= \|c_{R_i}\|^2 + \|c_B\|^2 - 2 \langle c_{R_i}, c_B \rangle. \end{aligned}$$

Il s'en suit que :

$$\begin{aligned} d^2(R_i \cup R_j, B) &= \langle c_{R_i \cup R_j} - c_B, c_{R_i \cup R_j} - c_B \rangle \\ &= \alpha^2 \langle c_{R_i}, c_{R_i} \rangle + (1 - \alpha)^2 \langle c_{R_j}, c_{R_j} \rangle \\ &\quad + 2\alpha(1 - \alpha) \langle c_{R_i}, c_{R_j} \rangle + \langle c_B, c_B \rangle \\ &\quad - 2\alpha \langle c_{R_i}, c_B \rangle - 2(1 - \alpha) \langle c_{R_j}, c_B \rangle \\ &= \alpha^2 \|c_{R_i}\|^2 + (1 - \alpha)^2 \|c_{R_j}\|^2 + 2\alpha(1 - \alpha) \langle c_{R_i}, c_{R_j} \rangle \\ &\quad + \|c_B\|^2 - 2\alpha \langle c_{R_i}, c_B \rangle - 2(1 - \alpha) \langle c_{R_j}, c_B \rangle \end{aligned}$$

En ajoutant et en soustrayant  $\alpha(\|c_{R_i}\|^2 + \|c_B\|^2)$  et  $(1 - \alpha)(\|c_{R_j}\|^2 + \|c_B\|^2)$ , on obtient :

$$\begin{aligned} d^2(R_i \cup R_j, B) &= \alpha^2 \|c_{R_i}\|^2 + (1 - \alpha)^2 \|c_{R_j}\|^2 + 2\alpha(1 - \alpha) \langle c_{R_i}, c_{R_j} \rangle \\ &\quad + \|c_B\|^2 + \alpha d^2(c_{R_i}, c_B) - \alpha \|c_{R_i}\|^2 - \alpha \|c_B\|^2 \\ &\quad + (1 - \alpha)d^2(c_{R_j}, c_B) - (1 - \alpha)\|c_{R_j}\|^2 - (1 - \alpha)\|c_B\|^2 \\ &= \alpha d^2(c_{R_i}, c_B) + (1 - \alpha)d^2(c_{R_j}, c_B) \\ &\quad - \alpha(1 - \alpha)\|c_{R_i}\|^2 - \alpha(1 - \alpha)\|c_{R_j}\|^2 + 2\alpha(1 - \alpha) \langle c_{R_i}, c_{R_j} \rangle \\ &= \alpha d^2(R_i, B) + (1 - \alpha)d^2(R_j, B) \\ &\quad - \alpha(1 - \alpha)(\|c_{R_i}\|^2 + \|c_{R_j}\|^2 - 2 \langle c_{R_i}, c_{R_j} \rangle), \end{aligned}$$

Par conséquent :

$$d^2(R_i \cup R_j, B) = \alpha d^2(R_i, B) + (1 - \alpha)d^2(R_j, B) - \alpha(1 - \alpha)d^2(R_i, R_j). \text{ [c.q.f.d]}$$

**Remarque :**

$$\begin{aligned} d^2(R_i \cup R_j, B) &= \alpha d^2(R_i, B) + (1 - \alpha)d^2(R_j, B) - \alpha(1 - \alpha)d^2(R_i, R_j) \\ &\leq \alpha d^2(R_i, B) + (1 - \alpha)d^2(R_j, B) \\ &\leq d^2(R_i, B) + d^2(R_j, B), \end{aligned}$$

Par conséquent, une distance  $d$  issue d'un produit scalaire est convexe et donc sous-additive.

**Preuve de la proposition 3.7 :** Soient  $R_i, R_j$  deux régions disjointes et adjacentes et  $B$  un blob de l'image.  $d$  une distance métrique issue d'un produit scalaire.

$$\begin{aligned} \delta_2(R_i \cup R_j, B) &= \beta \frac{d^2(R_i \cup R_j, B)}{|R_i \cup R_j| |B|} \\ &= \beta \frac{\alpha d^2(R_i, B) + (1 - \alpha)d^2(R_j, B) - \alpha(1 - \alpha)d^2(R_i, R_j)}{|R_i| |B| + |R_j| |B|} \\ &= \frac{|R_i| |B| \alpha \delta_2(R_i, B) + (1 - \alpha) |R_j| |B| \delta_2(R_j, B) - \beta \alpha(1 - \alpha)d^2(R_i, R_j)}{|R_i| |B| + |R_j| |B|} \\ &= \alpha \frac{|R_i|}{|R_i| + |R_j|} \delta_2(R_i, B) + (1 - \alpha) \frac{|R_j|}{|R_i| + |R_j|} \delta_2(R_j, B) \\ &\quad - \frac{\beta \alpha(1 - \alpha)}{|B|(|R_i| + |R_j|)} d^2(R_i, R_j) \\ &= \underbrace{\alpha^2}_{\leq \alpha} \delta_2(R_i, B) + \underbrace{(1 - \alpha)^2}_{\leq 1 - \alpha} \delta_2(R_j, B) - \underbrace{\frac{\beta \alpha(1 - \alpha)}{|B|(|R_i| + |R_j|)} d^2(R_i, R_j)}_{\geq 0} \\ &\leq \alpha \delta_2(R_i, B) + (1 - \alpha) \delta_2(R_j, B), \end{aligned}$$

$\Rightarrow \delta_2(., B)$  est convexe et par conséquent sous-additive [c.q.f.d].

Il apparaît donc que le potentiel attractif relatif  $\delta_2(., B)$  est sous-additif et peut par conséquent servir à la construction d'une énergie spatiale mono ou multi potentielle.

A l'inverse du premier potentiel  $\delta_1(., B)$ , le potentiel attractif inclut une distance spatiale entre la région et le blob  $B$ . Il se présente sous la forme de l'inverse de la force d'attraction entre deux objets.

Nous allons maintenant établir une forme plus générale de ce potentiel.

**Définition 3.13** On appelle une **fonction faiblement sur-additive relativement à un blob  $B$**  toute application  $f(., B) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}^+$  vérifiant les propriétés suivantes :

- $\forall R \in \mathcal{P}(\Omega), f(R, B) \geq 0,$
- $\forall R, R' \in \mathcal{P}(\Omega) \text{ t.q } R \leftrightarrow R', f(R \cup R', B) \geq \max(f(R, B), f(R', B)).$

**Proposition 3.13** Soit  $B$  un blob,  $d$  une distance métrique issue d'un produit scalaire et  $f$  une fonction faiblement sur-additive, alors

$$\forall R \in \mathcal{P}(\Omega), \quad \delta_2(R, B) = \beta \frac{d^2(R, B)}{f(R, B)},$$

est un potentiel convexe et par conséquent sous-additif relativement au blob  $B$ .

**Preuve :** Soient  $R_i$  et  $R_j$  deux régions disjointes et adjacentes et  $B$  un blob de l'image. Soit  $d$  une distance métrique issue d'un produit scalaire.

Puisque  $f$  est faiblement sur-additive alors

$$\begin{cases} f(R_i \cup R_j, B) \geq f(R_i, B), \\ f(R_i \cup R_j, B) \geq f(R_j, B). \end{cases}$$

$\Rightarrow$

$$\begin{cases} \frac{f(R_i, B)}{f(R_i \cup R_j, B)} \leq 1, \\ \frac{f(R_j, B)}{f(R_i \cup R_j, B)} \leq 1. \end{cases}$$

$$\begin{aligned} \delta_2(R_i \cup R_j, B) &= \beta \frac{d^2(R_i \cup R_j, B)}{f(R_i \cup R_j, B)} \\ &= \underbrace{\alpha \frac{f(R_i, B)}{f(R_i \cup R_j, B)} \delta_2(R_i, B)}_{\leq 1} + (1 - \alpha) \underbrace{\frac{f(R_j, B)}{f(R_i \cup R_j, B)} \delta_2(R_j, B)}_{\leq 1} \\ &\quad - \underbrace{\frac{\alpha(1 - \alpha)}{f(R_i \cup R_j, B)} d^2(R_i, R_j)}_{\geq 0} \\ &\leq \alpha \delta_2(R_i, B) + (1 - \alpha) \delta_2(R_j, B), \end{aligned}$$

$\Rightarrow \delta_2(., B)$  est convexe et par conséquent sous-additive [c.q.f.d].

**Proposition 3.14** Toute fonction additive, sur-additive ou croissante pour l'ordre d'inclusion est une fonction faiblement sur-additive.

**Preuve :** Soient  $R_i$  et  $R_j$  deux régions adjacentes d'une partition  $P$  et  $B$  un blob de l'image.

1- Si  $f(., B)$  est additive ou sur-additive alors

$$f(R_i \cup R_j, B) \geq f(R_i, B) + f(R_j, B) \Rightarrow \begin{cases} f(R_i \cup R_j, B) \geq f(R_i, B), \\ f(R_i \cup R_j, B) \geq f(R_j, B). \end{cases}$$

$$\Rightarrow f(R_i \cup R_j, B) \geq \max(f(R_i, B), f(R_j, B)).$$

2- Si  $f(., B)$  est croissante pour l'ordre d'inclusion alors

$$\forall R, R' \in \mathcal{P}(\Omega), \quad t.q \quad R \subseteq R', \quad f(R, B) \leq f(R', B),$$

on a donc

$$\begin{cases} f(R_i, B) & \leq f(R_i \cup R_j, B), \\ f(R_j, B) & \leq f(R_i \cup R_j, B). \end{cases} \Rightarrow f(R_i \cup R_j, B) \geq \max(f(R_i, B), f(R_j, B)).$$

[c.q.f.d].

Dès lors, nous pouvons proposer d'autres potentiels convexes (donc sous-additifs) qu'on appellera **pseudo-attractifs** en remplaçant le produit des masses de la région et du blob,  $|R| \times |B|$ , par une fonction faiblement sur-additive. Dans ce qui suit nous énumérerons des exemples de potentiels sous-additifs « pseudo-attractifs » à partir de fonctions toutes supra-additives en raison de leurs croissances ou de leurs sur-additivités :

1. pour  $f(R, B) = 1$ , on obtient  $\delta(R, B) = d^2(R, B)$ ,
2. si  $f(R, B) = |R \cup B|$ , on obtient  $\delta(R, B) = \frac{d^2(R, B)}{|R \cup B|}$ ,
3. si  $f(R, B) = \frac{|R|}{|B|}$ , on obtient  $\delta(R, B) = \frac{d^2(R, B)}{|R|/|B|}$ ,
4. ..., etc.

**Lemme 3.2** L'inverse d'une fonction supra-additive est une fonction faiblement sur-additive et réciproquement.

Ce lemme dont la démonstration est triviale nous permet alors d'utiliser l'inverse des fonctions supra-additives, définie dans la section précédentes, comme fonctions faiblement sur-additives. Pour exemple, nous pouvons citer :

1.  $f(R, B) = e^{K \cdot \frac{|R|}{|B|}}$ ,
2. si  $f(R, B) = e^{K \cdot \frac{|R \cap B|}{|B|}}$ ,
3. si  $f(R, B) = e^{-(K \cdot |B| - |R|)}$ ,
4. ..., etc.

Tous ces potentiels sous-additifs (car possédant la propriété de convexité) « pseudo-attractifs » peuvent être utilisés dans la construction d'une énergie spatiale sous-additive dont il convient de tester leurs pertinences. Il en existe d'autres mais nous nous contenterons de ceux cités ci-dessus.

### Cadre de blobs elliptiques : distance de Mahalanobis

Dans le cas de blobs elliptiques, tout blob  $B_k$  est caractérisé par la longueur de ses axes majeur et mineur :  $a_k, b_k$  ainsi que son orientation  $\alpha_k$ . Nous pouvons ainsi utiliser la distance de Mahalanobis dans la construction d'un potentiel pseudo attractif.

La distance de Mahalanobis est définie par :

$$d_{\Sigma}(x, y)^2 = (x - y)^t \Sigma^{-1} (x - y)$$

pour  $x, y \in \mathbb{R}^M$ .

La norme de Mahalanobis est définie par :

$$\|x\|_{\Sigma}^2 = x^t \Sigma^{-1} x$$

pour  $x \in \mathbb{R}^M$ .

Ainsi chaque blob elliptique  $B_k$  dans  $\mathbb{R}^2$  est caractérisé par sa matrice de covariance  $\Sigma_k^{-1}$  qui s'écrit en fonction de  $a_k$ ,  $b_k$  et  $\alpha_k$  comme suit :

$$\Sigma_k^{-1} = \begin{pmatrix} \frac{\cos^2(\alpha_k)}{a_k^2} + \frac{\sin^2(\alpha_k)}{a_k^2} & -\sin(2\alpha_k)/2(\frac{1}{a_k^2} - \frac{1}{b_k^2}) \\ -\sin(2\alpha_k)/2(\frac{1}{a_k^2} - \frac{1}{b_k^2}) & \frac{\sin^2(\alpha_k)}{a_k^2} + \frac{\cos^2(\alpha_k)}{a_k^2} \end{pmatrix}$$

Le potentiel pseudo-attractif d'une région relativement à un blob s'écrira :

$$\delta_{\Sigma_k^{-1}}(R, B_k) = \beta \frac{d_{\Sigma_k^{-1}}^2(R, B_k)}{f(R, B_k)}$$

tel que  $f(R, B_k)$  est une fonction faiblement sur-additive.

### 3.4.3 La sous-additivité

Comme nous l'avons vu précédemment, une distance métrique issue d'un produit scalaire possède la propriété de convexité. Ceci entraîne la convexité des potentiels étudiés dans la précédente section et par conséquent leurs sous-additivités.

Nous allons maintenant présenter une distance sous-additive connue sous le nom de « **distance du cosinus** ».

**Proposition 3.15** Soit  $\langle x, y \rangle = \sum_{i=1, \dots, M} x_i \cdot y_i$  le produit scalaire usuel défini sur  $\Omega \subseteq \mathbb{R}^+ \times \mathbb{R}^+$ . La distance du cosinus relativement au blob  $\mathbf{B}$ , définie par :

$$d_{\cos}(R, B) = d_{\cos}(c_R, c_B) = \frac{\langle c_R, c_B \rangle}{\|c_R\| \cdot \|c_B\|}$$

est sous-additive.

**Preuve :** Soient  $R_1$  et  $R_2$  deux régions adjacentes d'une partition  $P$  et  $B$  un blob de l'image. Posons  $\alpha = R_1/(R_1 + R_2)$

$$\begin{aligned} d_{\cos}(R_1 \cup R_2, B) &= d_{\cos}(c_{R_1 \cup R_2}, c_B) = \frac{\langle c_{R_1 \cup R_2}, c_B \rangle}{\|c_{R_1 \cup R_2}\| \cdot \|c_B\|} \\ &= \frac{\langle \alpha c_{R_1} + (1 - \alpha) c_{R_2}, c_B \rangle}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\| \cdot \|c_B\|} \\ &= \frac{\alpha \langle c_{R_1}, c_B \rangle + (1 - \alpha) \langle c_{R_2}, c_B \rangle}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\| \cdot \|c_B\|} \\ &= \frac{\alpha \cdot \|c_{R_1}\| \cdot d_{\cos}(R_1, B) + (1 - \alpha) \cdot \|c_{R_2}\| \cdot d_{\cos}(R_2, B)}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\|} \\ &= \frac{\alpha \cdot \|c_{R_1}\|}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\|} \cdot d_{\cos}(R_1, B) + \frac{(1 - \alpha) \cdot \|c_{R_2}\|}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\|} \cdot d_{\cos}(R_2, B). \end{aligned}$$

Il reste donc à démontrer que

$$\frac{\alpha \cdot \|c_{R_1}\|}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\|} \leq 1 \text{ et que } \frac{(1 - \alpha) \cdot \|c_{R_2}\|}{\|\alpha c_{R_1} + (1 - \alpha) c_{R_2}\|} \leq 1.$$

Nous allons tout d'abord développer le carré du dénominateur :

$$\begin{aligned}
 \|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2 &= \langle \alpha c_{R_1} + (1 - \alpha)c_{R_2}, \alpha c_{R_1} + (1 - \alpha)c_{R_2} \rangle \\
 &= \alpha^2 \langle c_{R_1}, c_{R_1} \rangle + (1 - \alpha)^2 \langle c_{R_2}, c_{R_2} \rangle + 2\alpha(1 - \alpha) \langle c_{R_1}, c_{R_2} \rangle \\
 &= \alpha^2 \|c_{R_1}\|^2 + (1 - \alpha)^2 \|c_{R_2}\|^2 + 2\alpha(1 - \alpha) \langle c_{R_1}, c_{R_2} \rangle \\
 &\Rightarrow \\
 \left\{ \begin{array}{l} \|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2 - \alpha^2 \|c_{R_1}\|^2 = (1 - \alpha)^2 \|c_{R_2}\| + 2\alpha(1 - \alpha) \langle c_{R_1}, c_{R_2} \rangle \geq 0 \\ \|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2 - (1 - \alpha)^2 \|c_{R_2}\|^2 = \alpha^2 \|c_{R_1}\| + 2\alpha(1 - \alpha) \langle c_{R_1}, c_{R_2} \rangle \geq 0 \end{array} \right. \\
 &\Rightarrow \\
 \left\{ \begin{array}{l} \|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2 - \alpha^2 \|c_{R_1}\|^2 \geq 0 \\ \|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2 - (1 - \alpha)^2 \|c_{R_2}\|^2 \geq 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \frac{\alpha^2 \|c_{R_1}\|^2}{\|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2} \leq 1 \\ \frac{(1 - \alpha)^2 \|c_{R_2}\|^2}{\|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|^2} \leq 1 \end{array} \right.
 \end{aligned}$$

Puisque  $\alpha$  et  $(1 - \alpha)$  sont positifs, la racine carré des deux inégalités précédentes donnent :

$$\left\{ \begin{array}{l} \frac{\alpha \|c_{R_1}\|}{\|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|} \leq 1 \\ \frac{(1 - \alpha) \|c_{R_2}\|}{\|\alpha c_{R_1} + (1 - \alpha)c_{R_2}\|} \leq 1 \end{array} \right.$$

Par conséquent :

$$d_{cos}(R_1 \cup R_2, B) \leq d_{cos}(R_1, B) + d_{cos}(R_2, B) \quad [\text{c.q.f.d}]$$

Dans le même esprit que la proposition 3.9, nous pouvons démontrer de la même manière le résultat suivant :

**Proposition 3.16** *Soit  $B$  un blob,  $\langle ., . \rangle$  le produit scalaire usuel défini sur  $\Omega \in \mathbb{R}^+ \times \mathbb{R}^+$  et  $f$  une fonction faiblement sur-additive, alors*

$$\forall R \in \mathcal{P}(\Omega), \quad \delta_3(R, B) = \beta \frac{d_{cos}(R, B)}{f(R, B)},$$

est un potentiel sous-additif relativement au blob  $B$ .

Ainsi la distance du cosinus combinée avec une fonction faiblement sur-additive permet de construire des potentiels sous-additifs qu'on peut aussi assimiler à des potentiels pseudo-attractifs.

#### 3.4.4 Conclusion

Dans les deux précédentes sections nous avons développé une méthode de construction d'une énergie spatiale sur une partition. Cette méthode est ascendante, elle commence par la définition d'un potentiel d'une région, relativement à un blob, qui mesure l'intensité d'attraction. Nous avons ensuite proposé de définir l'énergie d'une région relativement à un blob de trois différentes manières. Une énergie mono-potentielle mesure uniquement le degré d'attraction entre une région et un blob. Une énergie multi-potentielle mesure l'attraction d'une région à un blob en tenant compte aussi des potentiels des voisins de la région. L'énergie bi-potentielle s'appuie quand à elle sur la définition d'un potentiel d'attraction entre deux régions relativement à un blob. L'énergie de la région est alors la somme des bi-potentiels de la région avec toute région voisine.

Finalement, l'énergie spatiale d'une région  $D_B(R)$  est définie par la norme du vecteur des énergies relatives à chaque blob. La convexité d'un potentiel combinée à celle d'une norme nous assure que l'énergie spatiale d'une région est convexe et par conséquent sous-additive. On notera que si le potentiel est uniquement sous-additif alors la norme utilisée dans la définition de l'énergie spatiale d'une région et qui garantit sa sous-additivité est la norme 1 et la norme infini (voir la proposition 3.8).

## 3.5 RÉSULTATS EXPÉRIMENTAUX

Cette section met finalement en application l'approche proposée dans ce chapitre et présente des résultats expérimentaux. Une comparaison qualitative entre coupes issues de hiérarchies construites sans *a priori* et coupes issues de hiérarchies construites avec un *a priori* spatial est systématiquement proposée afin de mettre en évidence le rôle joué par l'information spatiale dans la construction d'une hiérarchie.

On utilisera comme partition d'initialisation la sur-segmentation obtenue par le même algorithme de partage des eaux utilisé dans les précédents chapitres. Les représentations ensembles-échelles sans *a priori* spatial seront construites sur la base du modèle constant par morceaux de Mumford-Shah.

Les représentations ensembles-échelles avec *a priori* spatial seront construites en combinant la même énergie radiométrique (Ecart quadratique) et une énergie spatiale qui peut être soit mono, multi ou bi-potentielle.

### 3.5.1 Cas d'un blob

Nous allons dans un premier temps nous restreindre à un seul objet d'intérêt dans l'image et par conséquent l'*a priori* spatial sera composé d'un seul blob.

#### 3.5.1.1 Energie spatiale mono-potentielle

La construction d'une énergie spatiale mono-potentielle nécessite de faire les choix suivants :

1. une distance métrique telle que la distance euclidienne, la distance de Mahalanobis ou la distance du cosinus,
2. une fonction faiblement sur-additive :  $|R| \cdot |B|$ ,  $|R \cap B| / |R|$ ,  $e^{K \cdot \frac{|R|}{|B|}}$  ou  $e^{K \cdot \frac{|R \cap B|}{|R|}}$ ,
3. une norme pour le vecteur des énergies relatives aux blobs :  $L^2$ ,  $L^1$  ou  $L^\infty$ .

Le nombre de combinaisons entre distance, fonction faiblement sur-additive et norme est important et nous ne pourrons pas les explorer toutes. On se limitera à quelques configurations pour tester et analyser le comportement de l'énergie spatiale mono-potentielle.

Nous allons dans un premier temps analyser le comportement de l'énergie spatiale sur une image de synthèse. La figure 3.7-(a) se compose de trois rectangles de même couleur (qu'on notera  $g_i$ ,  $i = 1, \dots, 3$ ) et de taille croissante (de droite à gauche). Ces trois rectangles sont englobés dans un grand rectangle de couleur blanche (qu'on notera  $b$ ). L'image blob est représentée dans la figure 3.7-(b) où nous avons entouré le plus petit rectangle ( $g_1$ ) par un blob. Le tableau 3.1 résume les énergies spatiales calculées sur chacun des rectangles ainsi que les énergies

spatiales de l'union de chaque rectangle gris avec le rectangle blanc. Rappelons que l'échelle d'apparition d'une région  $x$  s'écrit :

$$\lambda^+(x) = -\frac{D(x) - D(P)}{S(x) - S(P)},$$

tel que  $P$  représente la partition optimale de la hiérarchie partielle de la région  $x$ .

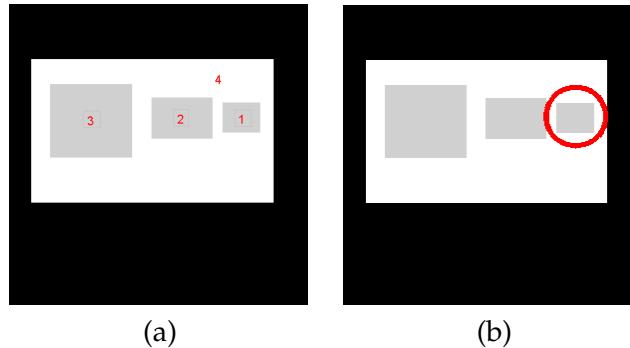


FIGURE 3.7 – a) image de synthèse, b) image-blob.

Le dénominateur  $\Delta S = S(x) - S(P)$  représente le coût de la fusion au sens de l'énergie spatiale des régions composant la partition  $P$  en une seule région  $x$ . Plus ce coût est faible et plus l'apparition de la région  $x$  est retardée.

Comme le montre le tableau 3.1 l'énergie spatiale du petit rectangle gris  $S(g_1)$  est la plus faible. Elle dénote la forte attraction de la région vis à vis du blob. Par contre l'énergie spatiale du grand rectangle est la plus forte indiquant une faible attraction entre la région et le blob. Le coût de la fusion,  $|\Delta S(g_1, b)|$ , entre le petit rectangle gris et le rectangle blanc est le plus faible parmi les fusions possibles entre rectangles gris et le rectangle blanc. Par conséquent, l'apparition  $\lambda^+(g_1)$  du rectangle gris sera élevée. Un coût de fusion spatial faible se traduit par une pénalisation plus forte sur l'apparition de la région tandis qu'un coût de fusion spatial élevé encourage la fusion de la région avec ses régions voisines. Normalement, la première fusion devrait impliquer le plus petit des rectangles gris  $g_1$  avec le rectangle blanc mais ici le rectangle  $g_1$  fusionnera après la fusion des autres rectangles,  $g_2$  et  $g_3$ .

La figure 3.8 représente les ensembles-échelles obtenus sur cette image de synthèse. La première colonne représente l'ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Les coupes successives de la hiérarchie sont présentées dans l'ordre inverse des fusions, de haut en bas. On explore la hiérarchie du détail le plus grossier au détail le plus fin. On constate que la construction hiérarchique respecte l'échelle spatiale des rectangles contenu dans l'image. Ainsi le petit rectangle est le premier à disparaître de la hiérarchie, il s'en suit la fusion du rectangle de taille intermédiaire puis celui qui a la taille la plus grande. Ce comportement est tout à fait satisfaisant et conforme au formalisme développé par Laurent Guigues à savoir un outil d'analyse en échelle du contenu de l'image. Maintenant, supposons que l'objet d'intérêt principal comme le montre l'image-blob 3.8-(2)-(b). La deuxième colonne représente l'ensemble-échelle obtenu en combinant l'énergie radiométrique de Mumford-Shah et l'énergie spatiale mono-potentielle basée sur le potentiel attractif (rapport entre la distance euclidienne et le produit des surfaces de la région et du blob). On observe, dans ce cas, que le dernier rectangle fusionné est le petit rectangle. À l'exception de ce changement dans l'ordre de fusion, les

<b>S</b>			
$g_1$	$g_2$	$g_3$	$b$
0.633425	60.2683	150.787	80.655
<b>S</b>			
	$g_1 \cup b$	$g_2 \cup b$	$g_3 \cup b$
	33.6875	78.7556	93.4526
$ \Delta S $			
	$(g_1, b)$	$(g_2, b)$	$(g_3, b)$
	47.6009	62.1677	137.9894

TABLE 3.1 – Le tableau illustre les énergies spatiales mono-potentielle des 3 rectangles gris et du rectangle blanc, les énergies de leur union et le coût de chaque fusion. Le coût de la fusion  $|\Delta S|(g_1, b)$  est le plus faible. Il correspond au petit rectangle gris  $g_1$  qui est entouré par un blob. La fusion du rectangle  $g_1$  avec le rectangle blanc sera plus fortement pénalisée par rapport aux autres fusions. Ainsi, ce rectangle sera fusionné ultérieurement contrairement à ce qui se produit dans l'approche ensemble-échelle classique où ce dernier est le premier absorbé par le rectangle blanc.

priorités des fusions sont respectées entre les deux rectangles restant, à savoir dans l'ordre croissant de la taille de l'objet. Dans la dernière colonne, on a encerclé le rectangle de taille moyenne et on relève le même comportement que celui noté sur la première image-blob.

Cet exemple trivial montre clairement le rôle de l'énergie spatiale dans la construction hiérarchique. Ainsi, l'utilisation d'une énergie spatiale à la place d'une énergie classique permet à un objet circonscrit (par un blob) de lui conférer un degré de pertinence plus élevé. L'énergie élève cet objet dans la hiérarchie des échelles et rend sa fusion avec une autre région plus difficile. Cette énergie permet aussi de se focaliser sur cet objet en fusionnant dans un premier temps les régions qui se trouvent en dehors du blob. En somme, l'énergie bouleverse l'ordre pré-établi de la pertinence des objets. Cet ordre était au départ guidé par la taille des objets, autrement dit par l'échelle spatiale des régions<sup>5</sup>

---

5. En toute rigueur, l'approche ensemble-échelle intègre l'échelle spatiale des objets à travers l'énergie d'attache aux données. Elle intègre également l'échelle de l'axe radiométrique (niveaux de gris) dans cette même énergie.

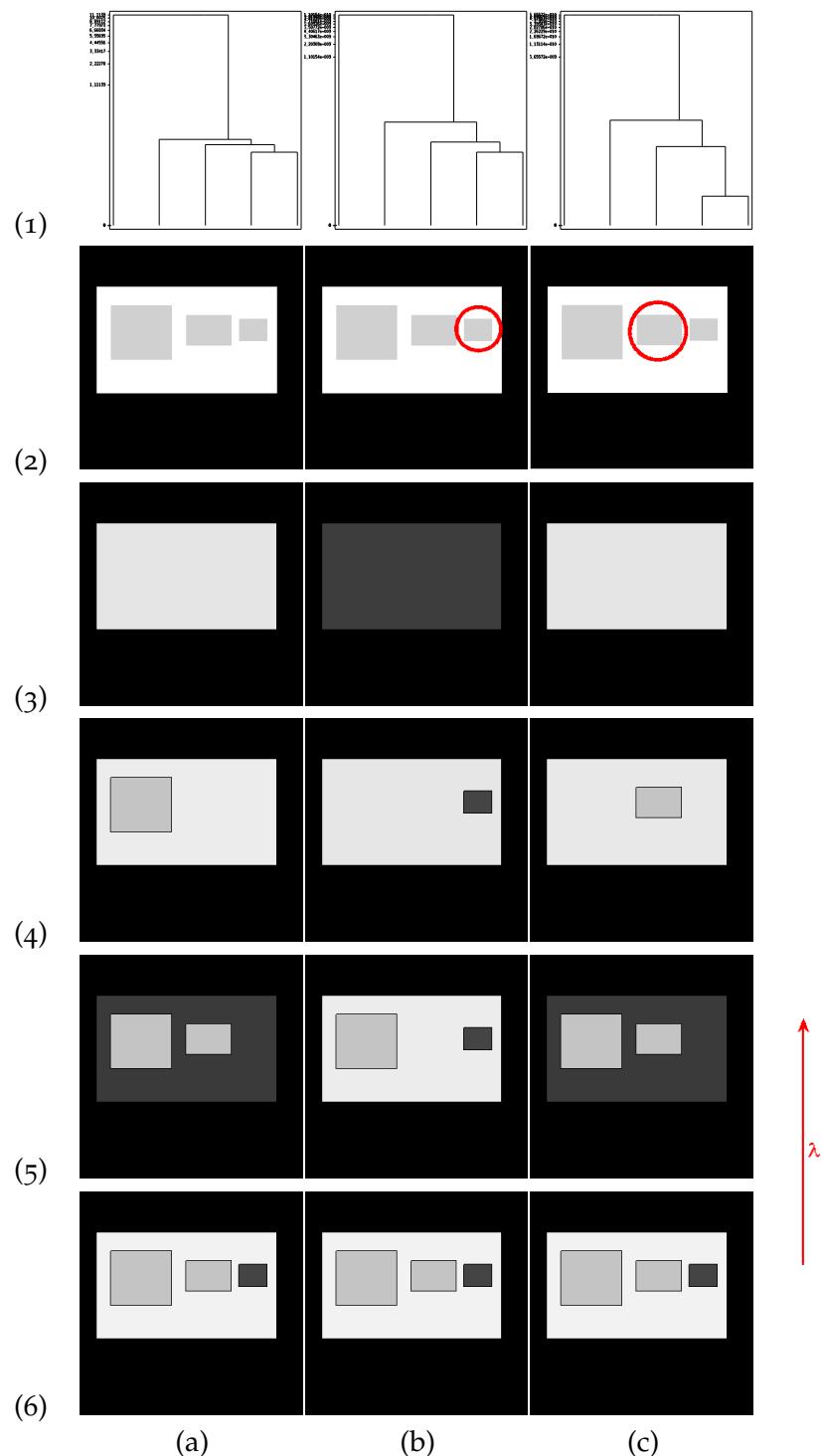


FIGURE 3.8 – La colonne (a) illustre l'ensemble-échelle obtenu avec l'énergie de Mumford-Shah. La colonne (b) celui obtenu avec l'énergie spatial mono-potentiel attractif partant d'un blob entourant le petit rectangle. La dernière colonne (c) utilise un blob entourant le rectangle de taille moyenne.

La figure 3.9 montre les différentes séquences de partitions obtenues avec une énergie spatiale utilisant une fonction faiblement sur-additive différente à chaque ligne (respectivement dans l'ordre  $|R \cap B|/|R|$ ,  $e^{|R \cap B|/|R|}$  et  $e^{|R|/|B|}$ ). Le comportement global est le même que celui relevé dans la figure précédente à savoir l'objet entouré par le blob devient plus pertinent que les objets de taille plus grande. Néanmoins, les deux premiers potentiels pseudo attractifs qui s'écrivent en fonction du degré d'inclusion de la région dans le blob n'attribuent pas d'importance à l'ordre de fusion entre les deux rectangles restants. Les deux objets sont fusionnés en même temps avec leur environnement direct (le rectangle blanc). Un phénomène d'inversion s'est produit lors de l'optimisation finale de l'escalade binaire où l'énergie spatiale n'a pas accordé une priorité plus forte à l'un ou l'autre des deux rectangles.

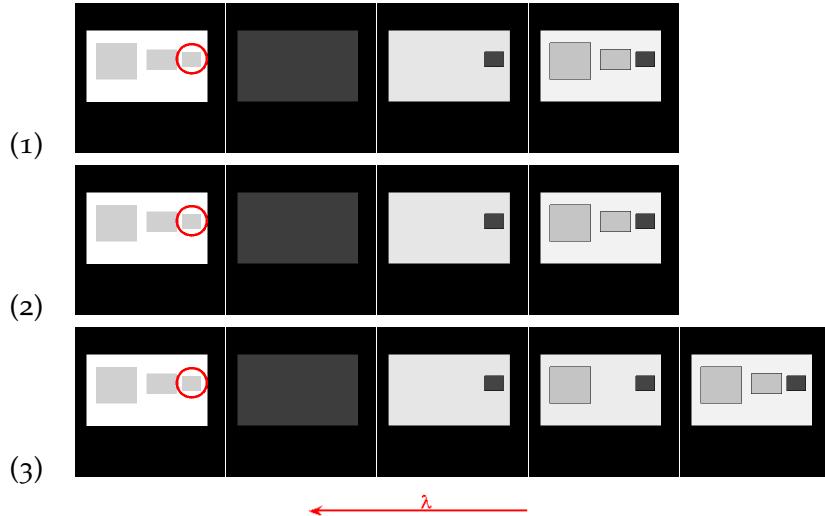


FIGURE 3.9 – (1) coupes obtenues avec l'énergie spatiale mono-potentielle à l'aide d'un potentiel pseudo-attractif avec la fonction  $|R \cap B|/|R|$ . (2) coupes obtenues avec la fonction faibl. sur-additive  $e^{|R \cap B|/|R|}$  et la ligne (3) avec la fonction  $e^{|R|/|B|}$ .



FIGURE 3.10 – (1) coupes obtenues avec l'énergie spatiale mono-potentielle à l'aide d'un potentiel pseudo-attractif avec la fonction  $|R \cap B|/|R|$ . (2) coupes obtenues avec la fonction faibl. sur-additive  $e^{|R \cap B|/|R|}$  et la ligne (3) avec la fonction  $e^{|R|/|B|}$ .

A présent, nous allons étudier le comportement de cette énergie spatiale sur une image réelle comportant de nombreux objets pertinents. La figure 3.11 représente l'image d'un trafic routier dans un centre ville. L'image comporte des voitures de différentes tailles et couleurs, des rues marquées par des lignes blanches, une partie d'immeuble et des arbres. Nous nous sommes focalisé sur le bus qui se trouve au centre de l'image comme le montre l'image blob dans la figure 3.11-(a)-(2).

On rappelle qu'une hiérarchie de régions est une entité volumique et par conséquent la navigation à l'intérieur nécessite l'extraction de coupes horizontales. Nous avons choisi de prélever des coupes horizontales par tranche de 20 régions.

La première colonne expose des coupes extraites d'une hiérarchie construite à partir de l'énergie de Mumford-Shah. Les partitions obtenues dans chaque colonne apparaissent dans l'ordre décroissant du nombre de régions de haut en bas de la colonne (respectivement, 20, 40, 60, 80 régions). Le bus comportant des structures hétérogènes il est évident, qu'elles ne peuvent être fusionnées en un seul objet d'intérêt en s'appuyant sur le modèle constant par morceaux de Mumford-Shah. La seule structure ayant survécu à de hautes échelles est le toit du bus qui apparaît dans la première partition comportant 20 régions. Dans les deux partitions suivantes, respectivement 40 et 60 régions, on voit apparaître deux structures supplémentaires appartenant au bus. Le reste des régions composant ces partitions se divisent en régions de grandes tailles telles que la chaussée, les rues mais aussi des régions de petites et moyennes tailles telles que des parties des voitures de couleurs blanches, des parties de l'immeuble, etc. Ces régions ont survécu soit pour leur radiométrie très différentes par rapport à leur environnement immédiat (régions de couleurs blanchâtres telles que des parties de certaines voitures) soit pour l'homogénéité de ces régions (telles que les chaussées ou certaines zones de l'immeuble). La partition comportant 80 régions exhibe quand à elle une partie importante du bus à savoir les fenêtres de couleur sombre. La dernière partition contient 150 régions et n'exhibe pas tous les détails du bus. La plupart des régions composant cette partition (et celles au dessus) sont de petites tailles et représentent soit des parties de l'immeuble soit des parties des lignes blanches du carrefour.

La deuxième colonne de la figure 3.11 représente les coupes extraites d'une hiérarchie construite à partir d'une énergie spatiale mono-potentielle issue du potentiel attractif. On a l'impression que les partitions obtenues via l'énergie spatiale comportent beaucoup plus de régions que les partitions issues de l'énergie de Mumford-Shah mais cette impression est fausse. En l'absence d'une énergie de régularisation, une région d'une partition issue de l'énergie spatiale peut avoir une frontière très sinuose et une forme complexe qui donne l'impression que la région est composée de plusieurs régions alors que c'est une seule et unique région. Inversement, les partitions issues de l'énergie de Mumford-Shah comporte de nombreuses petites régions noyées dans trois ou quatre grandes structures de contours réguliers ce qui renvoie une impression d'une partition faiblement chargée en termes de régions. On reviendra ultérieurement sur ce phénomène de frontières tortueuses et focalisons nous dans un premier temps sur la zone étudiée. La partition de 20 régions de la colonne (2) comporte plus de structures appartenant au bus que sa consœur de la première colonne. On constate qu'une coupe de 60 régions comporte plus de structures du bus qu'une coupe de 150 régions obtenue sans *a priori* spatial. La première observation que l'on peut tirer de l'analyse de cette deuxième colonne est que les structures se trouvant à l'intérieur du blob et aux abords de ce dernier survivent plus longtemps durant le processus de fusion.

**Influence de la non régularisation géométrique** La figure 3.12 illustre le rôle important que joue une énergie de régularisation telle que le périmètre dans le processus de fusion. La première colonne représente la coupe de 20 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. La deuxième colonne représente la coupe de 20 régions issue de la hiérarchie construite avec l'énergie spatiale mono-potentielle (potentiel attractif). La deuxième ligne de la figure représente ces coupes en couleurs afin de permettre une meilleure visualisation des régions construites. Comme nous l'avons dit précédemment, pour un nombre de régions équivalent, on a l'impression que la coupe extraite de la hiérarchie construite avec l'énergie de Mumford-Shah est plus légère en terme de nombre de régions que la coupe extraite avec l'énergie spatiale. En réalité, la première coupe (issue de Mumford-Shah) comporte de petites régions qui ont persistées dans la hiérarchie grâce à leur radiométrie qui se distingue fortement de leur voisinage. A l'inverse, la deuxième coupe (énergie spatiale) présente des régions de grandes tailles et de formes complexes. Pour exemple nous pouvons suivre l'évolution la région violète se trouvant en haut à droite de l'image. On remarque que cette région est de forme tortueuse. La progression de la surface de cette région alterne entre zones de grandes tailles et de petites zones qu'on assimilera à des goulets. Ces goulets donnent l'impression que la région est éclatée en plusieurs structures. Le processus de fusion est guidé par la radiométrie des régions et l'absence d'une énergie de régularisation a pour effet de ne plus pénaliser les fusions entre régions qui engendrent des zones de formes complexes ou comportant des frontières sinueuses. Ainsi l'absence d'une mesure sur la complexité géométrique des régions a pour effet de construire des régions aux formes complexes et aux frontières irrégulières. Ces formes renvoient alors la sensation que la partition contient un nombre important de régions. Cette perception est évidemment erronée pour les raisons que nous venons d'évoquées.

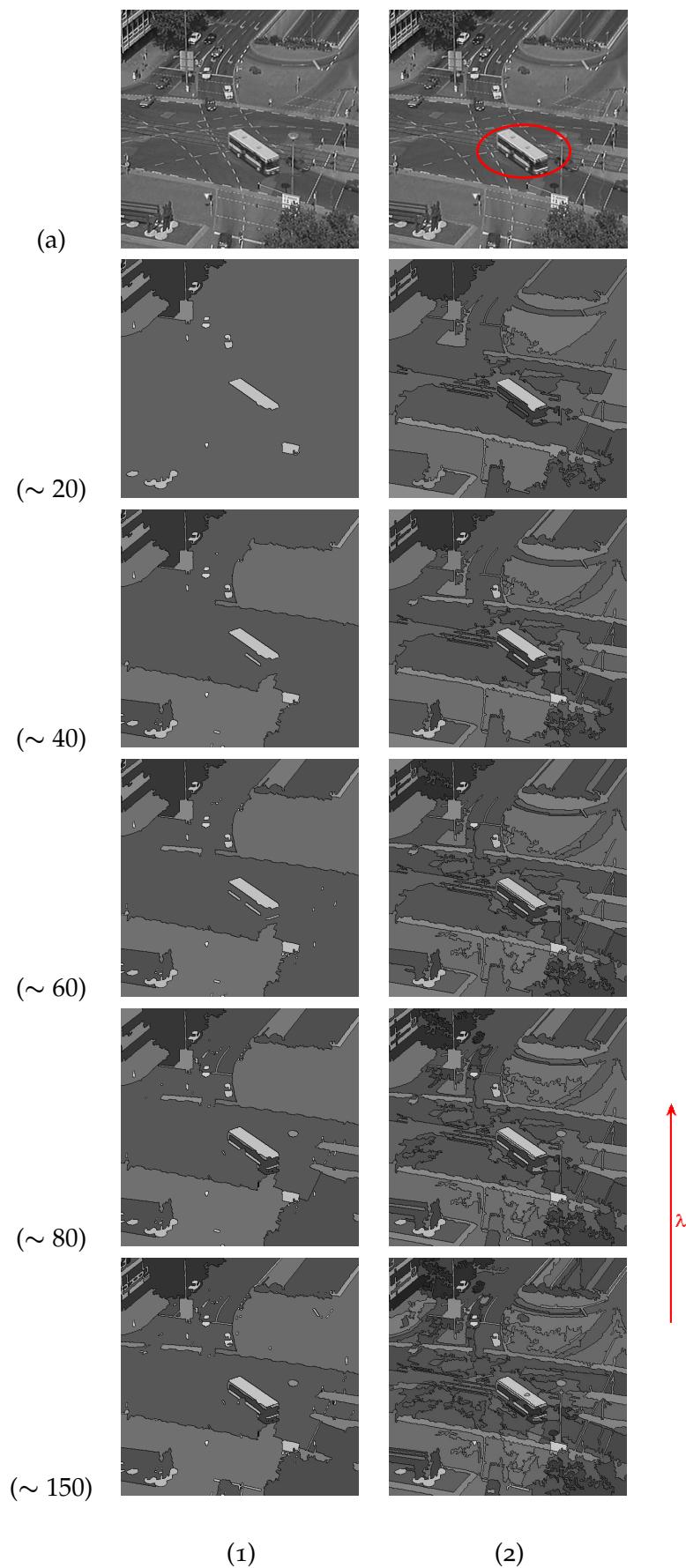


FIGURE 3.11 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) représente l'ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (attractive). La fonction faiblement sur-additive utilisée est :  $|R|.|B|$ .

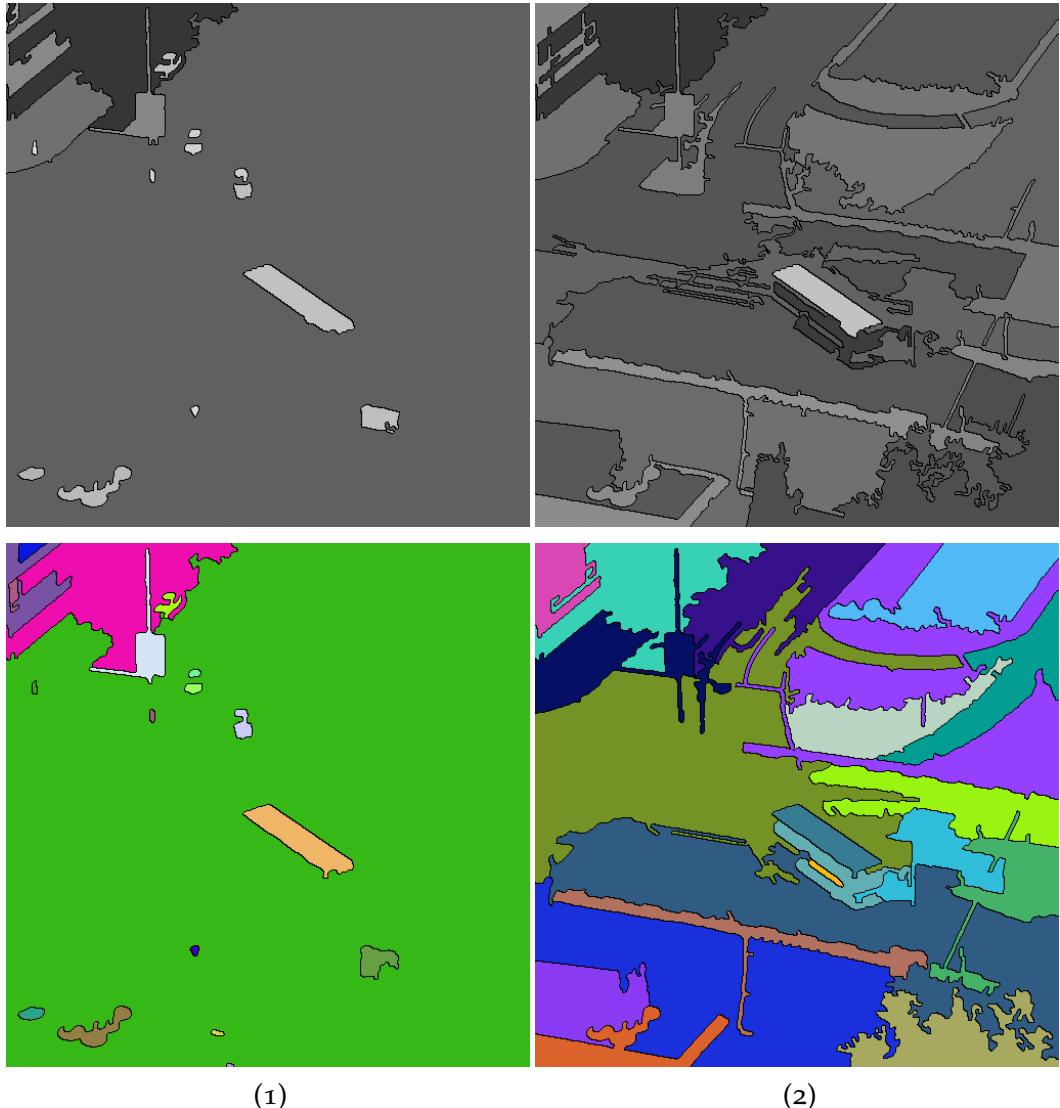


FIGURE 3.12 – Colonne (1) : coupe de 20 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 20 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R| / |B|}$ .

En se basant sur le même principe de comparaison, les figures 3.13, 3.15 et 3.17 présentent à chaque fois un parallèle entre l'ensemble échelle de Mumford-Shah et celui obtenu via une énergie spatiale mono-potentielle qui s'appuie sur les fonctions faiblement sur-additives suivantes :  $|R \cap B|/|R|$ ,  $e^{|R \cap B|/|R|}$  et  $e^{|R|/|B|}$ . Pour chaque figure : un même constat, les partitions construites avec un *a priori* spatial contiennent plus de détails sur l'objet d'intérêt que leurs consœurs obtenues sans *a priori* spatial. Une partition de 60 régions construite avec *a priori* spatial décrit l'objet d'intérêt avec plus de détails qu'une partition de 150 régions obtenue sans *a priori* spatial. Les petites régions de radiométrie différente par rapport à leur environnement qui caractérisent les coupes sans *a priori* spatial sont absentes dans les coupes construites avec *a priori* spatial. Elles sont plus rapidement absorbées. L'ordre des fusions des structures composant le bus est inverse, les régions sont lentement absorbées dans les hiérarchies construites avec *a priori* spatial et disparaissent rapidement dans la hiérarchie sans *a priori* spatial.

Qualitativement, on relèvera une certaine préférence pour les partitions issues des potentiels pseudo-attractif,  $|R \cap B|/|R|$  et  $e^{|R \cap B|/|R|}$ , par rapport à celles issues du potentiel dit attractif utilisé dans la figure 3.11. D'une part, elles contiennent plus de détails sur la composition de notre bus et d'autre part l'environnement immédiat du blob est plus dégagé et les contours sont moins tortueux. Cette différence est due à l'introduction de l'intersection de la région avec son blob dans la fonction faiblement sur-additive qui atténue la pénalisation des régions intersectant le blob et encourage de ce fait leur fusion avec des régions se trouvant à l'extérieur du blob.

Les figures ?? permettent de mieux visualiser les résultats obtenus sur une coupe comportant 60 régions. Ces coupes montrent clairement l'apport de l'énergie spatiale dans le processus de fusion.

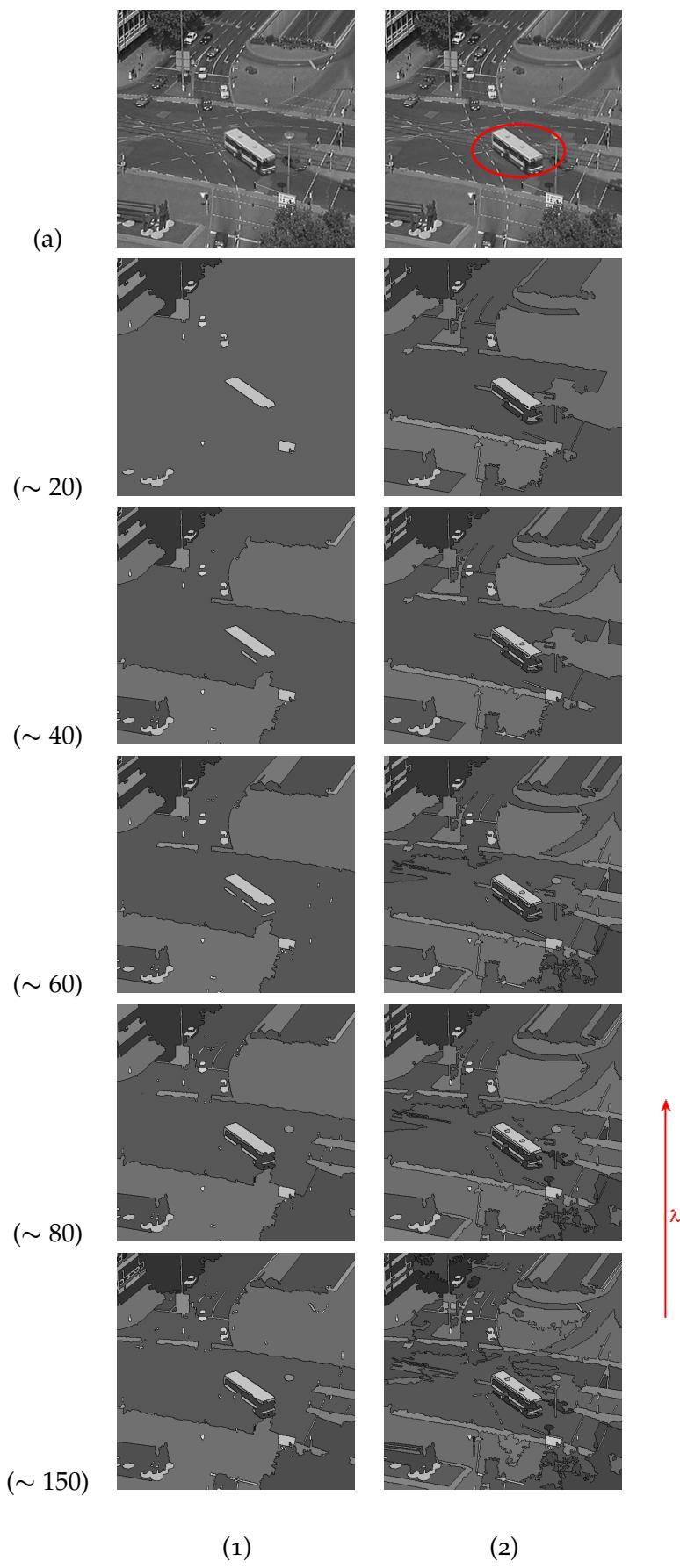


FIGURE 3.13 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle dont la fonction faiblement sur-additive utilisée est :  $|R \cap B| / |R|$ .

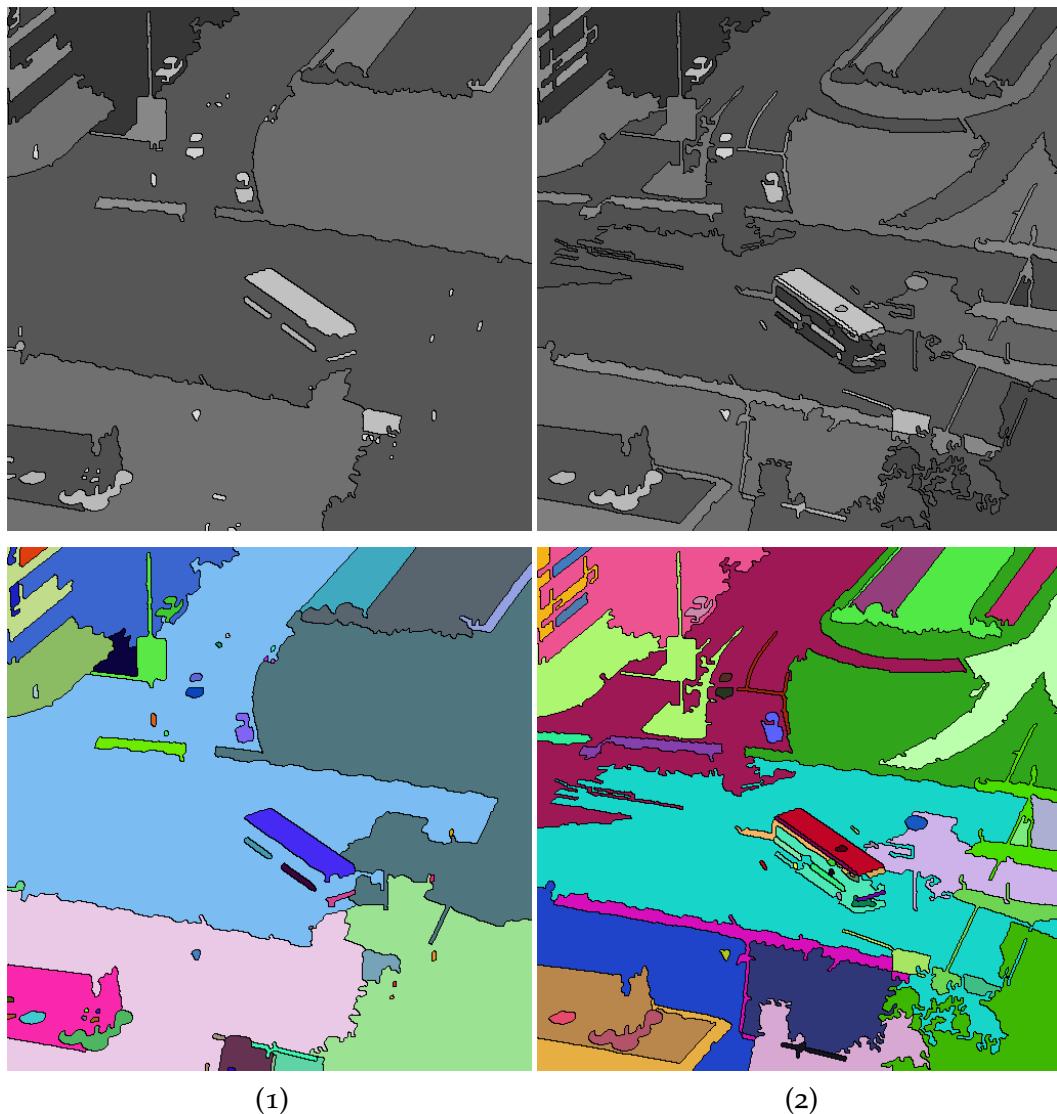


FIGURE 3.14 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R| / |B|}$ .

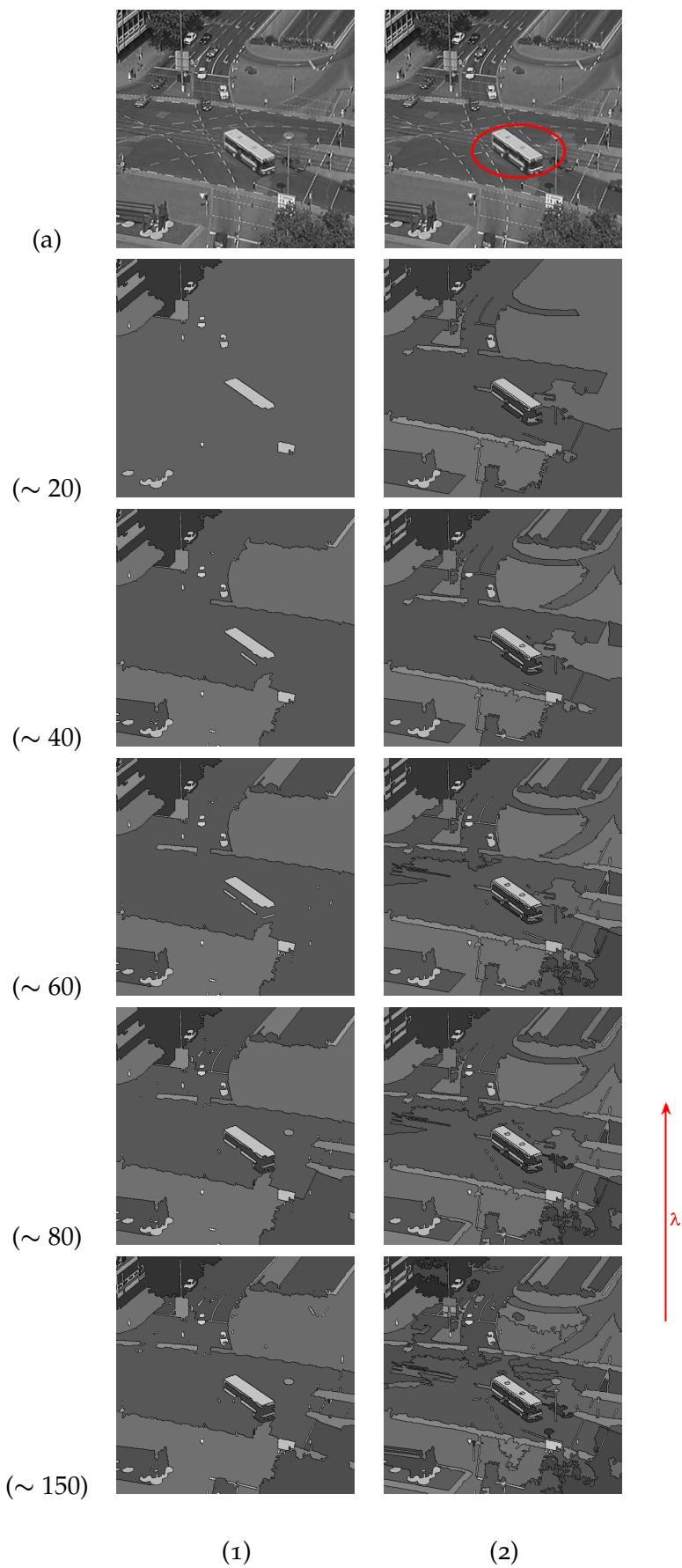


FIGURE 3.15 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle avec la fonction faiblement sur-additive  $e^{|R \cap B|/|R|}$ .

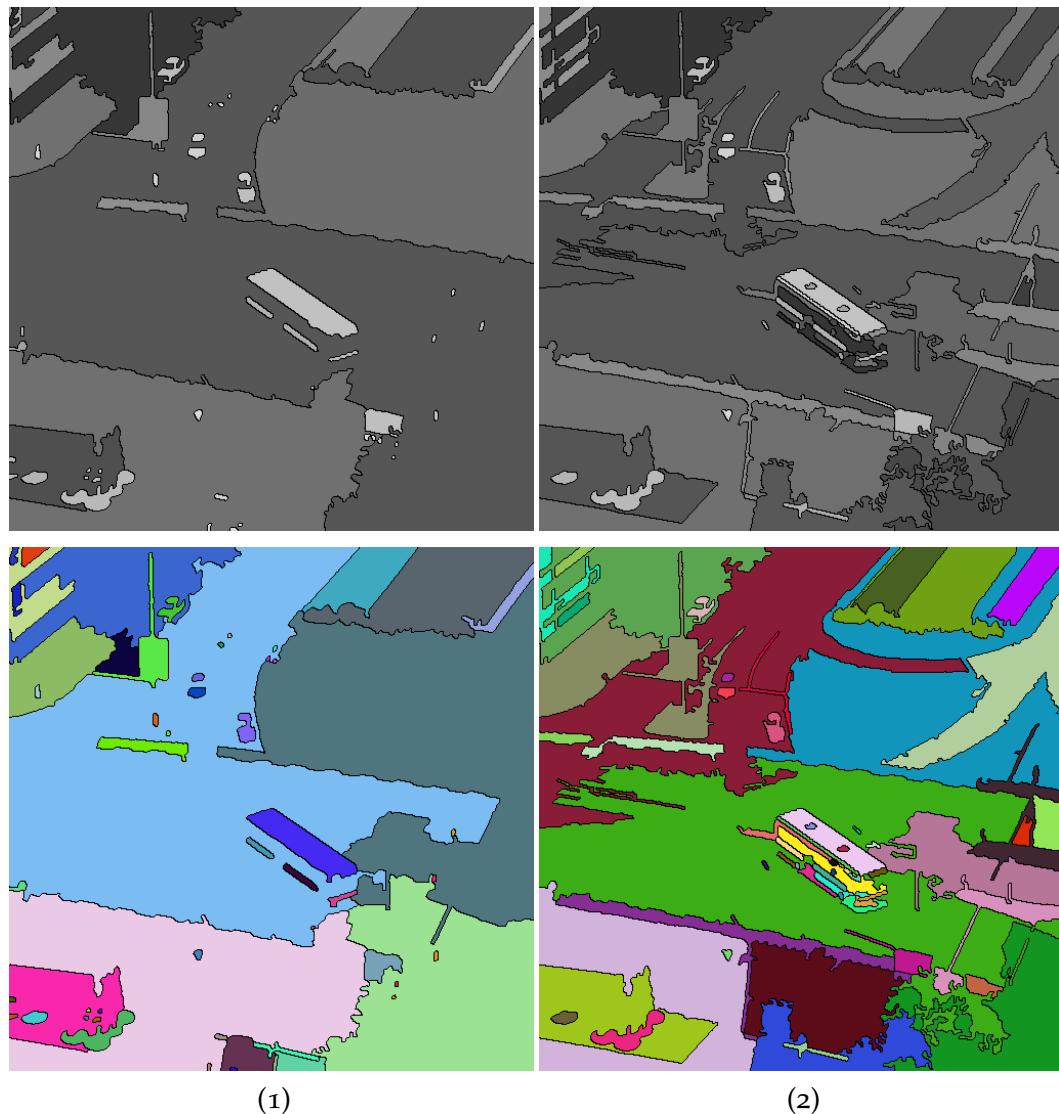


FIGURE 3.16 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R| / |B|}$ .

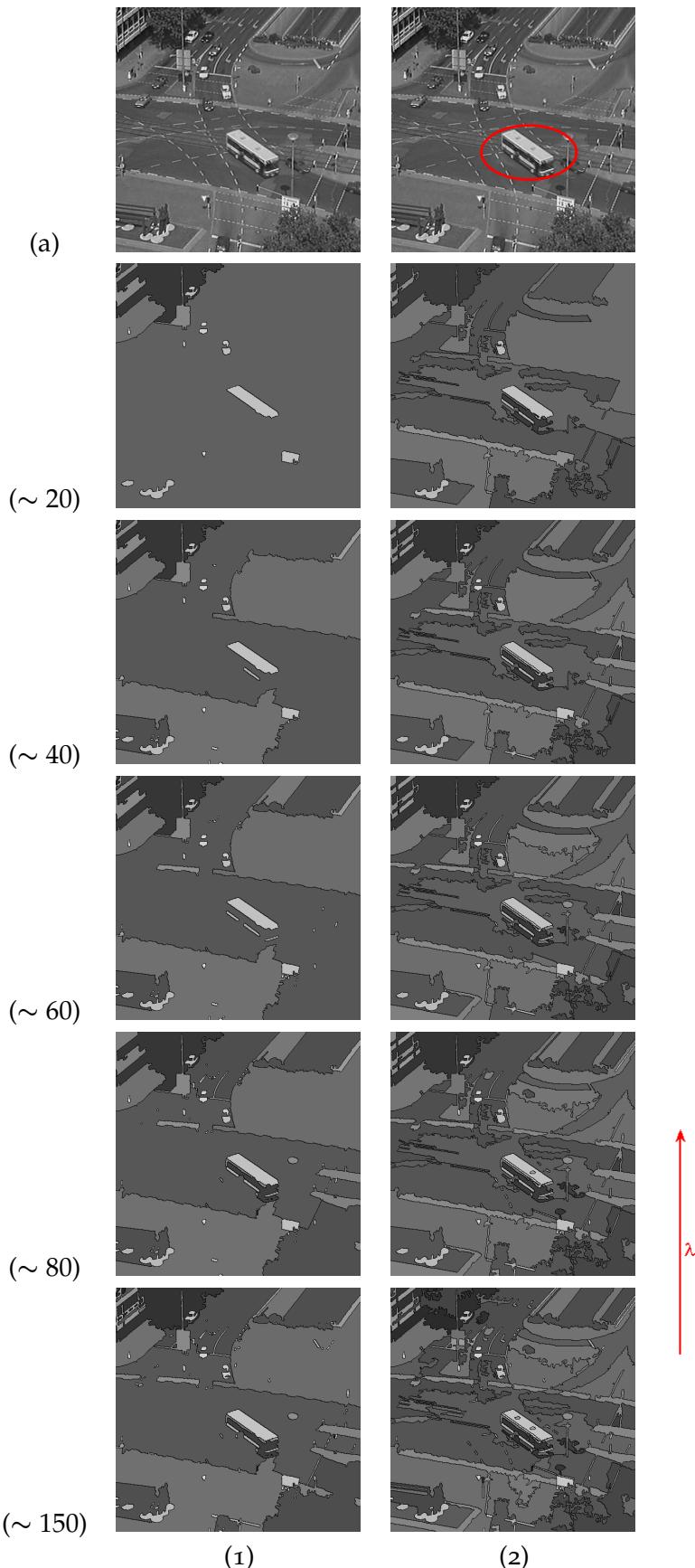


FIGURE 3.17 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle avec la fonction faiblement sur-additive  $e^{|R|/|B|}$ .

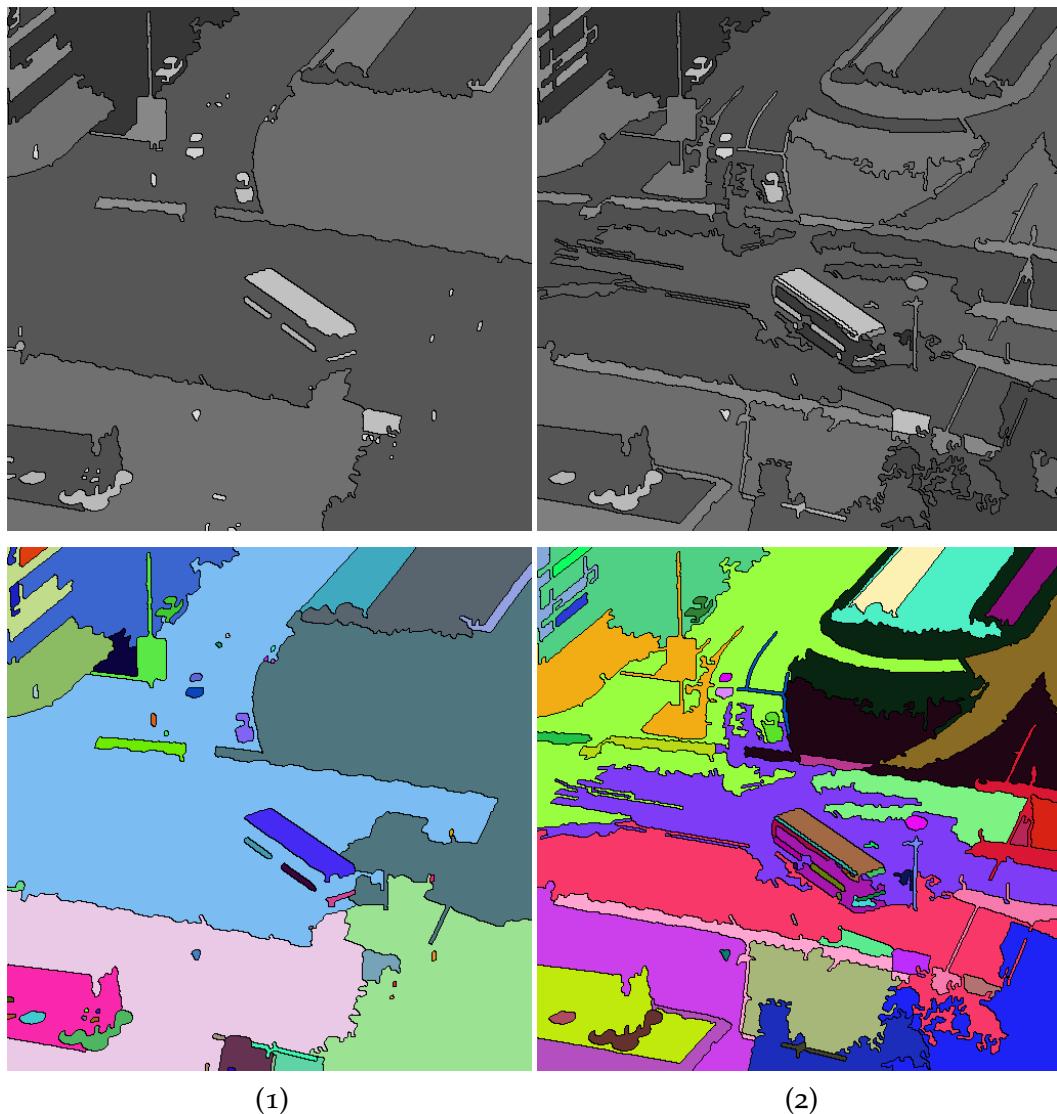


FIGURE 3.18 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R| / |B|}$ .

La figure 3.19 illustre également un trafic routier sur un périphérique de ville. L'image exhibe une voiture sur le point de passer sous un pont comportant deux panneaux. L'arrière plan est composé d'immeubles et d'arbres. Les panneaux, quand à eux portent des lettres de l'alphabet de tailles différentes. On s'intéresse à la voiture qui occupe un espace non-négligeable dans l'image mais qui porte une radiométrie proche du niveau de gris de la route sur laquelle elle se trouve.

Les deuxièmes colonnes des figures 3.19, 3.21, 3.23 et 3.25 représentent toutes des coupes de hiérarchies construites avec un *a priori* spatial utilisant des fonctions faiblement sur-additives différentes dans le même ordre que celles utilisées dans la précédente expérience. La première colonne présente des coupes extraites de la hiérarchie construite avec l'énergie de Mumford Shah. Les coupes de cette colonne exhibent très peu de détails sur la voiture même pour une partition comportant 80 régions. La plupart des régions composant cette partition se concentrent sur les caractères portés par inscrits dans les deux panneaux. Il faudra alors explorer profondément la hiérarchie afin d'y découvrir l'existence et les structures de la voiture qui, de notre point de vue, est une information très pertinente dans l'image et devrait par conséquent figurer dans les hautes échelles. La dissemblance radiométrique assez faible entre les structures de la voiture et la route n'ont pas permis aux zones d'intérêt de persister à de hautes échelles. Par ailleurs, la différence radiométrique importante qui existe entre la couleur blanchâtre des caractères et les panneaux a permis à ces régions de survivre longtemps grâce à leur fort contraste. Les 4 séquences de partitions des colonnes (2), issues de l'énergie spatiale, exhibent un tout autre comportement. Le phénomène est inversé, les structures de la voiture survivent plus longtemps dans la hiérarchie tandis que les caractères sont absorbées plus rapidement par les panneaux. Une partition comportant 60 régions contient suffisamment de détails pour distinguer la voiture et quelques unes de ces structures. Néanmoins, comme précédemment, les contours de l'image sont de moins bonne qualité que ceux qui décrivent les partitions issues de la fonctionnelle de Mumford-Shah.

En analysant les partitions issues d'un potentiel pseudo-attractif, par rapport à celles issues du potentiel dit attractif, on notera que ces dernières sont qualitativement de moins bonnes qualité en terme de frontières et de contenu. Les panneaux de la partition de 80 régions sont divisés en régions aux frontières sinueuses ne contenant que très peu de régions représentant les lettres pourtant très contrastées par rapport à leur environnement. Les partitions de mêmes tailles issues des potentiels dit pseudo-attractifs exhibent des panneaux dont le fond n'est pas sursegmenté et les lettres même de petites tailles survivent à la fusion grâce à leur dissemblance par rapport au fond.

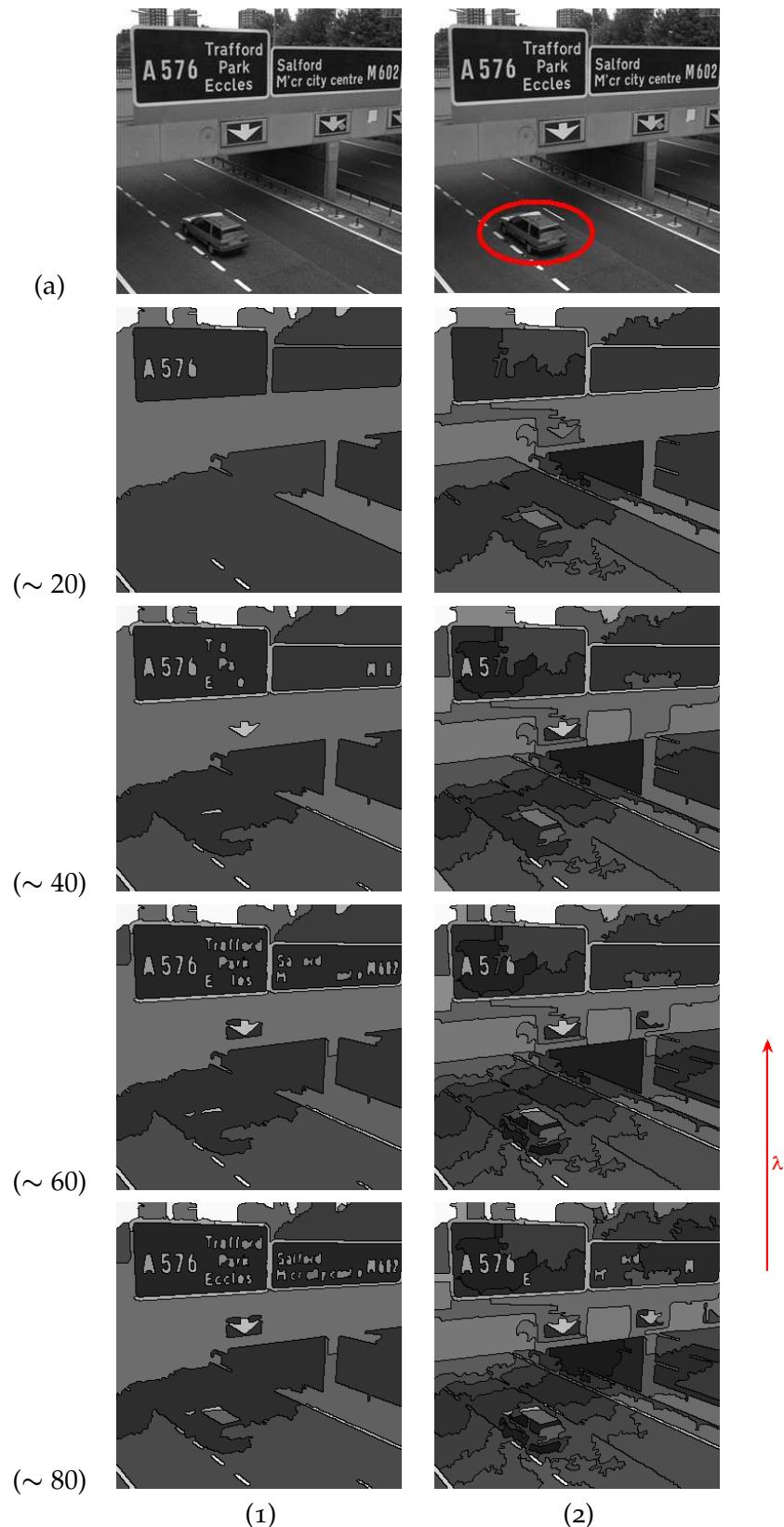


FIGURE 3.19 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle avec la fonction faiblement sur-additive utilisées  $|R| \cdot |B|$ .

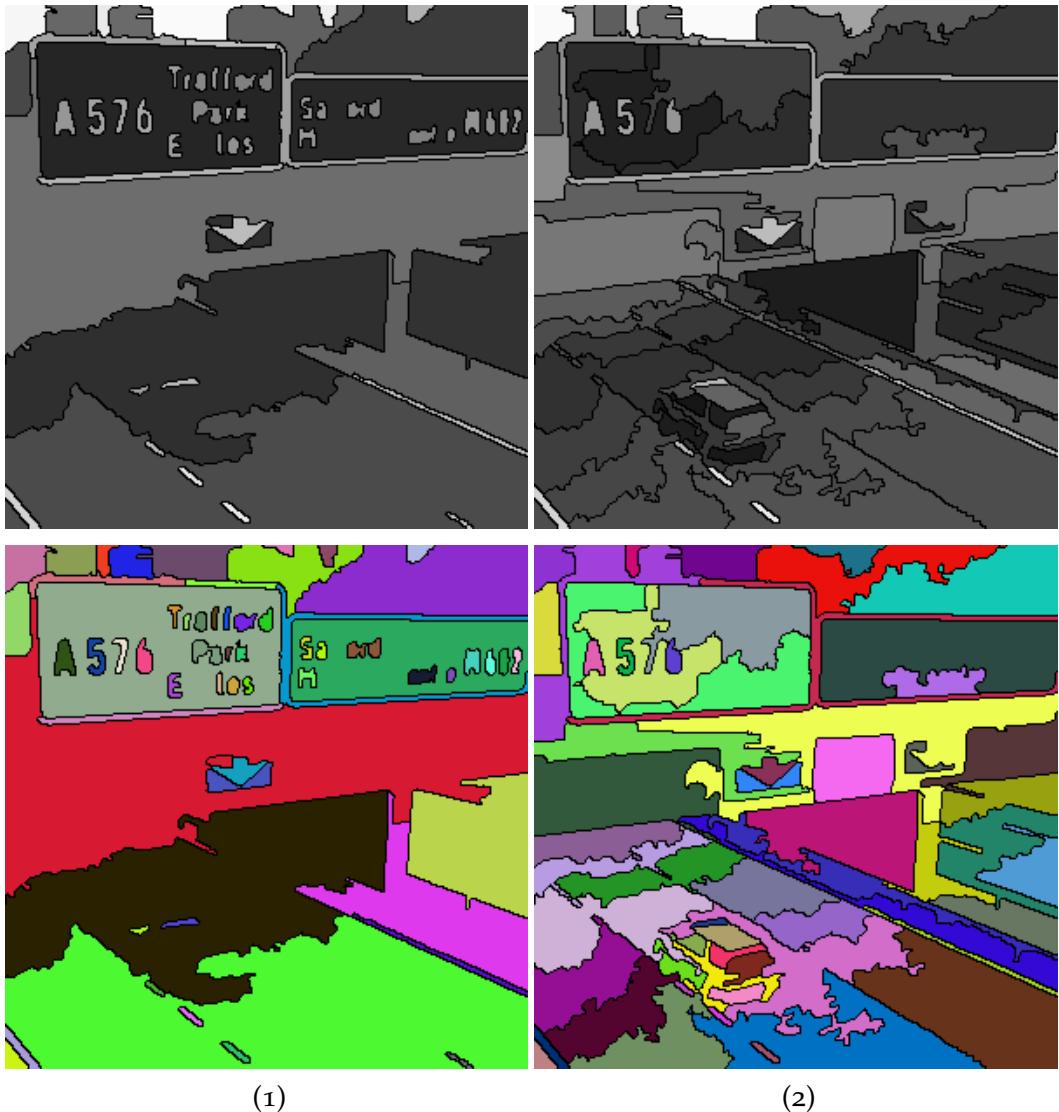


FIGURE 3.20 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R|/|B|}$ .

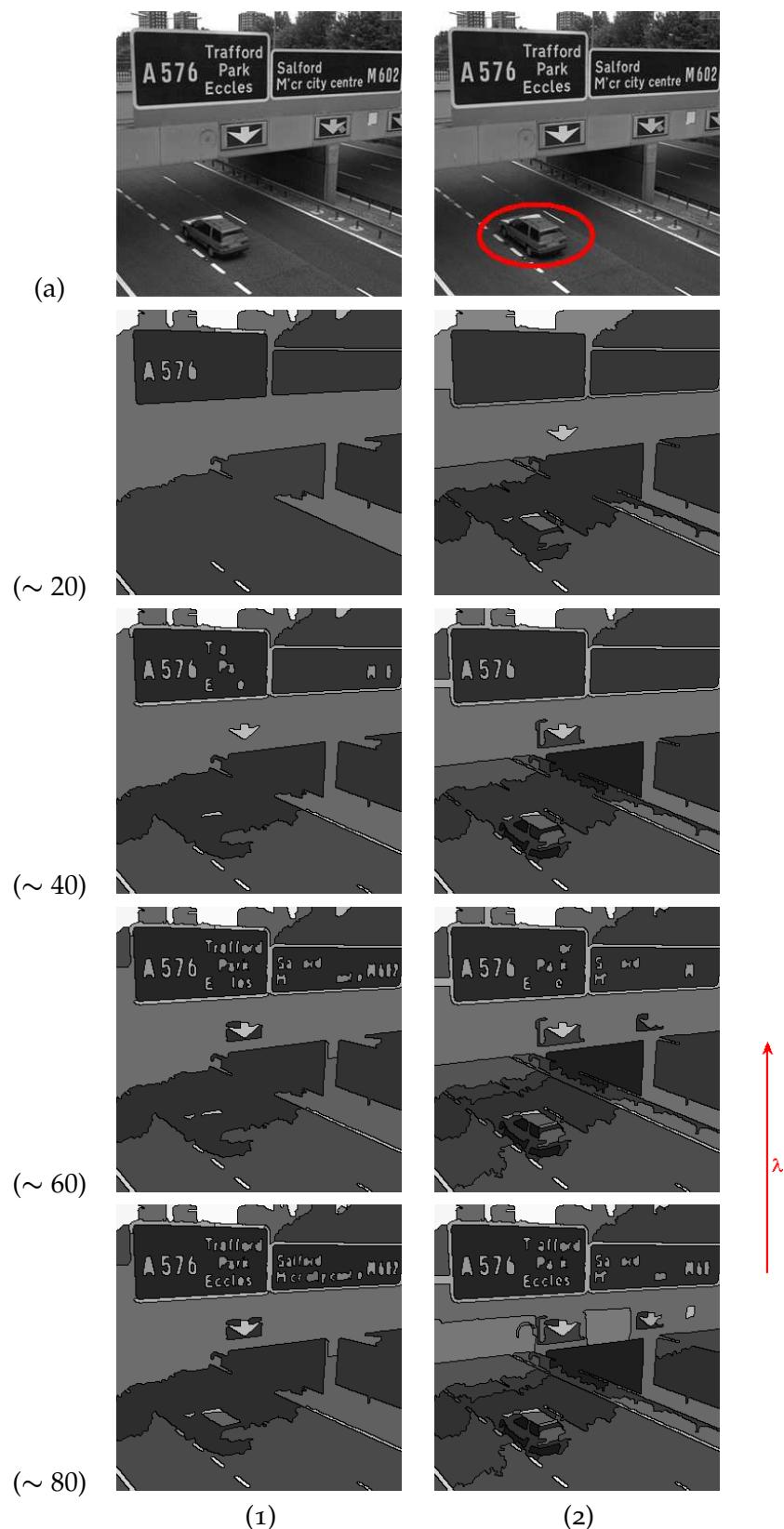


FIGURE 3.21 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle en utilisant la fonction faiblement sur-additive  $|R \cap B|/|R|$ .

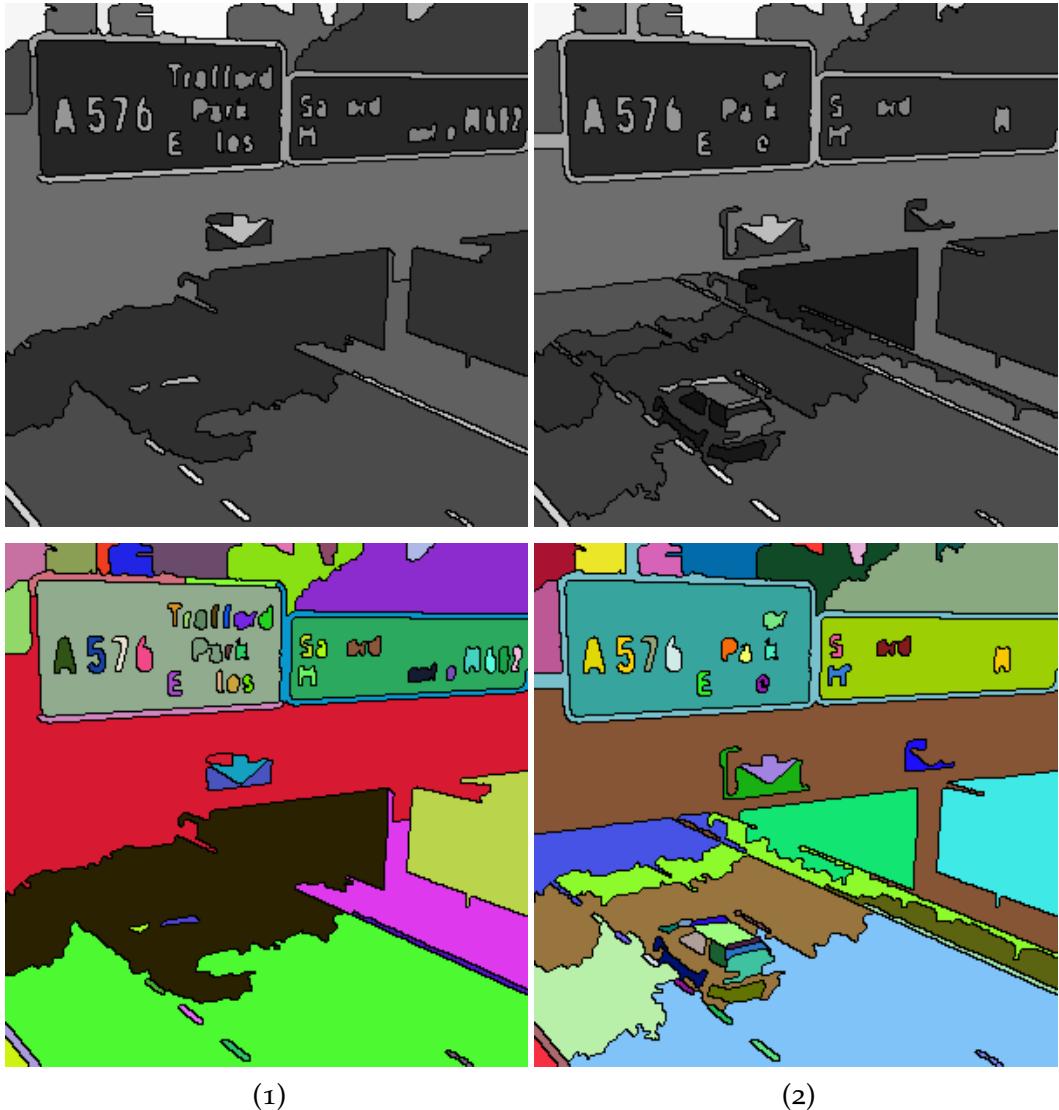


FIGURE 3.22 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R|/|B|}$ .

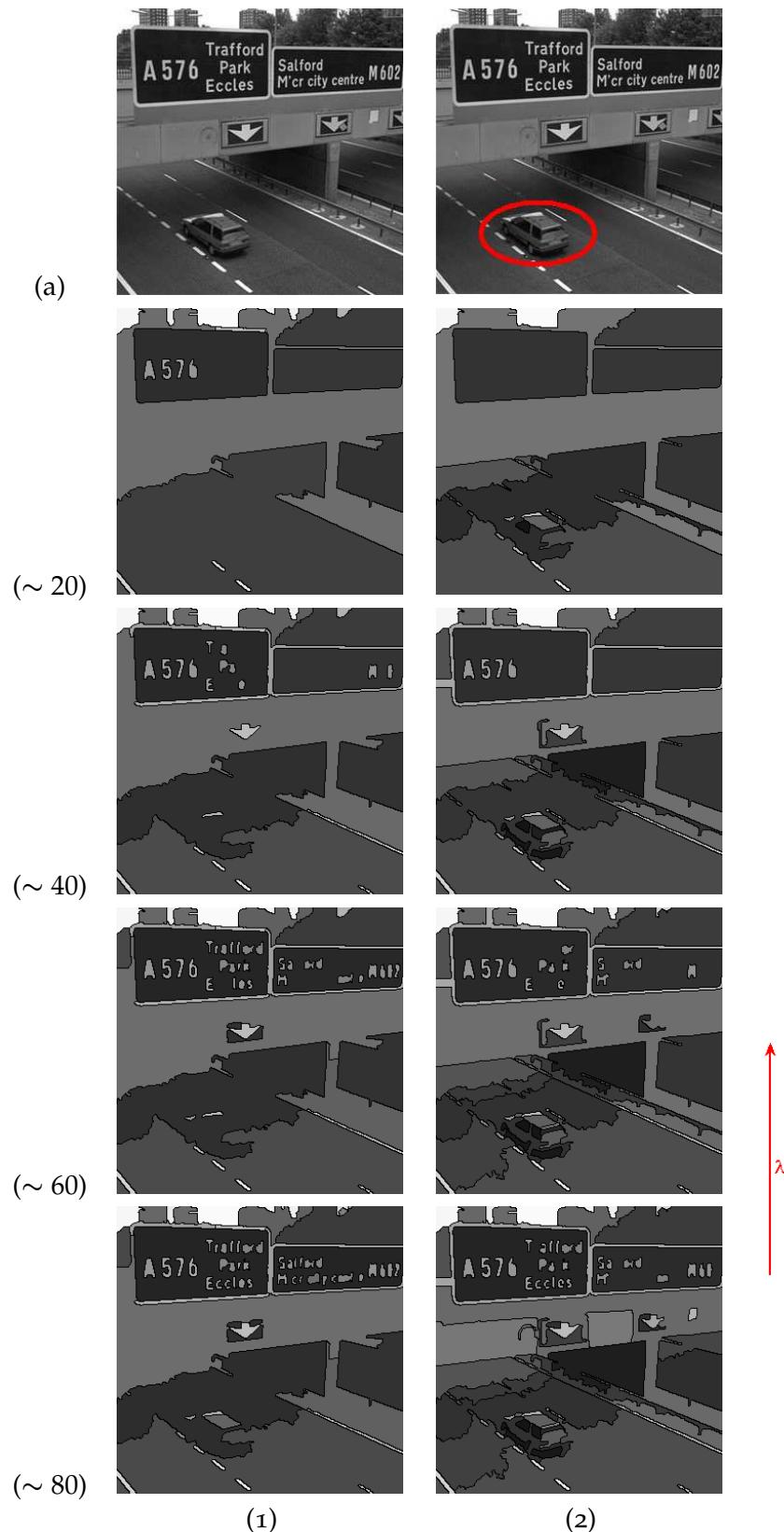


FIGURE 3.23 – Colonne (1) illustre l'ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle en utilisant la fonction faiblement sur-additive  $e^{|R \cap B| / |R|}$ .

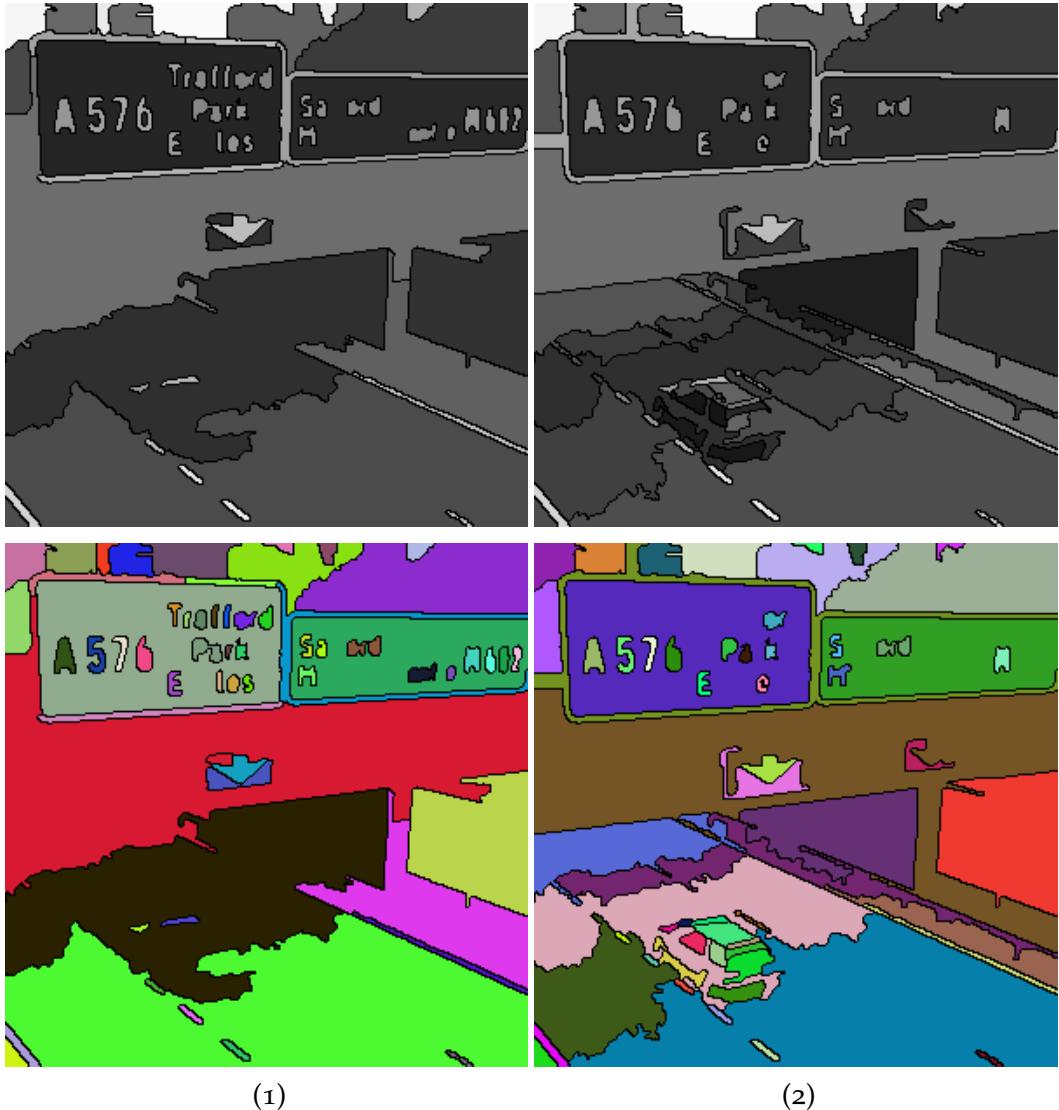


FIGURE 3.24 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R|/|B|}$ .

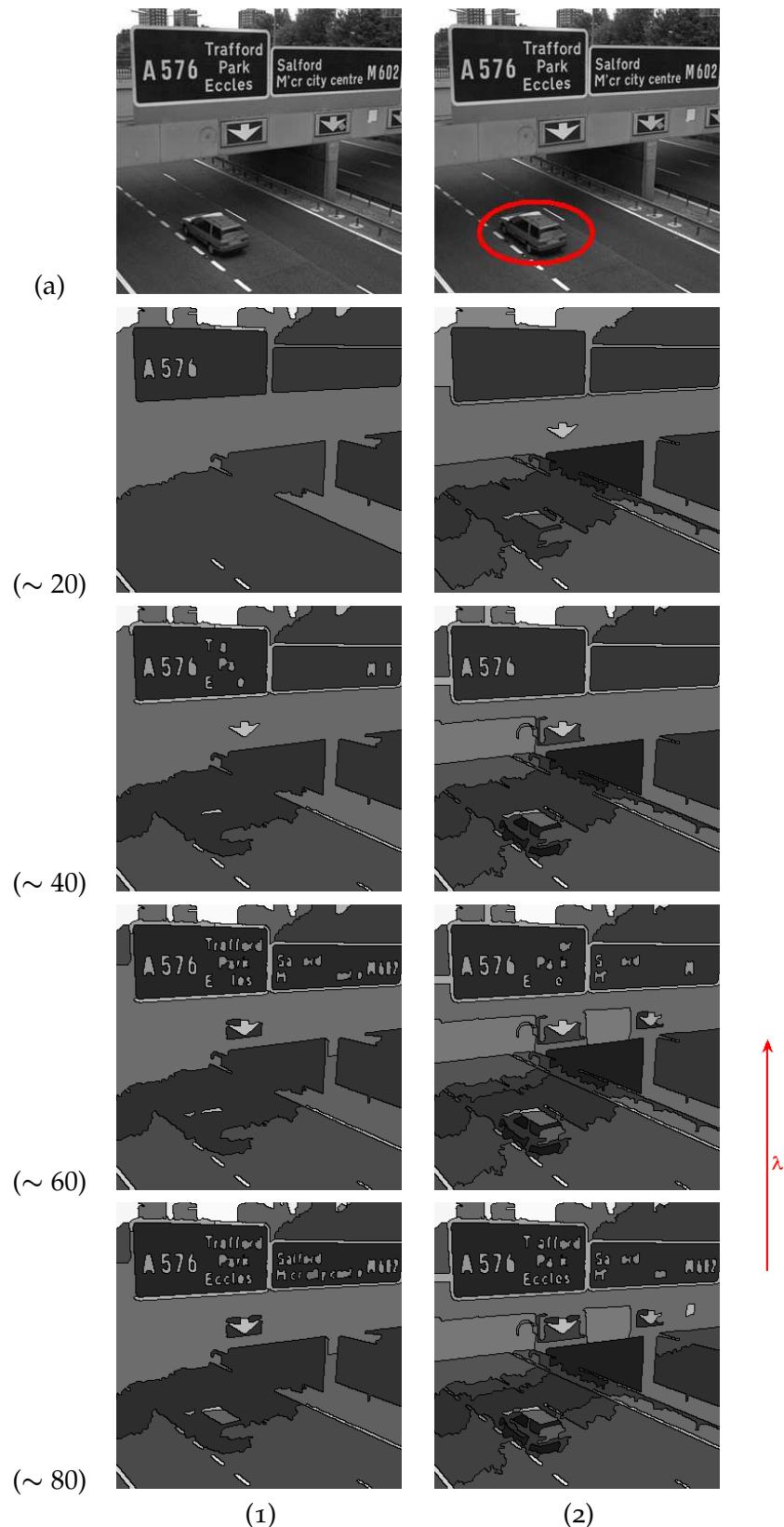


FIGURE 3.25 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle en utilisant la fonction faiblement sur-additive  $e^{|R|/|B|}$ .

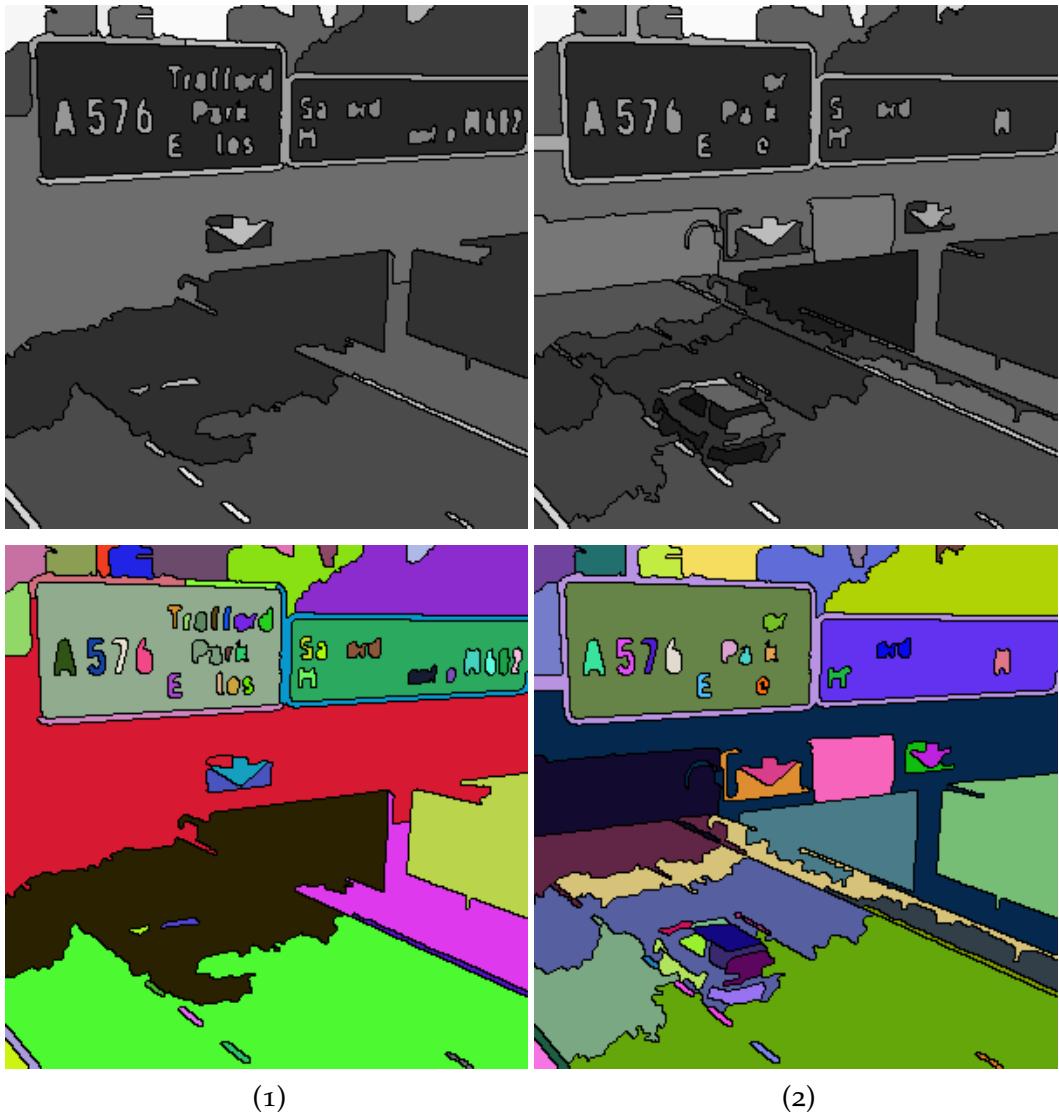


FIGURE 3.26 – Colonne (1) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie de Mumford-Shah. Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne). Colonne (2) : coupe de 60 régions extraite de la hiérarchie construite avec l'énergie spatiale mono-potentielle attractive ( $|R| \cdot |B|$ ). Visualisation en niveaux de gris (1ère ligne) et en couleur (2ème ligne) fonction faiblement sur-additive  $e^{|R|/|B|}$ .

### 3.5.1.2 Combinaison linéaire d'énergies sous-additives

Dans les précédentes expériences, l'énergie spatiale remplace terme pour terme l'énergie de régularisation de Mumford-Shah qui n'est autre que le périmètre. Ainsi la construction de la hiérarchie se base uniquement sur la radiométrie de l'image et l'*a priori* spatial. Aucune information « *a priori* » sur les contours n'a été introduite d'où l'aspect sinueux des frontières observées. L'énergie de régularisation de Mumford-Shah permet d'obtenir des frontières beaucoup plus lisses. Implicitement, une énergie de régularisation permet également de construire des régions de formes moins complexes.

Pour pallier cet inconvénient, nous proposons de combiner les deux énergies sous-additives afin d'obtenir des partitions avec des frontières plus régulières.

**Lemme 3.3** *Soient  $C_1, C_2, \dots, C_L$  une famille de régions sous-additives et soient  $\beta_1, \beta_2, \dots, \beta_L$  des nombres réels positifs tels que  $\sum_{i=1}^L \beta_i = 1$  alors toute combinaison linéaire des énergies  $C = \sum_{i=1}^L \beta_i \cdot C_i$  est une énergie sous-additive.*

Dans notre cas la nouvelle énergie s'écrit  $\beta \mathbf{S} + (1 - \beta) \mathbf{L}$  tel que  $\beta \in ]0, 1[$ . Toute la difficulté réside dans le choix du paramètre  $\beta$  qui jouera un rôle de pondération entre les deux énergies.

Nous avons conduit différentes expérimentations sur des images en modifiant le paramètre  $\beta$  légèrement et de manière croissante. Cependant, nous présenterons ici les résultats obtenus pour les valeurs  $\beta = 0.2, 0.5, 0.8$  car elles démontrent clairement l'évolution de la qualité des partitions obtenues pour ces différentes valeurs.

La première colonne des figures 3.27, 3.28 et 3.29 montre la série de coupes obtenues pour l'énergie de Mumford-Shah classique. Les trois dernières colonnes de ces figures sont des coupes acquises en combinant l'énergie spatiale et la longueur du périmètre pour différentes valeurs du coefficient  $\beta = 0.2, 0.5, 0.8$ . L'énergie spatiale mono-potentielle utilisée dans la deuxième colonne est basée sur la fonction faiblement sur-additive  $|R \cap B| / |R|$  celle de la colonne (3) est basée sur la fonction  $e^{|R \cap B| / |R|}$  et la dernière colonne utilise la fonction  $e^{|R| / |B|}$ .

Dans la colonne (2) de la figure 3.27, le poids donné à l'énergie spatiale est moins important que celui donné à l'énergie de régularisation de Mumford-Shah. Les partitions de 20, 40, 60 et 80 régions comportent peu de différences par rapport aux partitions obtenues sans *a priori* spatial. On retrouve quasiment le même ordre de fusion des régions dans les 4 colonnes. Les structures de la voiture sont absorbées plus rapidement mais les frontières obtenues avec cette combinaison convexe des deux énergies sous-additives permet d'obtenir des frontières plus lisses. Les partitions de 100 régions issues de l'énergie spatiale montrent quand à elles plus de structures dans la zone d'intérêt à l'inverse des partitions de même niveau obtenues sans *a priori* spatial. On notera que les frontières sont plus régulières par rapport à celles obtenues en utilisant uniquement l'énergie spatiale.

La figure 3.28 exhibe dans ses trois dernières colonnes les partitions obtenues avec un coefficient de convexité  $\beta = 0.5$ . Ce choix pondère l'action des deux énergies d'*a priori* de manière équitable. L'énergie de régularisation de Mumford-Shah joue pleinement son rôle en nous offrant des coupes avec des contours réguliers. Par ailleurs, l'énergie spatiale permet de protéger plus efficacement les zones en interaction avec le blob. Les structures qui nous intéressent résistent mieux au processus de fusion. Elles sont certes, situées plus bas mais les partitions de 80 et 100

régions exhibent des détails de la voiture qui sont absents dans les partitions de Mumford-Shah.

Pour terminer, nous avons porté le paramètre de pondération  $\beta$  à 0.8 donnant ainsi un rôle prépondérant à l'énergie spatiale dans la conduite du processus de fusion. Ce choix se traduit par des partitions dont les frontières sont plus sinueuses que ses homologues obtenues avec un coefficient plus faible. La contrepartie se traduit par des partitions exhibant une description plus détaillée de la zone étudiée. On dira que c'est le prix à payer si l'on souhaite obtenir des partitions de plus faibles cardinalités et contenant des objets d'intérêt complexes qui auraient du être absorbés par leur environnement à des niveaux plus bas si l'on n'avait pas inclus un *a priori* spatial.

### 3.5.1.3 Energie spatiale multi-potentielle

Les figures 3.30, 3.31 et 3.32 représentent les partitions obtenues en utilisant une énergie spatiale multi-potentielle. La première colonne de chaque figure représente des coupes obtenues par l'énergie de Mumford-Shah. La deuxième colonne celles construites avec une énergie spatiale mono-potentielle et la dernière colonne celles construites avec une énergie spatiale multi-potentielle. La comparaison entre la colonne (1) et (3) nous permet de constater que l'énergie multi-potentielle agit de la même manière que l'énergie mono-potentielle. Les structures incluses dans le blob sont préservées durant le processus de fusion plus longtemps que dans une hiérarchie construite sans *a priori* spatial. Par contre, la comparaison entre les résultats obtenus avec l'énergie spatiale mono-potentielle et multi-potentielle ne montre pas une différence majeure.

L'énergie multi-potentielle d'une région, relativement à un blob  $B$ , inclut le potentiel des régions voisines tandis que l'énergie mono-potentielle ignore les potentiels des régions adjacentes. La prise en compte du voisinage permet-elle de mieux préserver les structures ciblées ? Les partitions de 20 régions montrent que globalement les structures sont mieux préservées dans le cas mono-potentiel que dans le cas multi-potentiel. La prise en compte du voisinage d'une région, incluse dans un blob, induit un comportement de fusion local variable. Si la région est entourée de régions incluses dans le blob alors sa fusion est accélérée. A l'inverse, si la région possède un voisinage de régions qui se trouvent à l'extérieur du blob alors la région sera absorbée par ces régions et par conséquent certaines structures d'intérêt seront fusionnées plus rapidement que dans le cas mono-potentiel. Dans nos expérimentations, nous avons relevé des différences très minimes de l'ordre de deux ou trois régions.

Il s'avère donc que la prise en compte du voisinage de la région n'a pas d'effet significatif dans les résultats.

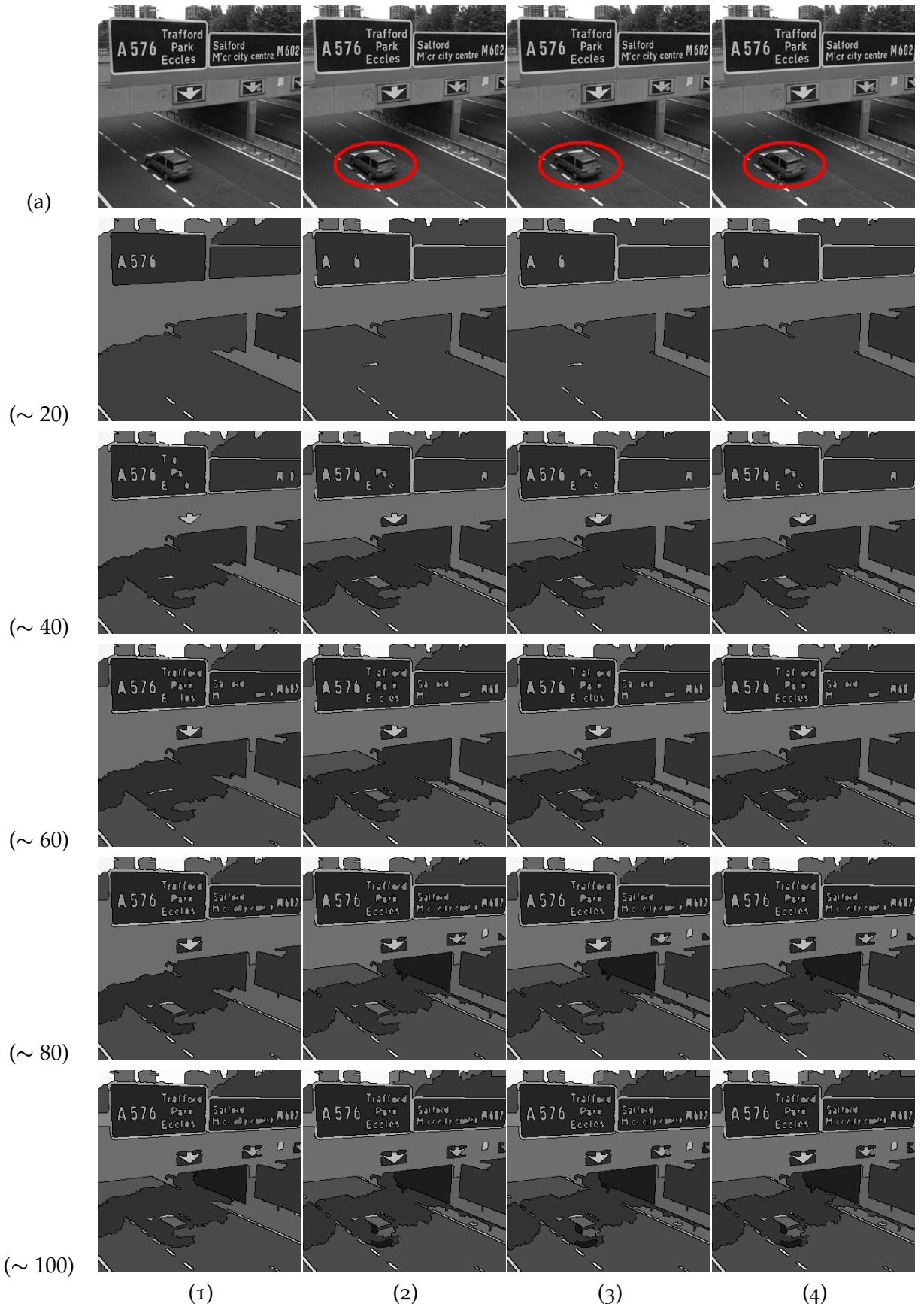


FIGURE 3.27 – Colonne (1) : coupes d'un ensemble-échelle construit avec l'énergie de Mumford-Shah. Colonnes (2), (3) et (4) représentent les ensembles-échelles obtenus avec l'énergie  $\beta \mathbf{S} + (1 - \beta) \mathbf{L}$  pour  $\beta = 0.2$  et pour différentes fonctions faibl. sur-additives ( $|R \cap B|/|R|$ ,  $e^{|R \cap B|/|R|}$  et  $e^{|R|/|B|}$ ).

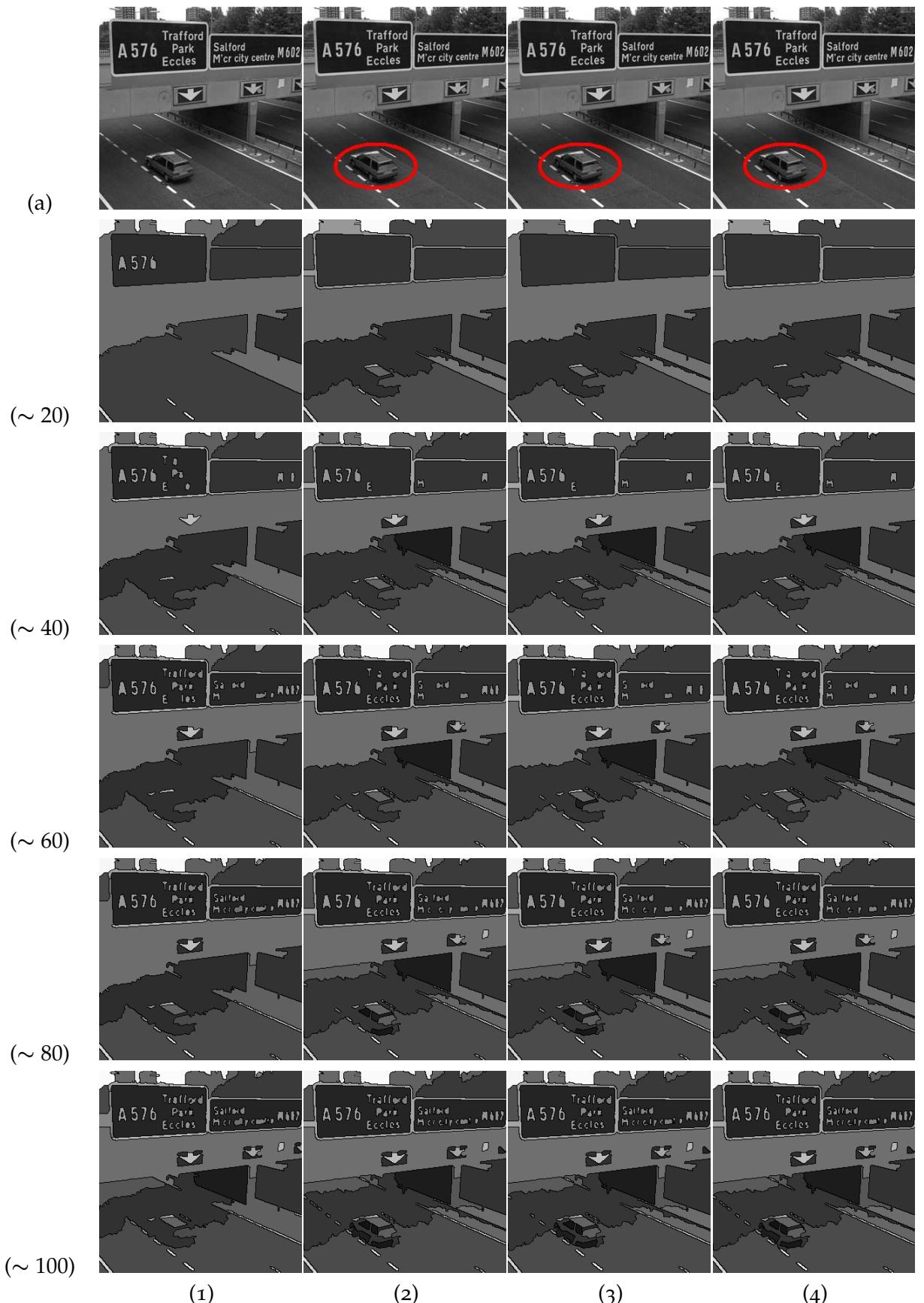


FIGURE 3.28 – Colonne (1) : coupes d'un ensemble-échelle construit avec l'énergie de Mumford-Shah. Colonnes (2) et (3) et (4) représentent les ensembles-échelles obtenus avec l'énergie  $\beta \mathbf{S} + (1 - \beta) \mathbf{L}$  pour  $\beta = 0.5$  et pour différentes fonctions faibl. sur-additives ( $|R \cap B| / |R|$ ,  $e^{|R \cap B| / |R|}$  et  $e^{|R| / |B|}$ ).

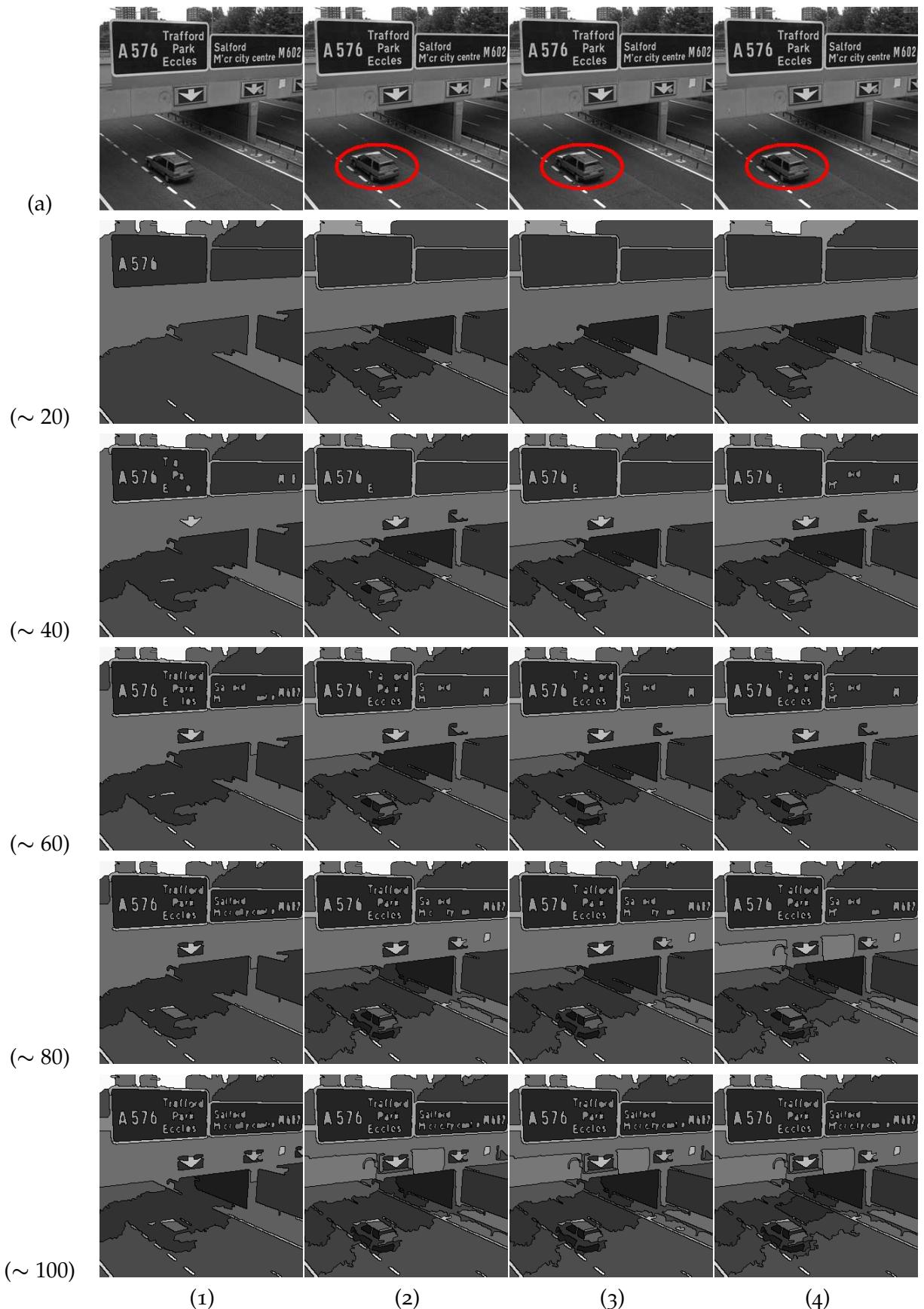


FIGURE 3.29 – Colonne (1) : coupes d'un ensemble-échelle construit avec l'énergie de Mumford-Shah. Colonnes (2) et (3) et (4) représentent les ensembles-échelles obtenus l'énergie  $\beta \mathbf{S} + (1 - \beta) \mathbf{L}$  pour  $\beta = 0.8$  et pour différentes fonctions faibl. sur-additives ( $|R \cap B|/|R|$ ,  $e^{|R \cap B|/|R|}$  et  $e^{|R|/|B|}$ ).

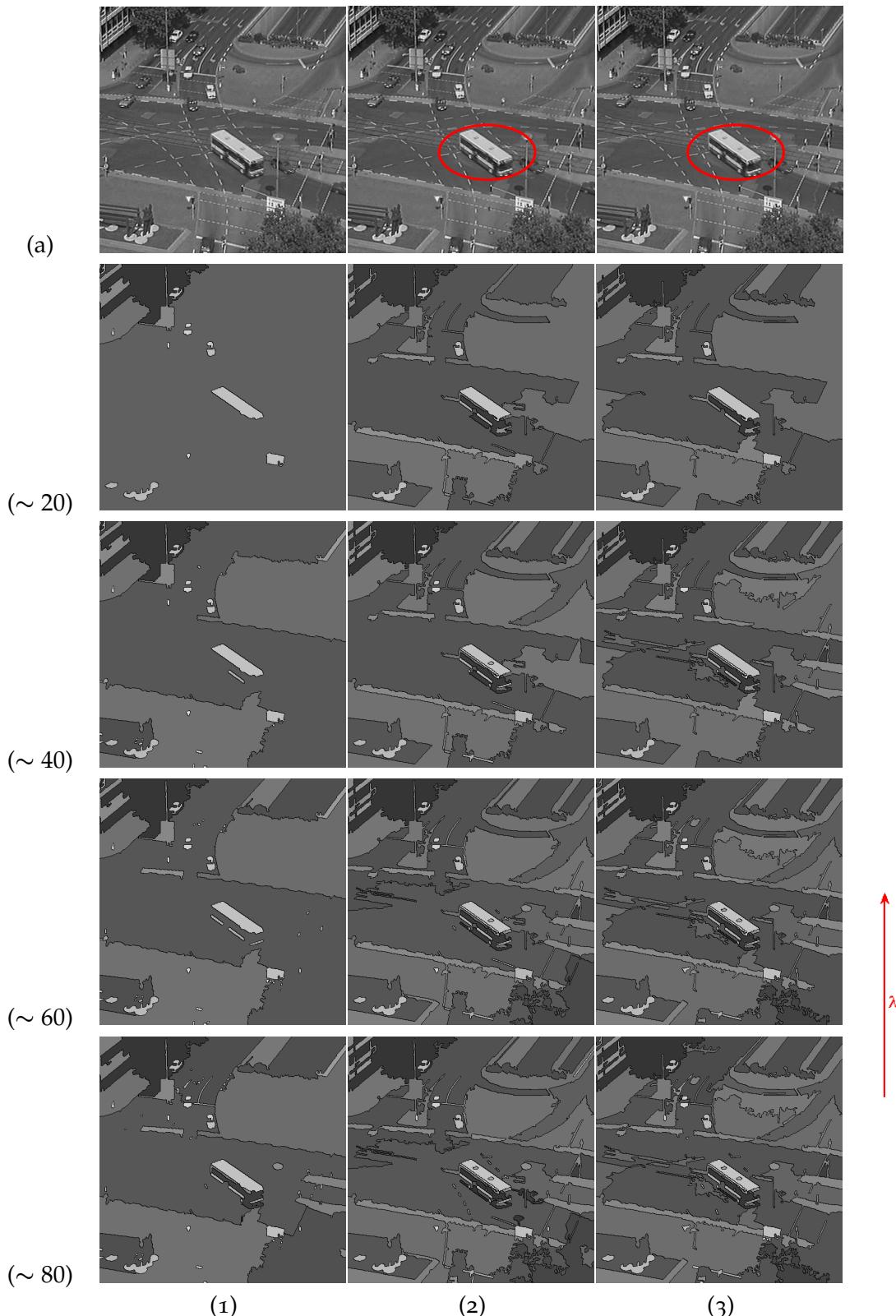


FIGURE 3.30 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (pseudo-attractrice) issue de la fonction faiblement sur-additive  $|R \cap B|/|R|$ . La dernière colonne utilise une énergie multi-potentielle.

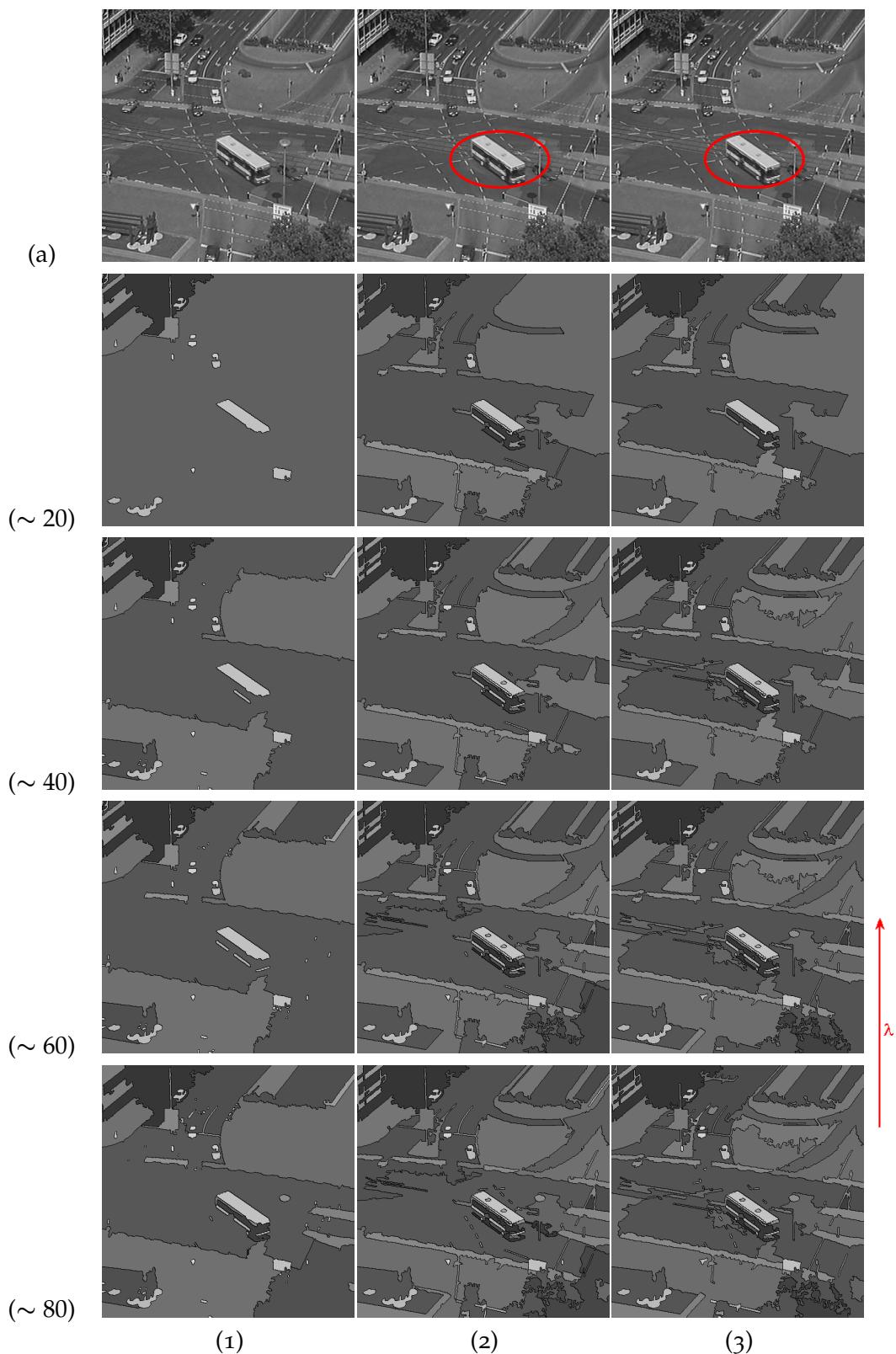


FIGURE 3.31 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (pseudo-attractrice) issue de la fonction faiblement sur-additive  $e^{|R \cap B|/|R|}$ . La dernière colonne utilise l'énergie multi-potentielle.

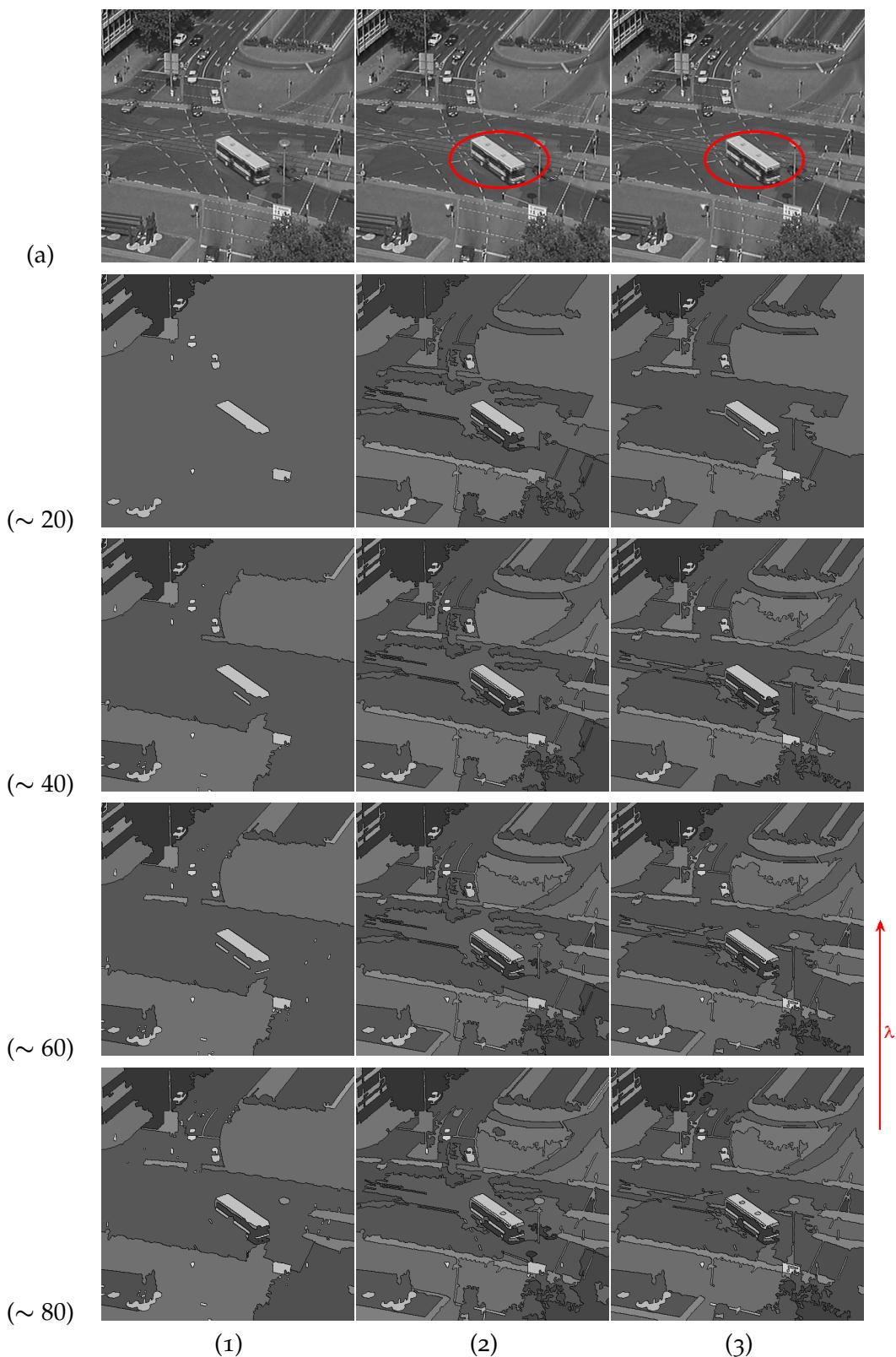


FIGURE 3.32 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (pseudo-attractrice) issue de la fonction faiblement sur-additive  $e^{|R|}/|B|$ . La dernière colonne utilise l'énergie multi-potentielle.

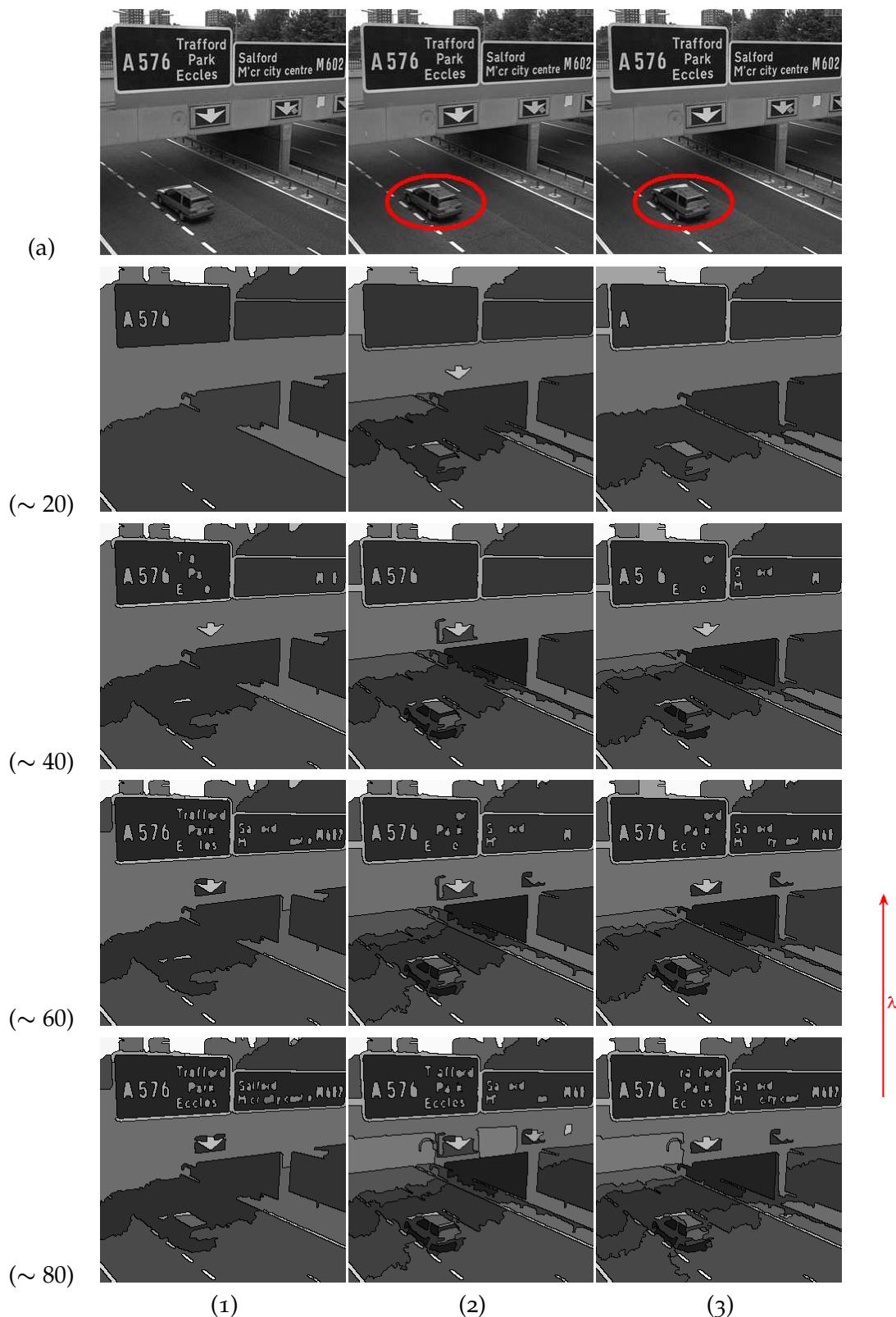


FIGURE 3.33 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (pseudo-attractrice) issue de la fonction faiblement sur-additive  $|R \cap B| / |R|$ . La dernière colonne utilise l'énergie multi-potentielle.

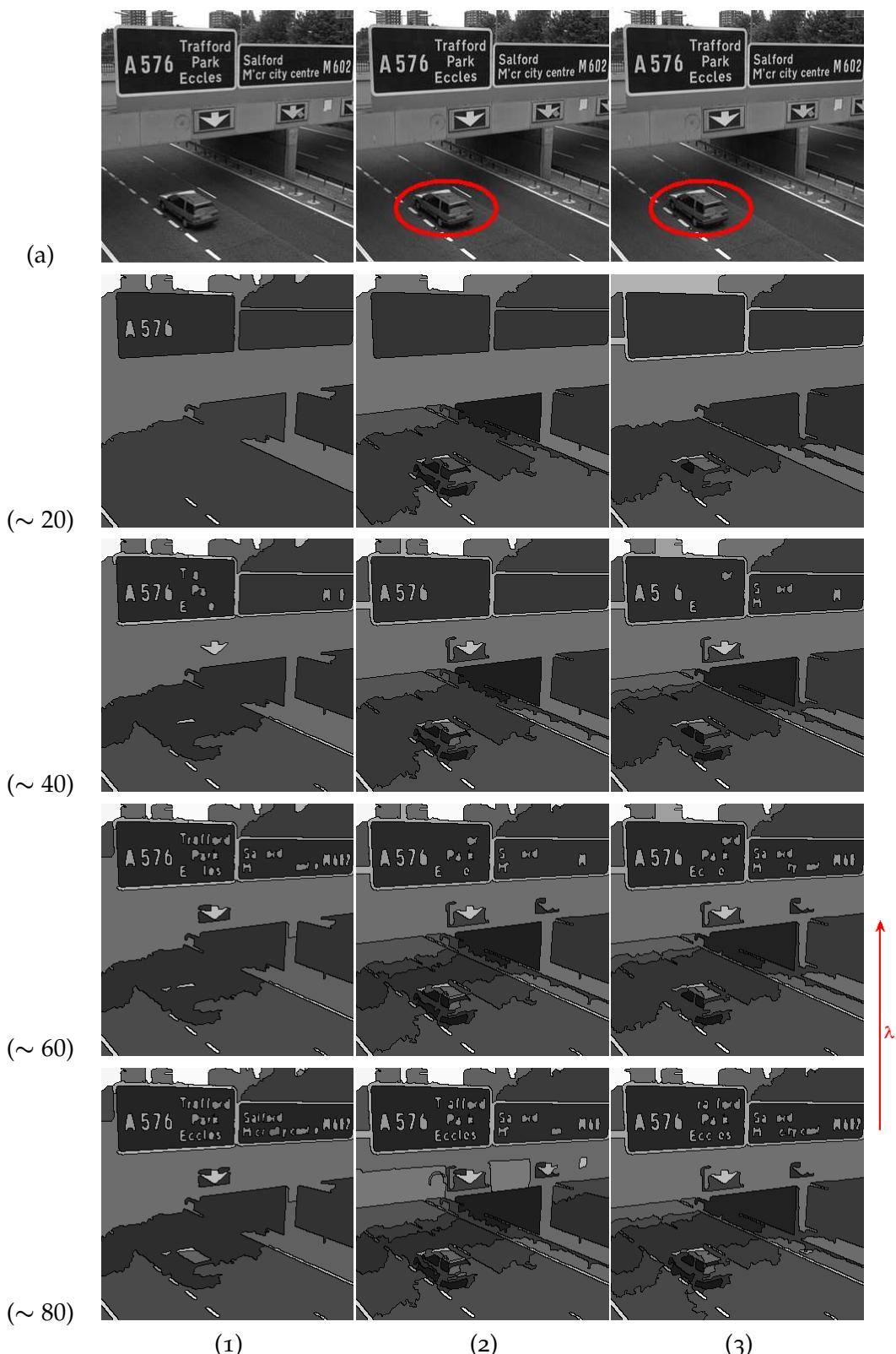


FIGURE 3.34 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (pseudo-attractrice) issue de la fonction faiblement sur-additive  $e^{|R \cap B|/|R|}$ . La dernière colonne utilise l'énergie multi-potentielle.

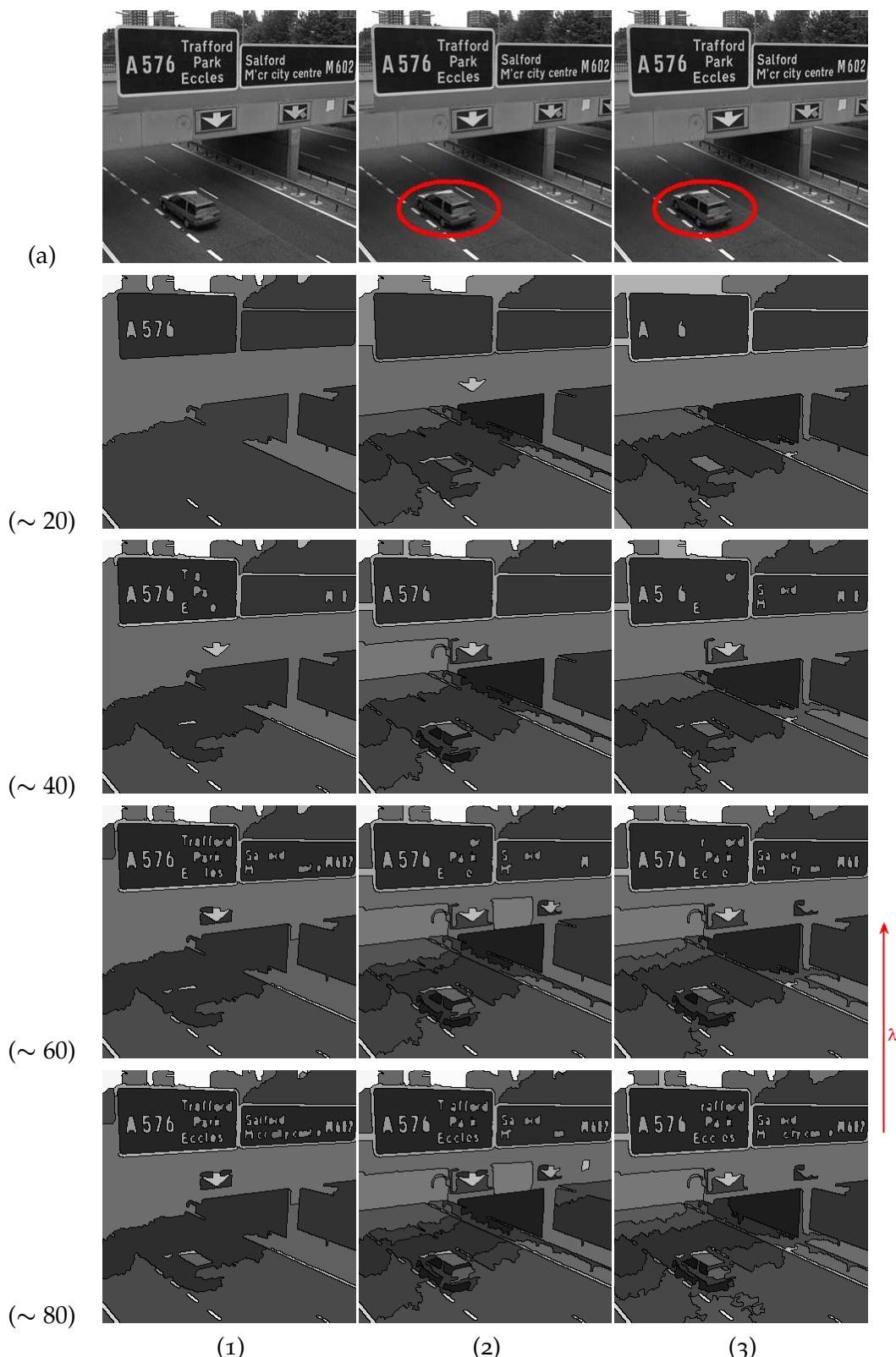


FIGURE 3.35 – Colonne (1) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : ensemble-échelle obtenu avec l'énergie spatiale mono-potentielle (pseudo-attractrice) issue de la fonction faiblement sur-additive  $e^{|R|/|B|}$ . La dernière colonne utilise l'énergie multi-potentielle.

### 3.5.1.4 Energie spatiale bi-potentielle

Pour terminer le cas de figure où l'*a priori* spatial porte sur un seul blob, nous allons examiner le comportement d'une énergie bi-potentielle lors de la construction hiérarchique.

On utilisera un bi-potentiel multiplicatif construit à partir de la distance métrique de lien minimum :

$$d_{min}(R, B) = \min_{x \in R} \{d(x, c_B)\},$$

tel que  $d$  est la distance euclidienne classique.

Cette distance sera combinée avec les fonctions supra-additive suivantes :

- $\delta_s(R, B) = \frac{1}{|R||B|}$ ,
- $\delta_s(R, B) = \frac{1}{(1+|R \cap B|)}$ ,
- $\delta_s(R, B) = e^{-\frac{|R|}{|B|}}$ ,
- $\delta_s(R, B) = e^{-\frac{|R \cap B|}{|R|}}$ ,

Nous ferons le parallèle entre les coupes obtenues avec une énergie mono-potentielle (illustrées dans les colonnes 1) et celles issues de l'énergie bi-potentielle (représentées par les colonnes 2). La fonction supra-additive utilisée dans l'énergie bi-potentielle sera l'inverse de la fonction faiblement sur-additive utilisée dans l'énergie mono-potentielle.

L'analyse de la figure 3.36 laisse apparaître une certaine différence entre les coupes de même taille. Les partitions de 80 régions décrivent l'objet d'intérêt de manière quasi-identique. Néanmoins, la partition de 40 régions construite avec l'énergie bi-potentielle décrit la voiture avec plus de détails que son homologue obtenue avec l'énergie mono-potentielle. L'énergie bi-potentielle impose une plus forte pénalité entre régions appartenant à un blob par rapport à une énergie mono-potentielle. Les structures composant le blob seront alors plus difficile à fusionner et persisteront plus longuement dans la hiérarchie.

La différence est plus remarquable dans la figure 3.37 où les coupes issues de l'énergie bi-potentielle décrivent l'objet avec plus de détails que les coupes issues des autres énergies bi-potentielles. La fonction supra-additive utilisée est  $1/(|R \cap B| + 1)$  combinée avec la distance du lien minimum. L'exploration en profondeur de la hiérarchie à partir du sommet donne l'impression que l'on se focalise dans un premier temps uniquement sur l'objet circonscrit. A chaque tranche de 20 régions on voit apparaître davantage de structures de la zone étudiée par rapport aux coupes extraites de hiérarchies construites avec une énergie mono ou bi-potentielle quelconque. Le reste des zones d'intérêt qui compose l'image est très vite absorbé. La fusion paraît asynchrone puisqu'une partition de 80 régions sur-segmente l'objet d'intérêt tout en sous-segmentant le reste des régions d'intérêt de l'image. Un déséquilibre se crée entre l'objet étudié et le reste des objets composant la scène.

Les coupes issues des figures 3.38 et 3.39 ont quand à elles un comportement similaire aux coupes construites avec la fonction supra-additive  $1/|R||B|$ . Les partitions de 40 régions exhibent plus de détails que leurs consœurs issues d'une énergie mono-potentielle. Certes, la région étudiée est mieux représentée mais cela n'empêche pas l'exploration en profondeur de la hiérarchie d'exhiber d'autres détails de l'image qui se trouvent à l'extérieur du blob.

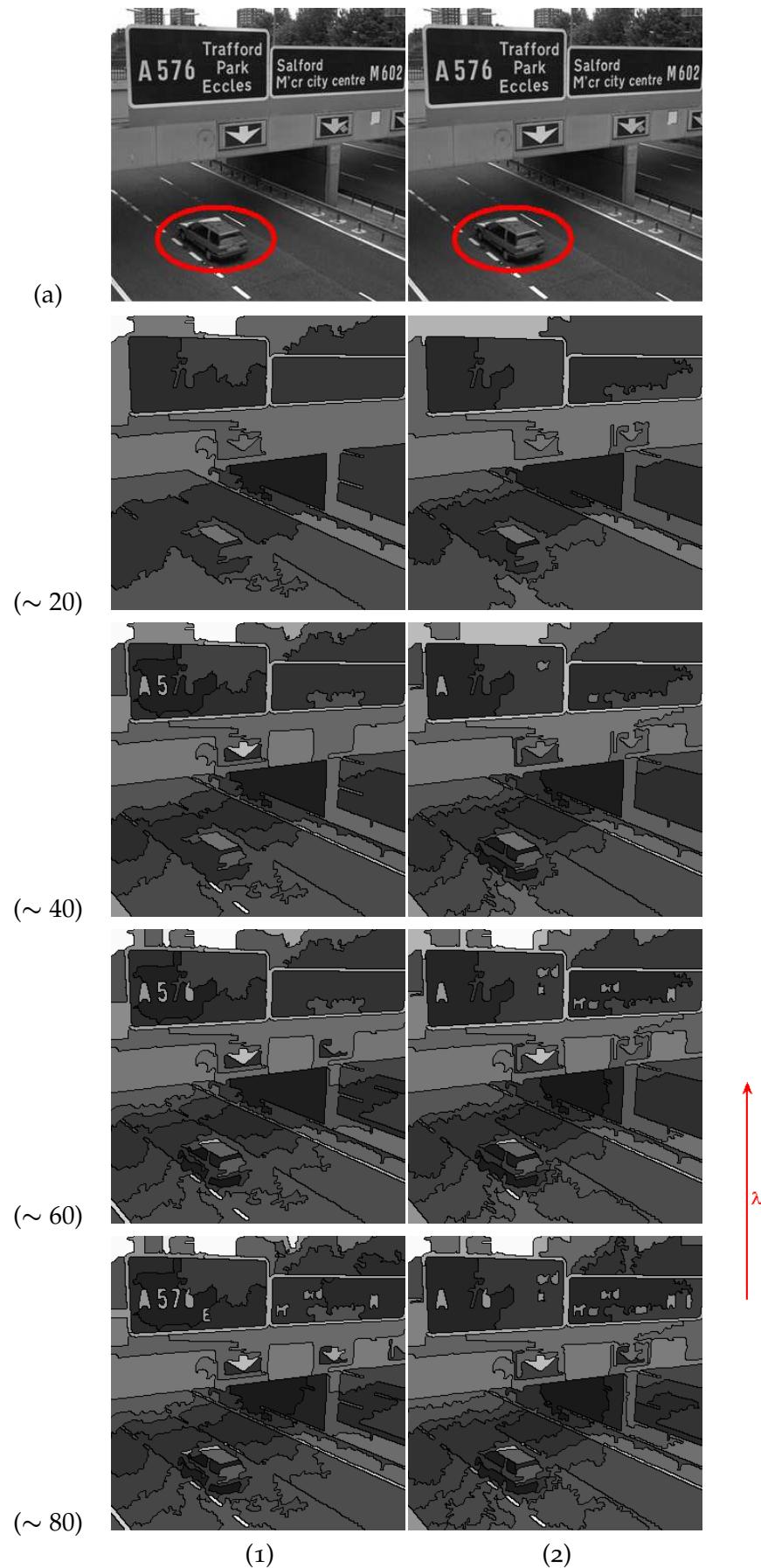


FIGURE 3.36 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $|R| \cdot |B|$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $|R| \cdot |B|$ .

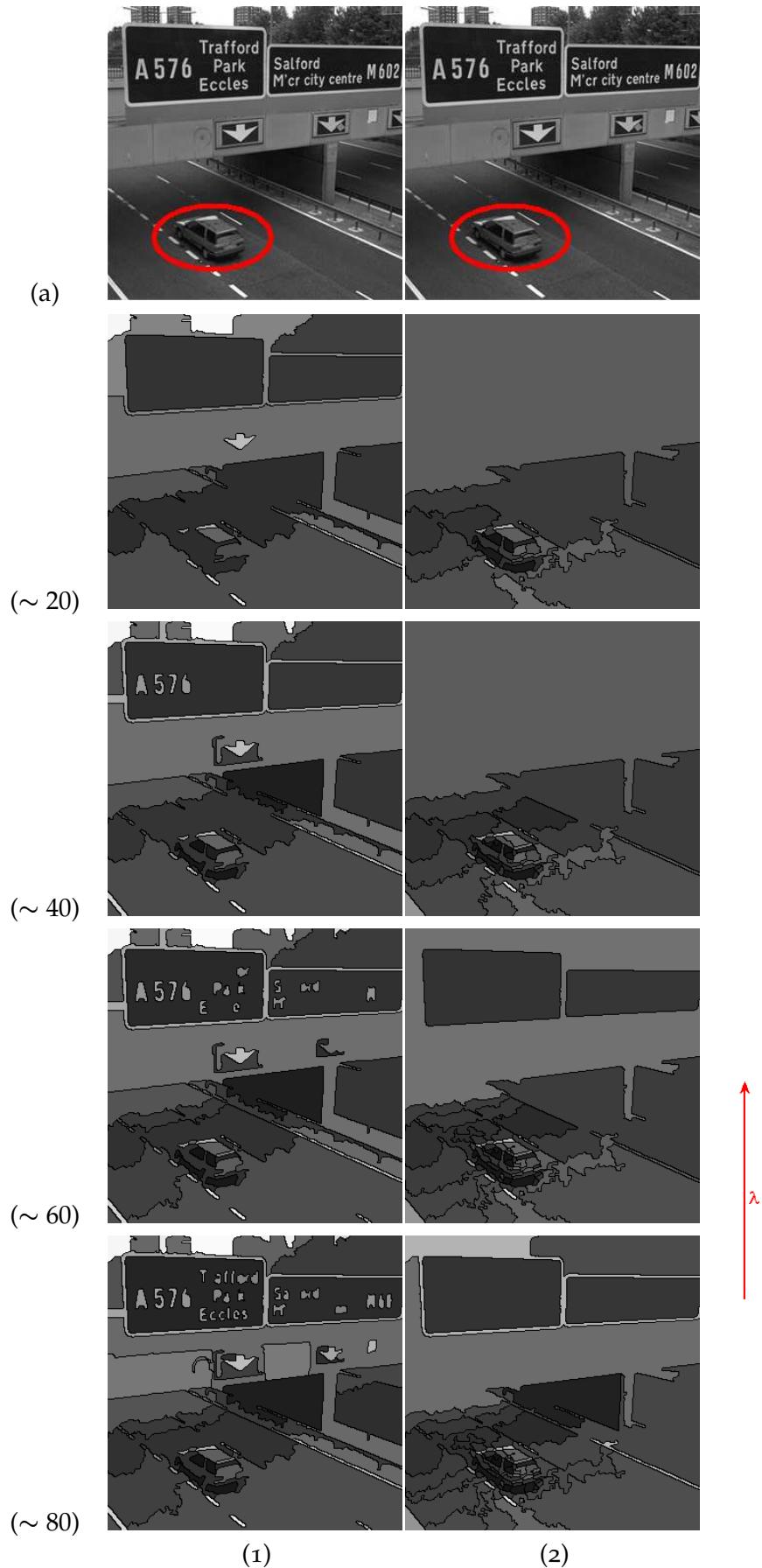


FIGURE 3.37 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $|R \cap B| / |R|$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $1 / (1 + |R \cap B|)$ .

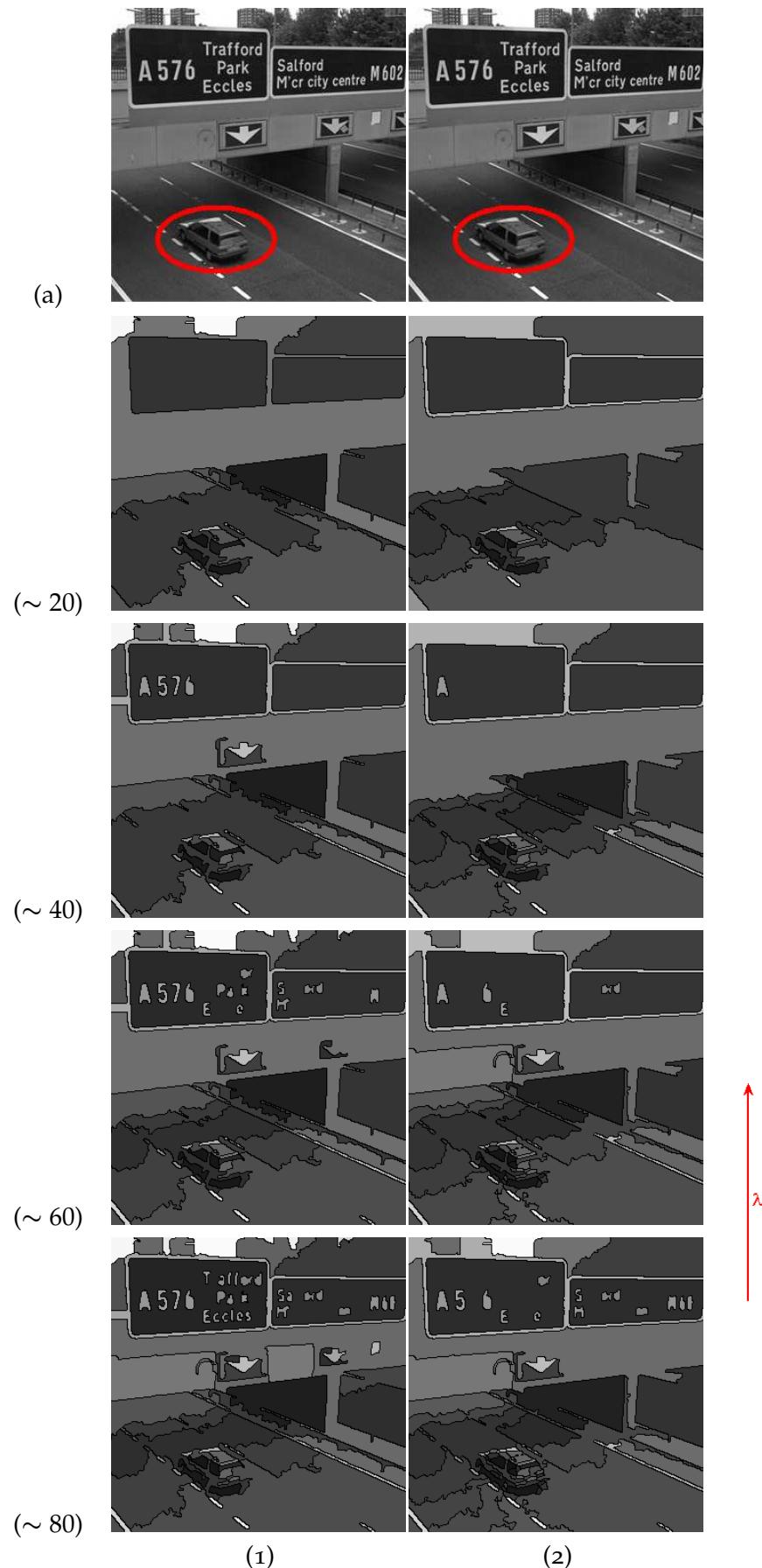


FIGURE 3.38 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $e^{|R \cap B| / |R|}$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $e^{-|R \cap B| / |R|}$ .

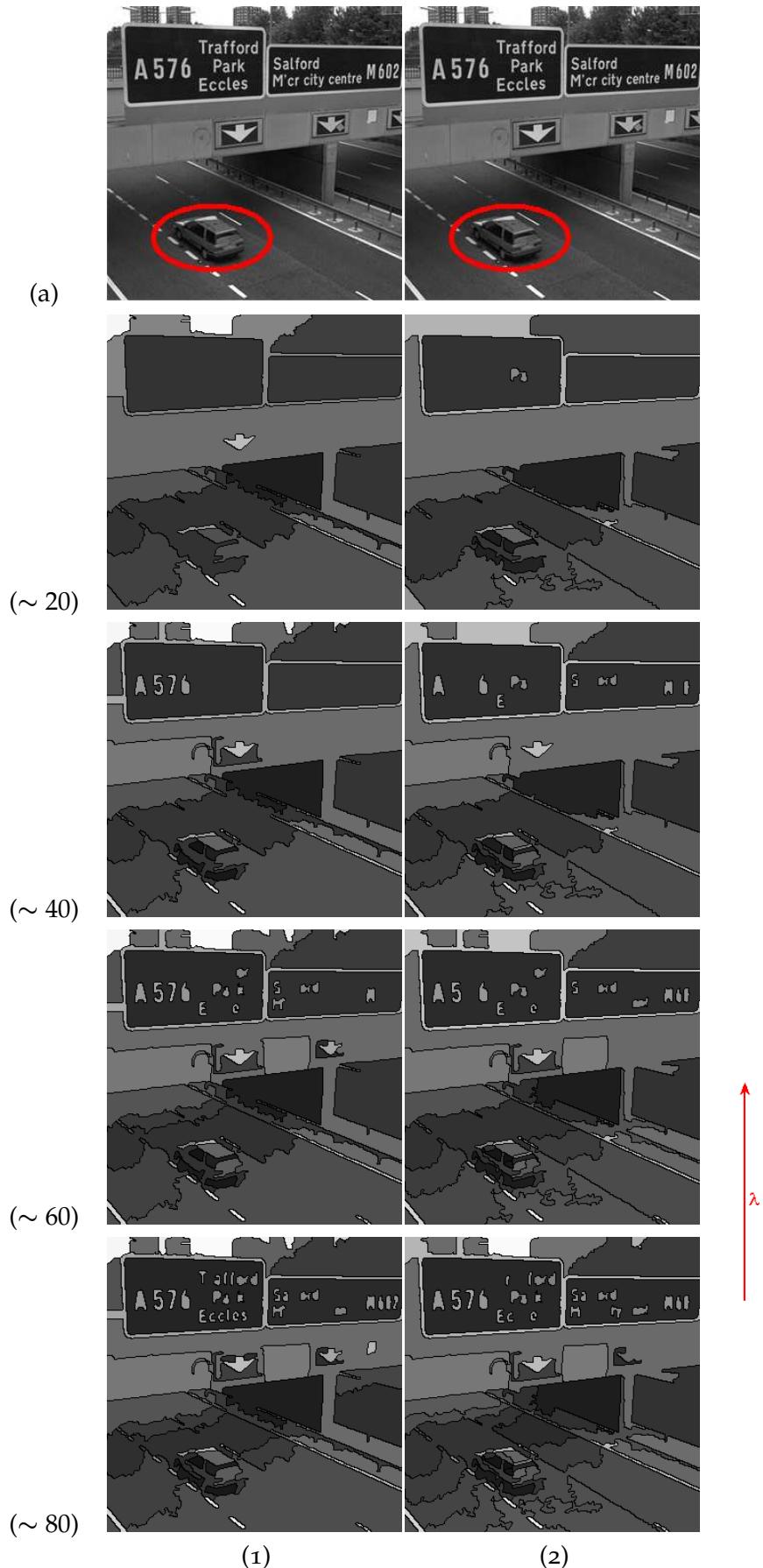


FIGURE 3.39 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $e^{|R|/|B|}$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $e^{-|R|/|B|}$ .

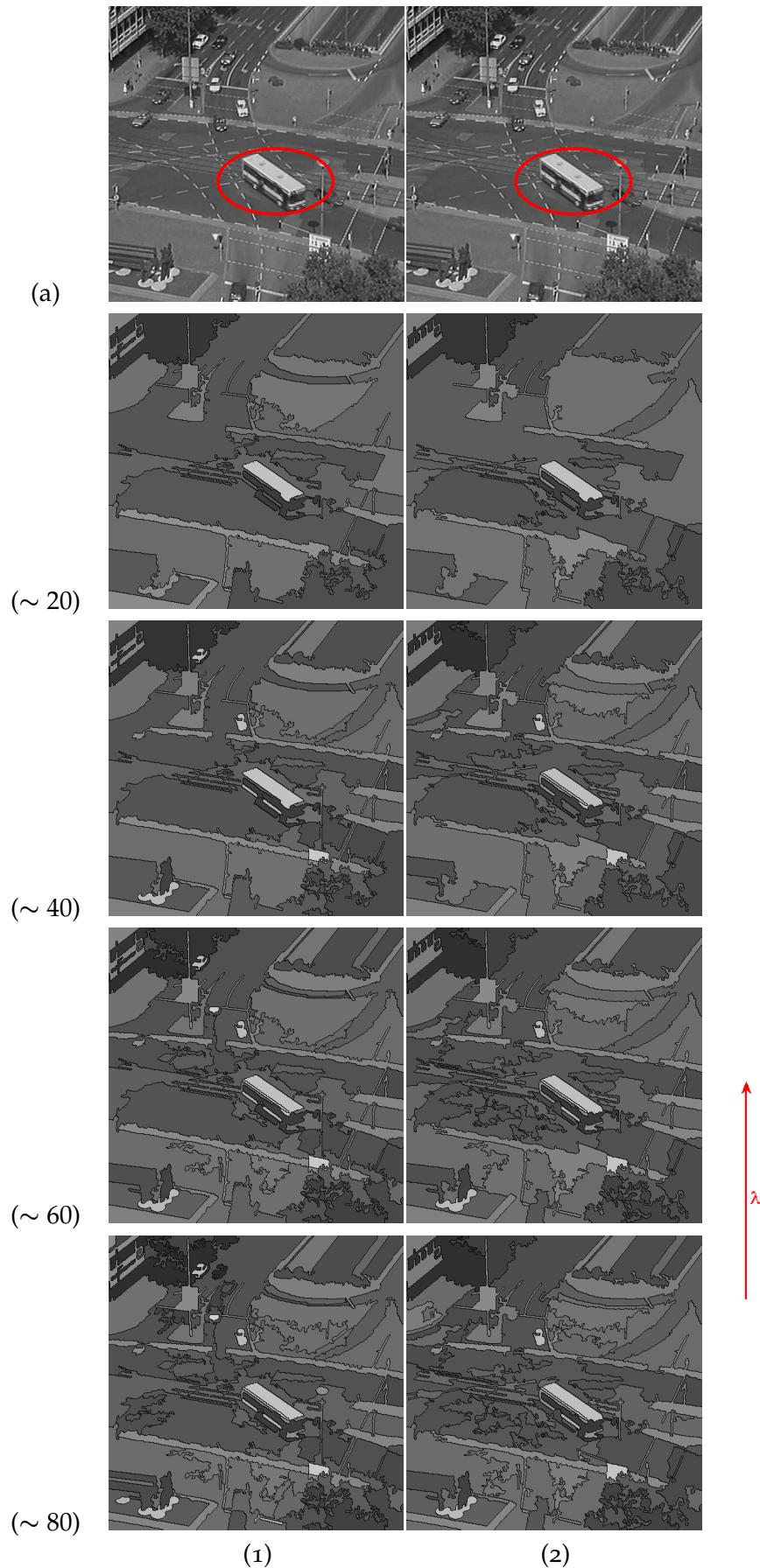


FIGURE 3.40 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $|R| \cdot |B|$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $|R| \cdot |B|$ .

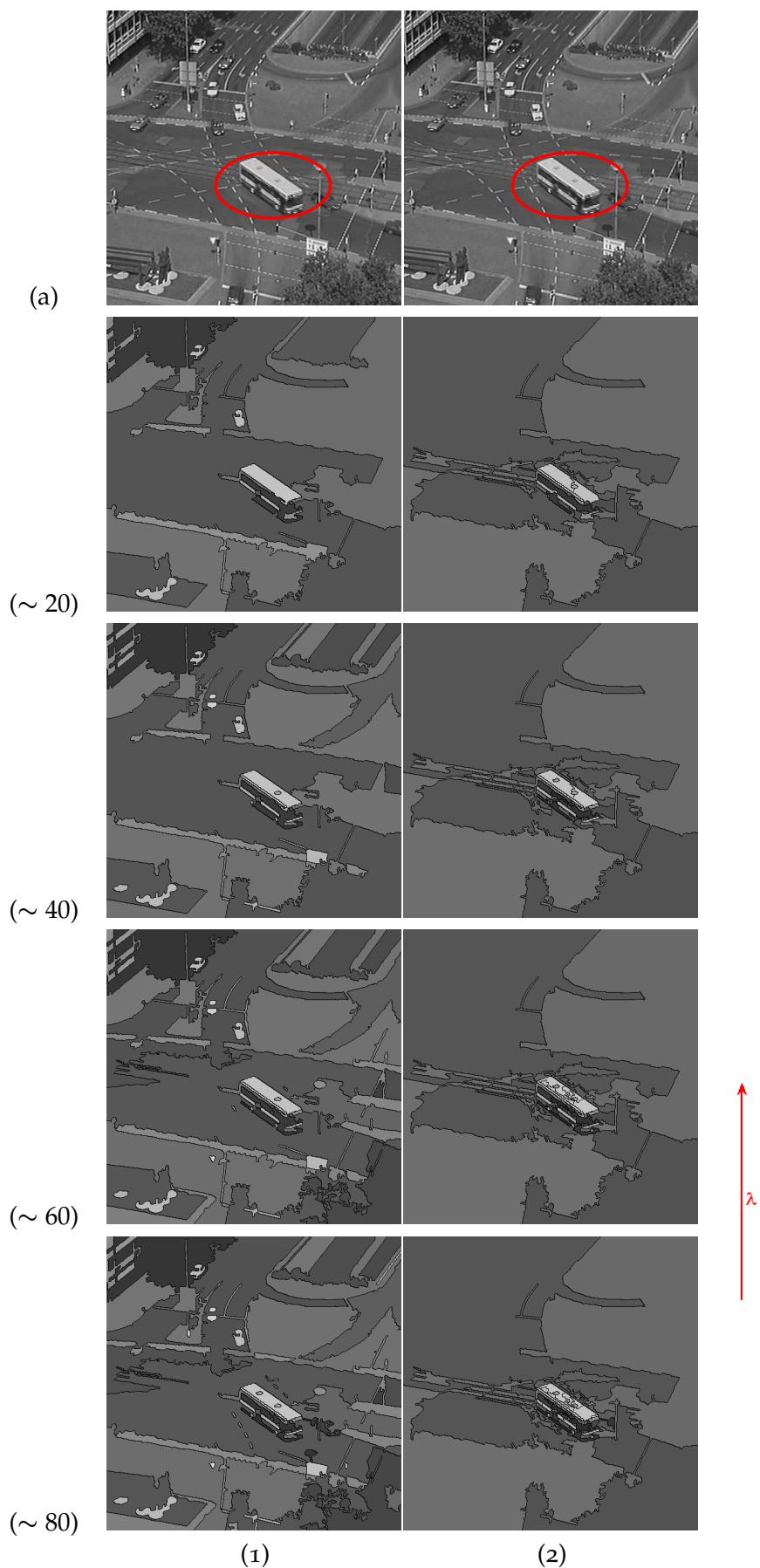


FIGURE 3.41 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $|R \cap B| / |R|$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $1 / (1 + |R \cap B|)$ .

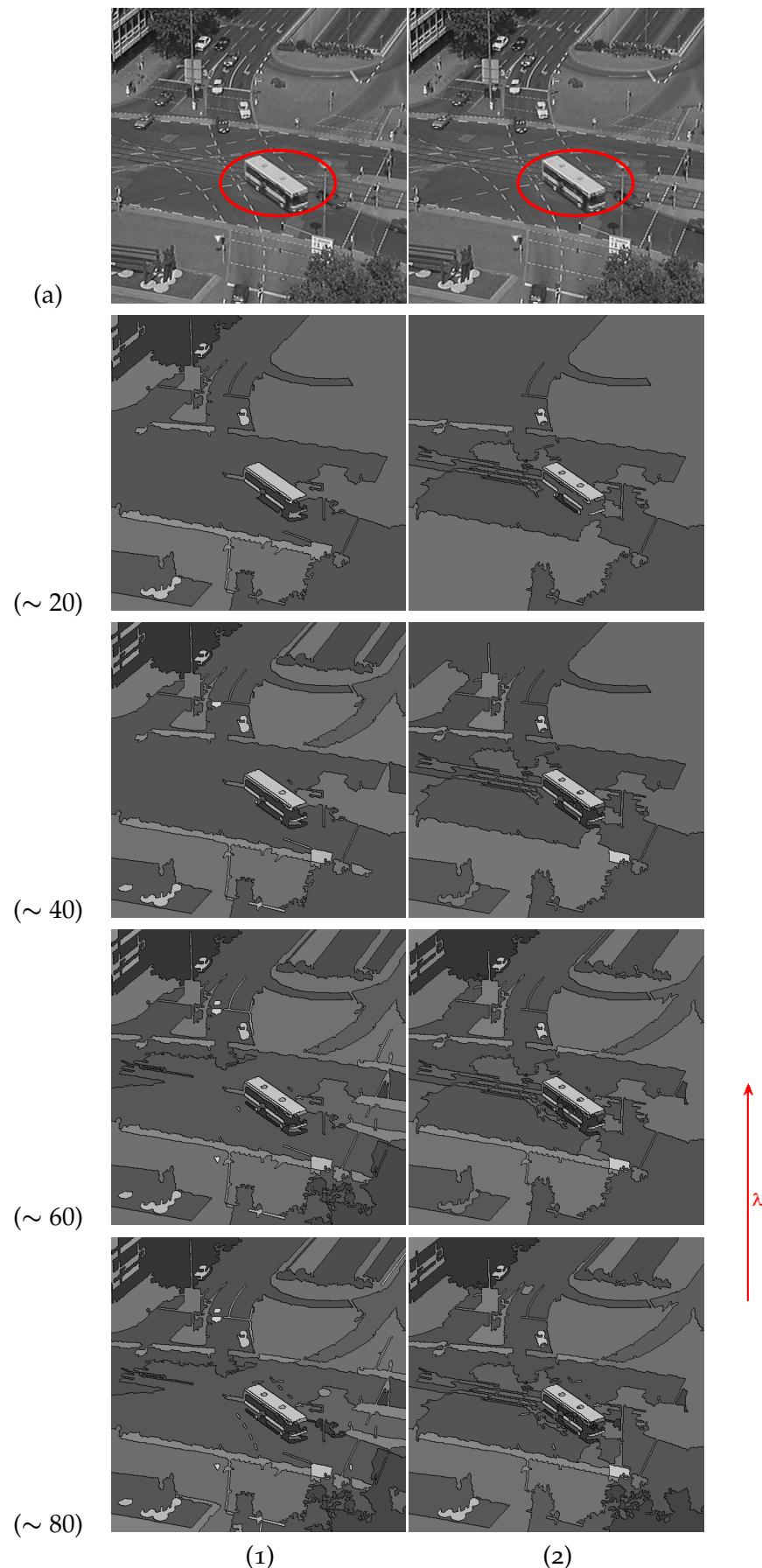


FIGURE 3.42 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $e^{|R \cap B| / |R|}$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $e^{-|R \cap B| / |R|}$ .

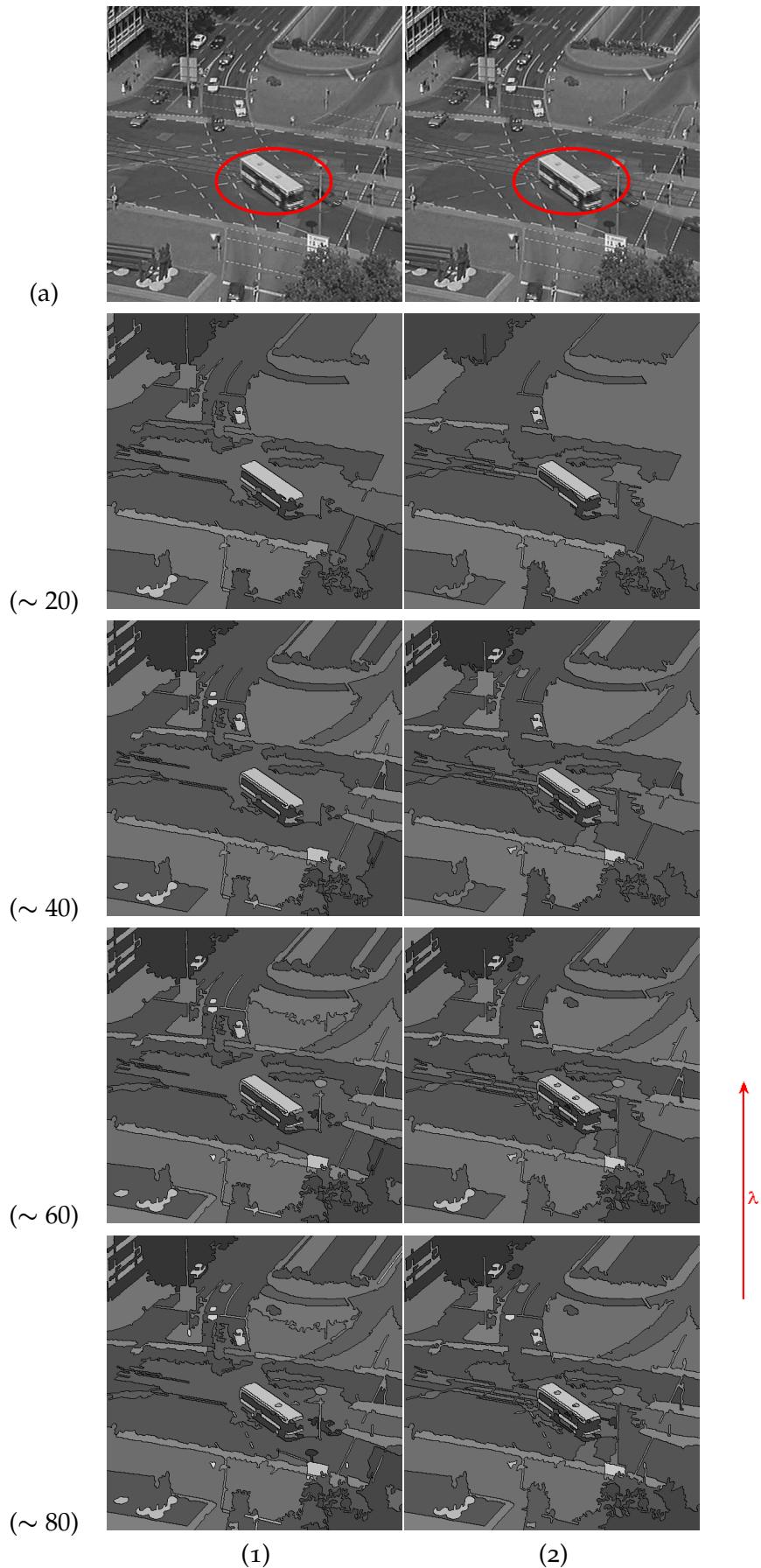


FIGURE 3.43 – Colonne (1) ensemble-échelle obtenu avec l'énergie mono-potentielle ( $e^{|R|/|B|}$ ). Colonne (2) : celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $e^{-|R|/|B|}$ .

### 3.5.2 Cas de plusieurs blobs

Nous allons maintenant étudier le comportement de l'énergie spatiale en présence de plusieurs blobs dans l'image.

L'énergie spatiale issue d'un *a priori* comportant  $K$  blobs se présente sous la forme d'un vecteur d'énergies spatiales relatives à chaque blob, i.e :

$$\overrightarrow{D_B(R,.)} = \{D(R, B_k)\}_{k=1, \dots, M}$$

Il nous faut choisir une norme afin de calculer l'intensité du vecteur des énergies, comme il a été précisé dans la précédente section. Nous avons montré dans cette dernière que des normes telles que  $L^2$ ,  $L^1$  ou  $L^\infty$  permettaient de conserver la propriété de sous-additivité de l'énergie spatiale. Néanmoins, la direction de l'attraction n'étant pas prise en considération, l'intensité du vecteur d'attraction ne peut pas suffire pour caractériser correctement l'interaction entre une région et un blob. Une région incluse dans un blob sera caractérisée par une faible énergie spatiale relativement à ce blob tandis que ses autres composantes seront relativement élevées. Le calcul de l'intensité du vecteur, en attribuant le même poids à chaque blob, diminue l'influence de la composante pertinente. L'énergie de la région, supposée être faible afin d'exprimer la forte attraction de la région vis à vis du blob, se trouve biaisée par les autres composantes du vecteur potentiel. Il en résulte une pénalisation moins forte sur la fusion de cette région avec ses régions voisines.

Ces inconvénients nous ont conduit à choisir une autre fonction afin de calculer l'intensité du vecteur d'attraction :

$$D_B(R) = \min\{\overrightarrow{D_B(R,.)}\} = \min_{k=1, \dots, K}\{D(R, B_k)\}$$

Cette énergie n'est pas sous-additive mais localement, c'dà à l'intérieur des blobs, elle respecte cette propriété. L'énergie d'une région sera exprimée par sa composante énergétique la plus faible. Quand cette région est à l'intérieur d'un blob, la composante relative à ce blob représentera l'énergie de cette région tandis que si la région est à l'extérieur de la région, la composante la moins élevée représentera le blob le plus proche de cette région. La sous-additivité sera forcée dans le cas où elle est transgessée.

#### 3.5.2.1 Energie mono-potentielle

La figure 3.44 représente deux rectangles blancs de même taille dans lesquels sont imbriqués trois rectangles gris de surface croissante de gauche à droite. L'image 3.44-(b)-(2) représente une image-blob contenant deux blobs de même taille entourant les deux petits rectangles. L'image 3.44-(b)-(3) représente une autre image-blob qui se focalise sur deux régions d'intérêt, la première concerne le petit rectangle gris du premier rectangle blanc et la deuxième concerne le rectangle intermédiaire du second rectangle blanc. Nous utiliserons ici l'énergie mono-potentielle attractive. La colonne (a) de la figure 3.44 représente les coupes successives de l'image issues de la hiérarchie construite avec l'énergie de Mumford-Shah. La hiérarchie 3.44-(a)-(1) se scinde en deux hiérarchie partielles quasiment symétrique<sup>6</sup> représentant chacune le contenu des deux rectangles blancs de l'image.

6. L'apparition et la disparition de deux rectangles gris de même taille appartenant à deux rectangles blancs différents sont légèrement différentes. La différence est de l'ordre de  $10^{-2}$ .

L'exploration en profondeur de la hiérarchie révèle l'ordre dans lequel les régions sont fusionnées. La hiérarchie exhibe dans un premier temps les plus grands rectangles gris dans les deux sous hiérarchies. Les rectangles gris intermédiaires sont les suivants à apparaître dans la hiérarchie. Il s'en suit l'apparition des deux petits rectangles gris. Ce comportement est parfaitement concordant avec une analyse multi-échelle sans *a priori*.

La colonne (b) de la figure 3.44 montre quand à elle l'ensemble-échelle obtenu en utilisant l'*a priori* spatial décrit par l'image-blob 3.44-(a)-(2). Elle se focalise sur les deux petits rectangles gris de l'image. La hiérarchie obtenue présente, elle aussi, deux sous-hiéarchies quasiment symétriques. En revanche, l'ordre de fusion est différent. La hiérarchie laisse apparaître dans un premier temps successivement les deux petits rectangles circonscrits. Le reste des rectangles apparaissent ensuite dans l'ordre décroissant de leur taille dans chacun des deux rectangles blancs.

La dernière colonne (c) de la figure 3.44 illustre l'ensemble-échelle obtenu en utilisant l'*a priori* spatial décrit par l'image-blob 3.44-(a)-(3). L'image blob s'intéresse à deux objets d'échelle spatiale différente. Elle se focalise sur le petit rectangle gris du premier rectangle blanc et le rectangle de taille moyenne du deuxième rectangle blanc. La hiérarchie obtenue se décompose en deux sous-hiéarchies non-symétrique. La sous-hiéarchie du rectangle blanc du plus grand blob est la dernière à disparaître de la hiérarchie. Dans cette hiérarchie partielle le dernier objet à disparaître est celui qui a été entouré, la suite de l'exploration respecte quand à elle l'échelle des objets composant le rectangle blanc. La première hiérarchie partielle présente le même comportement. Le petit rectangle gris est le dernier à disparaître puisqu'il représente à travers l'*a priori* spatial introduit, l'objet le plus pertinent des structures composant ce rectangle blanc.

Les figures 3.45 et 3.46 illustrent également des coupes obtenues à partir des deux images blobs citées ci-dessus. Chaque ligne de chaque figure représente les premières coupes extraites de la hiérarchie en partant du sommet. Chaque hiérarchie a été obtenue avec différentes énergies spatiales mono-potentielles. Le comportement général est identique à celui décrit précédemment.

Ainsi, on retrouve un comportement similaire à celui constaté dans l'exemple d'un *a priori* spatial à un seul blob. Souligner des objets d'intérêt par des blobs et l'exprimer à travers une énergie spatiale permet à la fois de se focaliser sur ces zones et d'attribuer un degré de pertinence plus élevé à ces régions. A travers ces exemples sur une image synthétique, nous montrons le rôle de l'énergie spatiale dans la construction d'une hiérarchie. La pertinence des objets d'intérêt n'est plus guidée uniquement par la taille ou l'échelle spatiale des structures mais également par l'*a priori* spatial introduit à travers des blobs entourant ces objets.

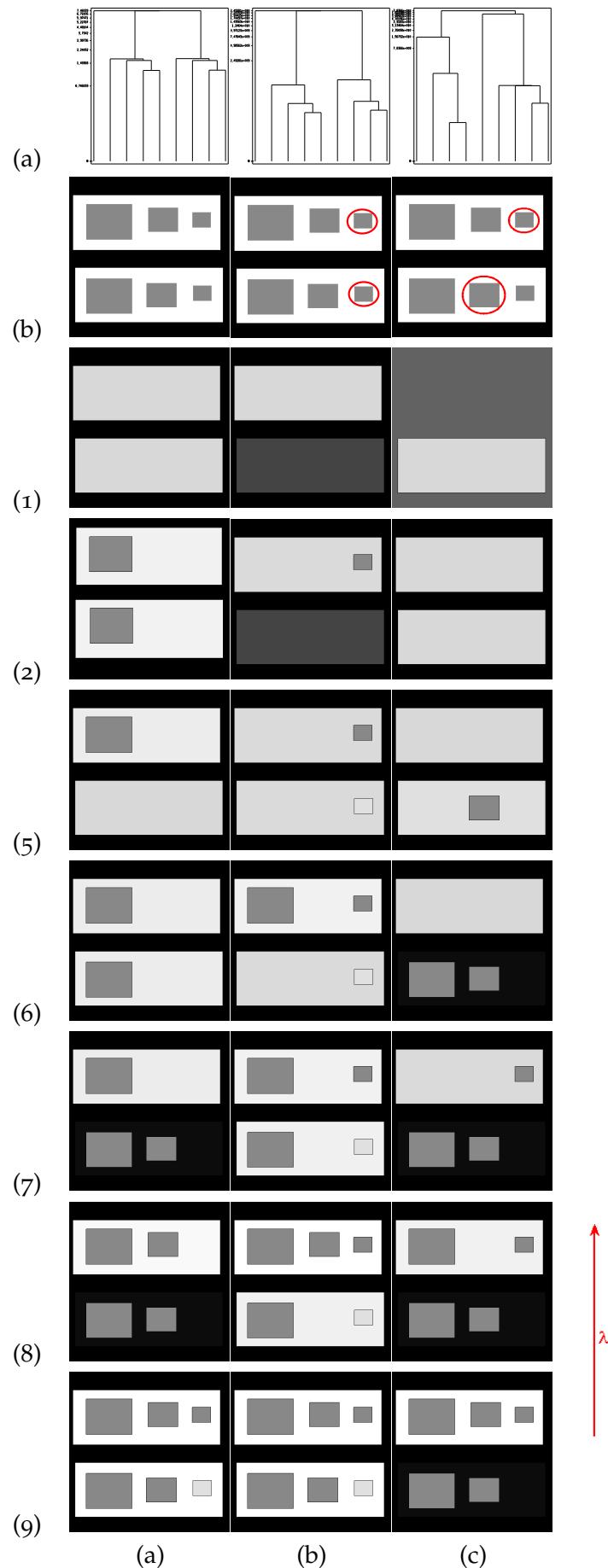


FIGURE 3.44 – Colonne (a) : ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (b) celui obtenu avec l'énergie spatial mono-potentielle attractif partant de deux blobs entourant les deux petits rectangles gris. Colonne (c) : utilise deux blobs entourant le petit rectangle et le rectangle de taille moyenne.

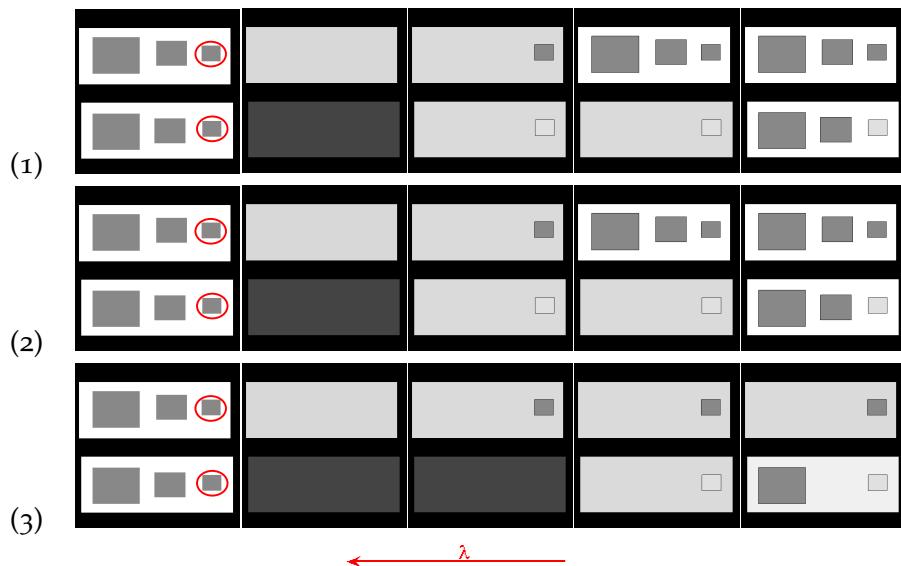


FIGURE 3.45 – (1) coupes obtenues avec l'énergie spatiale mono-potentielle à l'aide d'un potentiel pseudo-attractif utilisant la fonction  $|R \cap B| / |R|$ . (2) coupes obtenues avec la fonction faibl. sur-additive  $e^{|R \cap B| / |R|}$  et la ligne (3) avec la fonction  $e^{|R| / |B|}$ .



FIGURE 3.46 – (1) coupes obtenues avec l'énergie spatiale mono-potentielle à l'aide d'un potentiel pseudo-attractif avec la fonction  $|R \cap B| / |R|$ . (2) coupes obtenues avec la fonction faibl. sur-additive  $e^{|R \cap B| / |R|}$  et la ligne (3) utilisant la fonction  $e^{|R| / |B|}$ .

Nous examinons à présent l'effet de l'énergie spatiale sur une image réelle en présence d'un *a priori* contenant plusieurs blobs.

Nous reprenons pour cela l'image du carrefour traversé par un bus. On s'intéresse cette fois-ci à deux objets : le premier étant le bus et le deuxième représente une voiture noire visible en haut de l'image. Les deux objets ont des échelles spatiales très différentes.

La figure 3.47 représente dans sa première colonne des coupes extraites de l'ensemble échelle obtenu avec l'énergie de Mumford-Shah. Comme cité précédemment, les structures du bus ayant persisté lors du processus de fusion sont des zones dont la radiométrie est très différente par rapport à leur voisinage. On retrouve une partie significative des structures composant le bus qu'à partir d'une partition contenant 80 régions. Concernant la voiture noire elle n'apparaît dans aucune des partitions exhibées dans cette colonne<sup>7</sup>. La deuxième colonne de cette figure représente les coupes obtenues en utilisant l'*a priori* spatial contenant un blob entourant le bus. L'énergie spatiale employée utilise le potentiel attractif. Une partition de 20 régions décrit le bus aussi bien que la partition de 80 régions obtenue sans *a priori*. Toutes les coupes de la deuxième colonne exhibent plus de structures du bus mais aucune structure de la voiture noire n'apparaît sauf dans la partition de 80 régions. La dernière colonne représente les coupes obtenues avec la même énergie spatiale mais en s'appuyant sur les deux blobs comme le montre la figure 3.47-(a)-(3). On observe que l'introduction d'un second blob a permis de relever la pertinence de cet objet. La partition de 40 régions exhibe à la fois une partie importante du bus ainsi que la région principale qui délimite la voiture noire. La comparaison visuelle entre les deux dernières colonnes révèle une forte similitude dans le contenu des coupes à l'exception de l'apparition de la structure décrivant la voiture noire.

Les figures 3.48 et 3.49 représentent les résultats obtenus avec les fonctions faiblement sur-additive :  $|R \cap B|/|R|$ ,  $e^{|R \cap B|/|R|}$ . Dans ces deux figures la partition de 20 régions décrit les deux objets d'intérêt de manière identique à la partition de 40 régions obtenue avec le potentiel attractif. En outre, Les frontières obtenues avec ces deux potentiels pseudo attractif sont plus régulières que celles obtenues avec le potentiel dit attractif.

La dernière figure 3.50 représente les résultats de la construction hiérarchique avec le potentiel pseudo attractif  $e^{|R|/|B|}$ . On observe les contours des partitions sont sinueux comme celles obtenues avec le potentiel attractif. La structure entourée par le petit blob n'apparaît que dans la partition de 40 régions.

---

<sup>7</sup>. La première partition contenant des structures de la voiture noire se compose de 266 régions

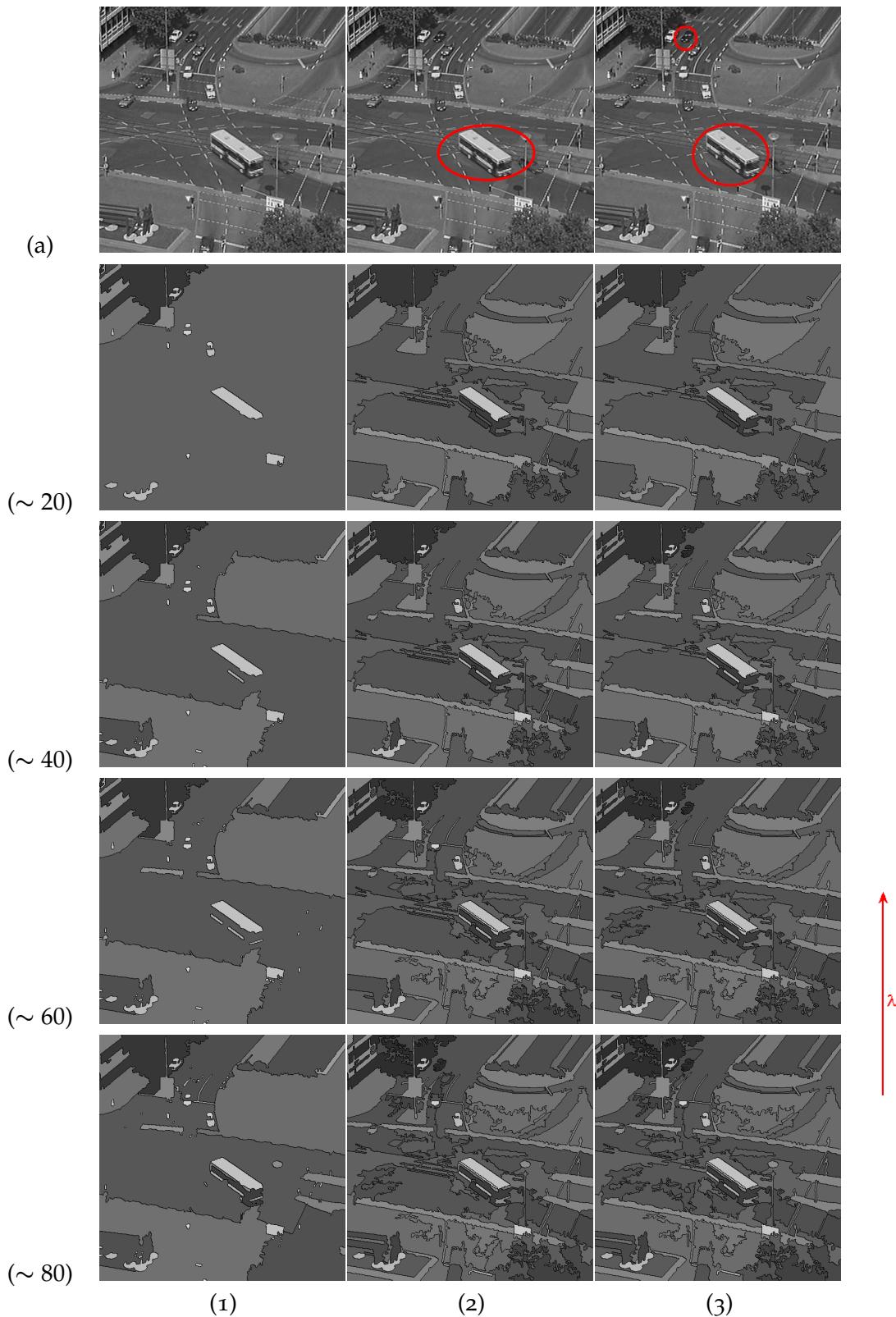


FIGURE 3.47 – Colonne (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) celui obtenu avec l'énergie mono-potentielle ( $|R| \cdot |B|$ ) à partir d'un seul blob. Colonne (3) : celui obtenu avec la même énergie mais en ajoutant un deuxième blob.

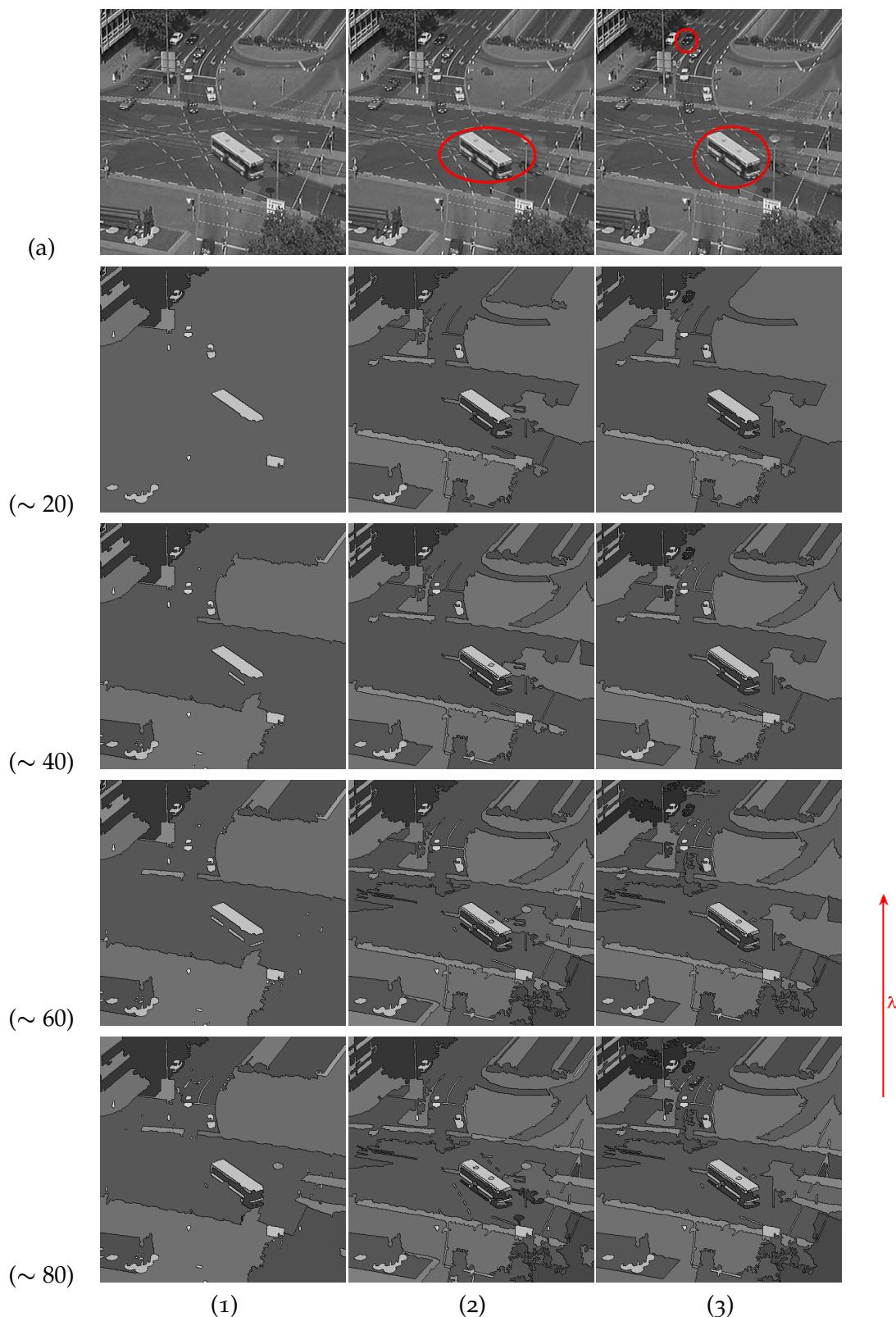


FIGURE 3.48 – Colonne (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) : celui obtenu avec l'énergie mono-potentielle ( $|R \cap B| / |R|$ ) à partir d'un seul blob. Colonne (3) : celui obtenu avec la même énergie mais en ajoutant un deuxième blob.

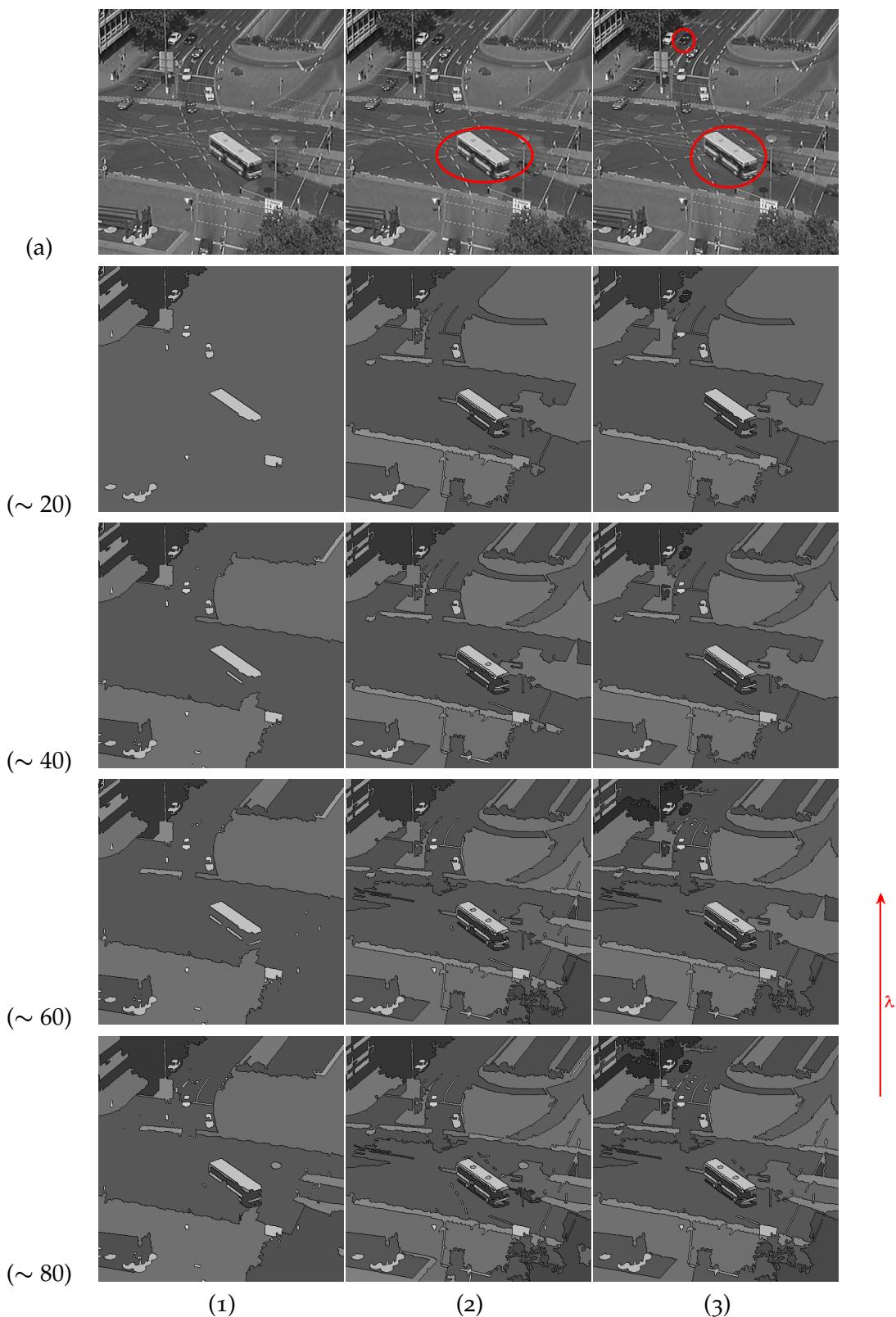


FIGURE 3.49 – Colonne (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) celui obtenu avec l'énergie mono-potentielle ( $e^{|R \cap B|}/|R|$ ) à partir d'un seul blob. Colonne (3) : celui obtenu avec la même énergie mais en ajoutant un deuxième blob.

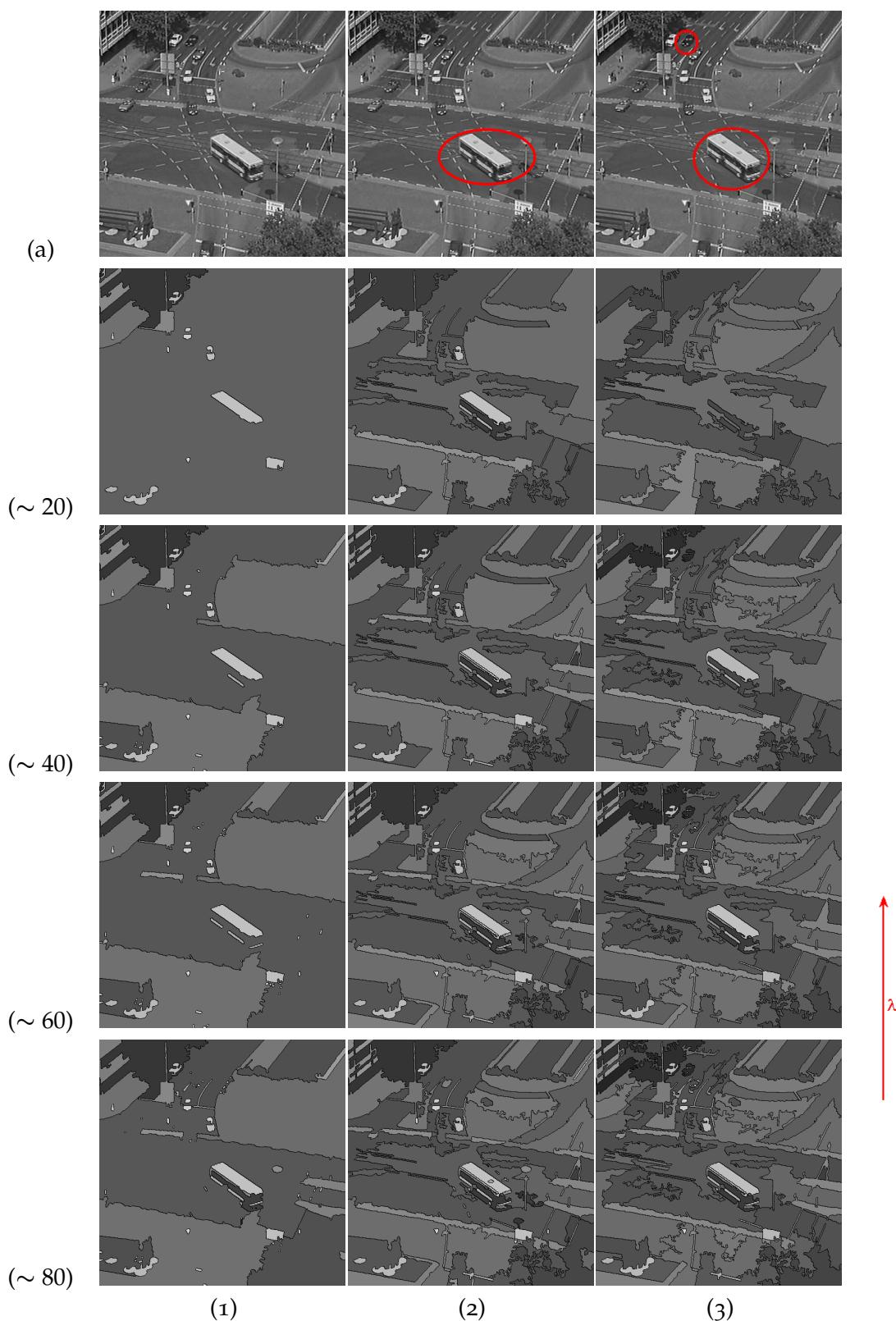


FIGURE 3.50 – Colonne (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. Colonne (2) celui obtenu avec l'énergie mono-potentielle ( $e^{|R|/|B|}$ ) à partir d'un seul blob. Colonne (3) : celui obtenu avec la même énergie mais en ajoutant un deuxième blob.

### 3.5.2.2 Energie bi-potentielle

Nous allons maintenant analyser les résultats obtenus en utilisant une énergie bi-potentielle. On se restreindra à l'étude du comportement de cette énergie construite soit avec la fonction supra-additive  $1/(1 + |R \cap B|)$ , soit avec la fonction  $e^{-|R \cap B|/|R|}$ . On utilisera le bi-potentiel multiplicatif. La distance utilisée est celle du lien minimum.

La colonne (2) de la figure 3.51 représente les coupes extraites de la hiérarchie construite avec un bi-potentiel utilisant la fonction supra-additive  $1/(1 + |R \cap B|)$ . L'*a priori* spatial introduit contient un seul blob entourant le bus. L'exploration profonde de la hiérarchie indique que les zones composant le bus sont de plus en plus morcelées par rapport aux autres régions composant l'image. La fonction supra-additive utilisée impose une forte pénalisation entre régions adjacentes appartenant au blob. Ainsi, les régions se trouvant à l'extérieur du blob sont fusionnées en priorité. Il est clair que l'on accède ainsi plus rapidement aux régions composant le blob mais on occulte en même temps le reste des régions d'intérêt composant l'image. Le processus de fusion devient asynchrone puisque les régions à l'intérieur du blob restent sur-segmentées tandis que les objets à l'extérieur du blob sont soit sous-segmentés, soit déjà agrégés avec d'autres régions. Toutes les coupes de la première colonne ne contiennent aucun détail sur la voiture noire. La colonne (3) de cette même figure construite avec la même énergie spatiale mais sur la base des deux blobs exhibe, quand à elle, les structures de nos objets d'intérêt dès la partition de 20 régions. Le second objet d'intérêt (voiture noire) est présevé de la fusion avec d'autres régions grâce à l'injection d'un nouveau blob.

La figure 3.52 représente les partitions obtenues avec un bi-potentiel construit avec la fonction supra-additive  $e^{-|R \cap B|/|R|}$ . Les coupes de la deuxième colonne, construites à partir du blob entourant le bus n'exhibent aucun détail sur la structure de la voiture noire. Les structures du bus sont présentes mais sont moins sur-segmentées que dans la colonne (2) de la figure 3.51 à cause de la fonction supra-additive utilisée. Cette dernière pénalise moins fortement les fusions entre régions incluses dans le blob que la fonction  $1/(1 + |R \cap B|)$ . La colonne (3) exhibe, quand à elle, la voiture noire dès la partition à 20 régions grâce à l'introduction d'un second blob entourant cette dernière. La différence avec les coupes construites avec l'énergie mono-potentielle est peu significative et ne permet pas de conclure si une énergie mono-potentielle est meilleure qu'une énergie bi-potentielle. Ce choix dépend notamment du choix de la fonction faiblement sur-additive dans le cas mono-potentiel et du choix de la fonction supra-additive dans le cas bi-potentiel. Ainsi, utiliser un bi-potentiel avec la fonction supra-additive  $1/(1 + |R \cap B|)$  induit une forte focalisation sur le blob et permet d'explorer plus rapidement les zones composant l'objet d'intérêt mais rend le processus de fusion relativement asynchrone entre les régions internes et externes aux blobs.

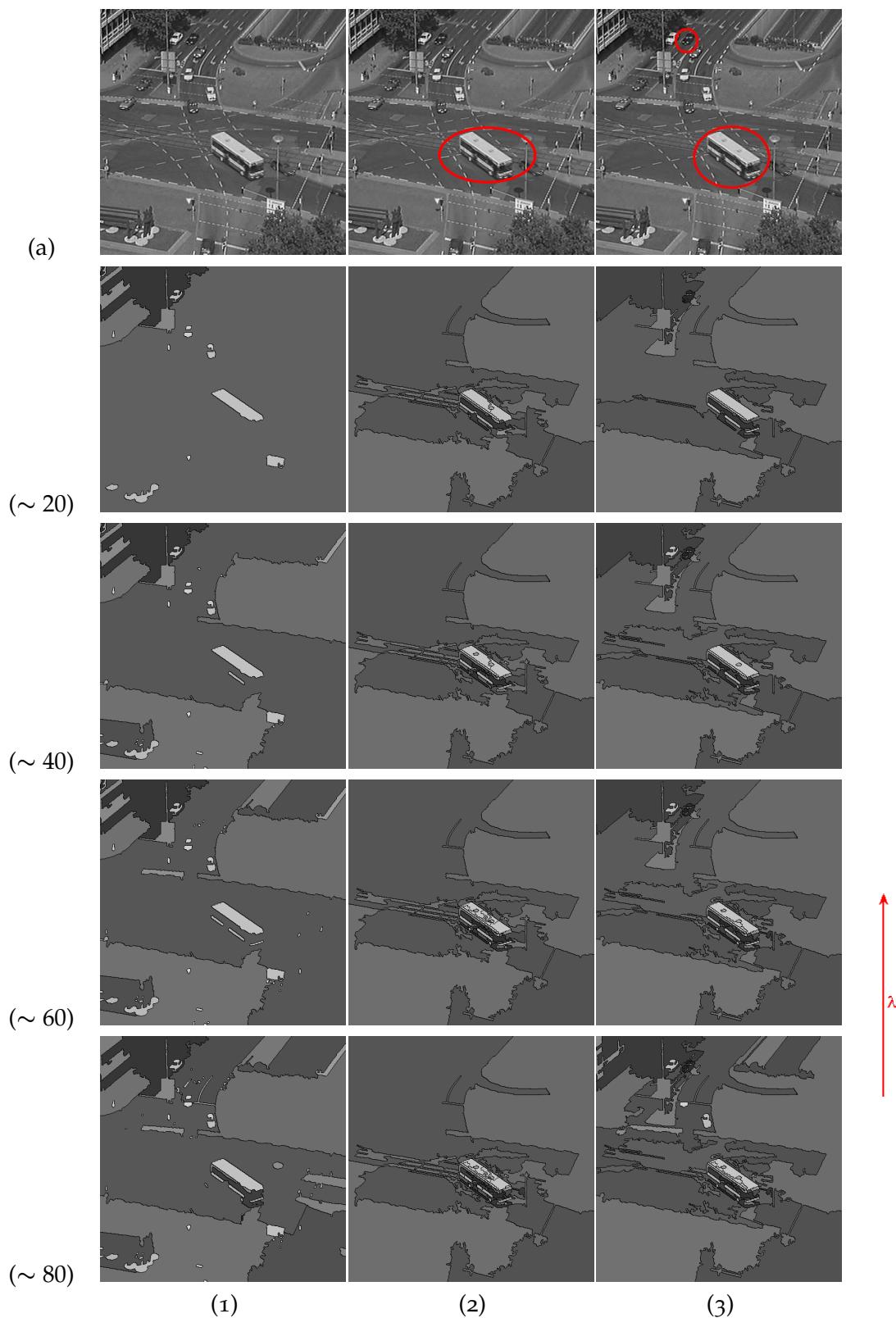


FIGURE 3.51 – (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. (2) celui obtenu avec l'énergie bi-potentielle utilisant la fonction supra-additive  $1/(1 + |R \cap B|)$ . (3) : celui obtenu avec la même énergie mais en ajoutant un deuxième blob.

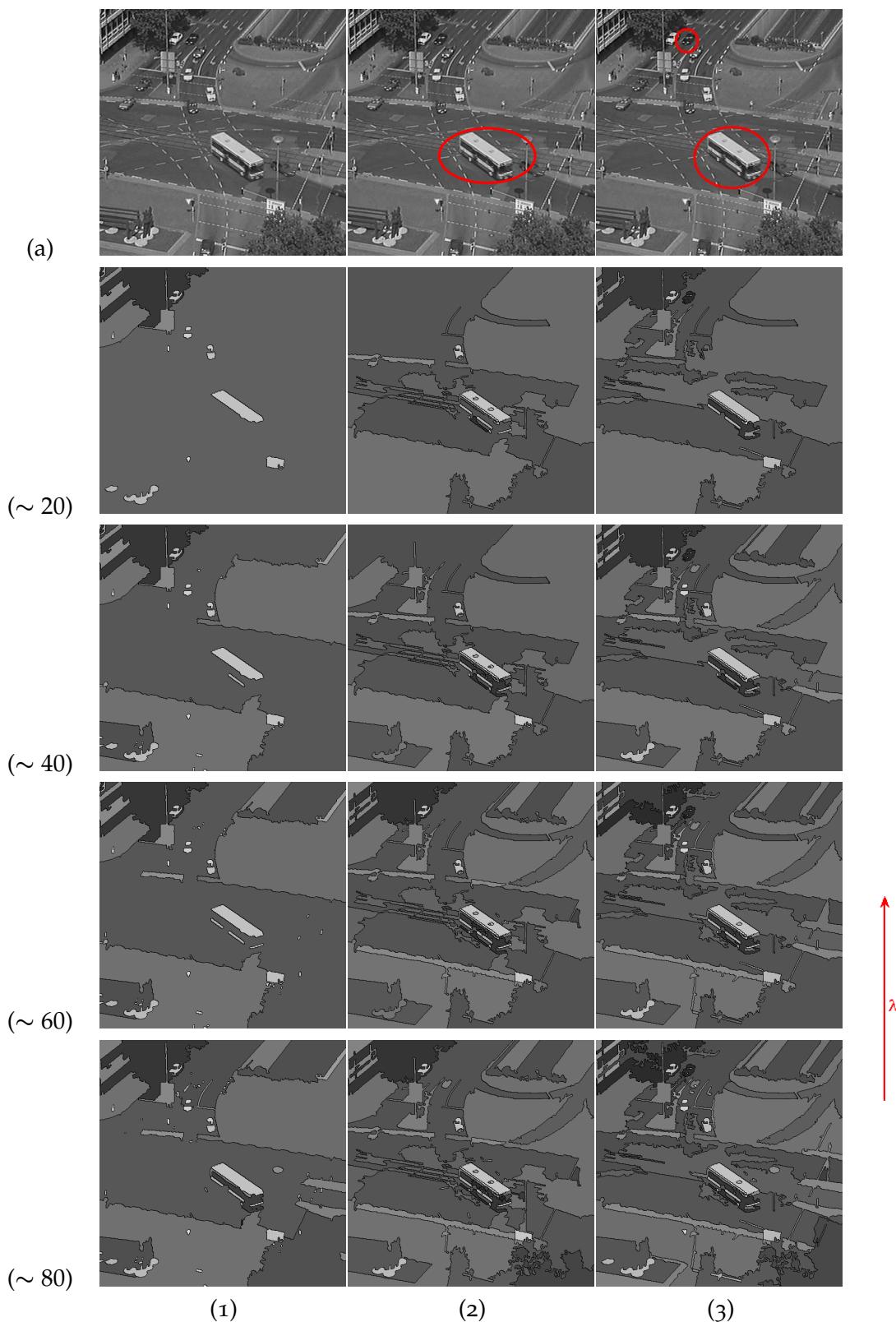


FIGURE 3.52 – (1) ensemble-échelle obtenu avec l'énergie de Mumford-Shah. (2) celui obtenu avec l'énergie bi-potentielle en utilisant la fonction supra-additive ( $e^{-|\mathcal{R} \cap \mathcal{B}| / |\mathcal{R}|}$ ). (3) : celui obtenu avec l'énergie bi-potentielle en utilisant la même énergie mais en ajoutant un deuxième blob.

## CONCLUSION DU CHAPITRE

Dans ce chapitre nous avons traité l'une des façons d'introduire des connaissances *a priori*. Nous avons proposé d'introduire un *a priori* spatial sous forme de blobs afin de souligner la pertinence de certaines zones de l'image dans l'interprétation de la scène. A partir d'une image-blob, nous avons proposé de construire une énergie spatiale « sous-additive » afin de respecter le formalisme de la théorie ensemble-échelle. On a exploré différentes manières de modéliser une énergie spatiale : mono-potentielle, multi-potentielle ou bi-potentielle. Toute région sera caractérisée par une énergie spatiale qui se présente sous forme d'un vecteur dans le cas où l'*a priori* comprend plusieurs blobs. Chaque composante exprimera le degré d'attraction de la région vis à vis d'un blob. L'attraction est d'autant plus forte que l'énergie est faible. En présence de plusieurs blobs c'est l'intensité du vecteur des énergies qui représentera l'énergie de la région. Bien que les normes  $L^p$  sont des applications qui permettent de préserver la sous-additivité de l'énergie, elle ne reflètent pas l'attraction d'une région envers un blob en particulier. Dans le cas où l'image comporte plusieurs blobs nous avons utilisé la fonction du minimum qui renvoie la composante énergétique la plus faible. Les différentes expériences menées sur des images de synthèse ou des images naturelles ont montré l'efficacité et le rôle d'une telle énergie dans la construction hiérarchique. Une énergie spatiale confère aux objets circonscrits un degré de pertinence plus élevé, leurs structures résistent aux processus de fusion et persistent plus longuement dans la hiérarchie. L'exploration de haut en bas de la hiérarchie permet de retrouver plus rapidement ces objets dans des coupes comportant très peu de régions. Le parallèle réalisé avec des hiérarchies construites sans *a priori* spatial conforte notre idée de départ. Des objets pertinent aux yeux d'un utilisateur peuvent ne pas l'être dans la hiérarchie et se retrouver noyées dans des partitions de grande taille en terme de régions. L'introduction d'un *a priori* spatial permet de protéger ces objets et/ou leurs structures de la fusion avec les régions se trouvant à l'extérieur des blobs.

Concernant les différents modes d'énergie, les expérimentations n'ont pas révélé de différence significative entre énergie mono et multi potentiel. La prise en compte du voisinage de la région dans le calcul de son attraction vis à vis d'un blob donne lieu à des résultats sensiblement similaires à ceux obtenus avec l'énergie mono-potentielle. Par contre, l'utilisation d'une énergie bi-potentielle montre que les structures se trouvant à l'intérieur des blobs sont mieux préservées des fusions avec l'extérieur. Cela est dû à une plus forte pénalisation des fusions entre les structures internes au blob ce qui permet aux régions externes aux blobs de fusionner en premier.

La construction d'une énergie spatiale exige dans un premier temps le choix d'un mode, c'est à dire mono, multi ou bi potentiel. Il faut se munir d'une distance métrique et d'une fonction supra-additive ou faiblement sur-additive selon le mode énergétique. Expérimentalement, le mode mono ou bi potentiel se révèle plus pertinent. Quand à la fonction supra-additive ou faiblement sur-additive, on priviliera des fonctions qui prennent en compte le degré d'inclusion de la région vis à vis du blob telle que l'intersection  $|R \cap B|$  ou les ratios  $|R \cap B|/|R|$ ,  $|R \cap B|/|B|$ .

Nous avons relevé durant ce chapitre que la qualité des frontières était inférieure à celles obtenues avec une énergie de Mumford-Shah. Il serait alors recommandé de combiner deux énergies sous-additives la première étant le périmètre de la région qui favorisera la construction de contours réguliers et la deuxième

sera notre énergie spatiale qui permettra l'émergence d'objets pertinents occultés à cause de leur complexité et de leur faible dissemblance avec leur voisinage.

# COUPE OPTIMALE PAR INJECTION D'UN *a priori* SPATIAL

Ce chapitre présente une approche heuristique d'extraction d'une partition pertinente à partir d'une hiérarchie de régions. Il ne s'inscrit pas dans la même thématique que les précédents chapitres puisqu'il ne s'intéresse pas à la construction d'une hiérarchie de régions par injection d'un *a priori*. On suppose dès le départ que la hiérarchie de régions a été construite par un algorithme de segmentation quelconque. La recherche d'une coupe pertinente dans une hiérarchie est fortement liée à l'application visée. Par conséquent, les connaissances *a priori* sont des informations pertinentes pourachever le processus d'extraction. On se restreindra ici à la recherche d'une partition pertinente dans une hiérarchie de régions à partir d'une information *a priori* spatiale. L'approche proposée est considérée comme semi-interactive si cette information spatiale est fournie par l'utilisateur. Elle exploite la même idée que celle exposée dans le chapitre précédent. Elle se base sur l'introduction d'un *a priori* spatial sous forme de blobs dans le but cette fois-ci de sélectionner une partition pertinente.

Dans la première partie de ce chapitre nous allons motiver et développer notre approche de sélection de coupes. Nous présenterons dans une seconde partie, à travers les différentes expérimentations menées, le comportement et la pertinence de cette approche.

## 4.1 CONSTRUCTION D'UNE ÉNERGIE SPATIALE SÉPARABLE

### 4.1.1 Présentation de la démarche

Certaines applications de vision requièrent ou ne s'opposent pas à une fusion d'information ou à une interaction entre l'utilisateur et la structure représentant l'image afin d'extraire une segmentation de l'image satisfaisante. La représentation hiérarchique d'une image, quand elle est associée à un outil de navigation, permet à l'utilisateur de se déplacer à travers une suite monotone de partitions. Cependant, le nombre de coupes horizontales est généralement très élevé et peut atteindre  $2N - 1$  coupes où  $N$  représente le nombre de régions de

la sur-segmentation de départ<sup>1</sup>. Ce nombre ne prend pas en compte les coupes non-horizontales que l'on peut extraire. Par conséquent, la tâche d'extraction peut s'avérer longue et fastidieuse, spécialement lorsque des régions d'intérêt correctement segmentées ne se trouvent pas à une même échelle<sup>2</sup>.

Si l'application visée ne s'oppose pas à une interaction entre l'utilisateur et l'image alors nous pouvons envisager une manière d'interagir avec la hiérarchie différente de la navigation à travers ses coupes horizontales. Notre approche consiste à déplacer l'interaction vers l'image et non à travers la hiérarchie. De la même manière que dans le chapitre précédent, on supposera que l'utilisateur s'intéresse à certaines zones spatiales de l'image. Nous souhaitons extraire une coupe à partir d'une hiérarchie de régions (obtenue par une quelconque méthode) et d'un ensemble de blobs (dessiné par l'utilisateur ou obtenu par un dispositif de fusion multicapteur). L'extraction de la coupe se fera par minimisation d'une énergie spatiale sur l'espace des partitions de la hiérarchie. Cette énergie intégrera l'*a priori* spatial introduit par l'utilisateur. Cette approche peut être considérée comme un post-traitement de la segmentation hiérarchique qui permet à l'utilisateur, grâce à des connaissances *a priori*, d'extraire une partition pertinente pour son application.

#### 4.1.2 Coupe minimale d'une hiérarchie

Soit  $I$  une image dont le support spatial est noté  $\Omega$  et soit  $\mathcal{H}$  une hiérarchie de régions de l'image  $I$ .

L'ensemble des blobs disjoints sera noté :

$$\mathcal{B} = \{B_1, \dots, B_M\}.$$

La recherche d'une partition optimale au sens d'une énergie séparable  $E$  dans une hiérarchie  $\mathcal{H}$  se résout par un algorithme de programmation dynamique à condition que l'énergie  $E$  soit séparable (Guigues [2003]). La coupe extraite sera composée des régions contenues dans la hiérarchie. Nous avons abordé ce point dans le premier chapitre lorsque que nous avons rappelé le formalisme ensemble-échelle (voir la section 1.4).

Désignons par  $E_s : \mathbb{P}(\Omega) \longrightarrow \mathbb{R}^+$  l'énergie spatiale d'une partition  $P$ . On supposera que  $E_s$  est **séparable** :

$$E_s(P) = \sum_{R \in P} E_s(R).$$

L'algorithme de programmation dynamique s'appuie sur une relation de récurrence qui lie l'énergie d'une région  $R$  appartenant à la hiérarchie aux énergies de ses sous-structures :

$$\forall R \in \mathcal{H}, \quad E_s^*(R) = \inf \left\{ E_s(R), \sum_{R' \in \mathcal{S}(R)} E_s^*(R') \right\} \quad (4.1)$$

où  $\mathcal{S}(R)$  représente l'ensemble des nœuds fils de la région  $R$  dans la hiérarchie  $\mathcal{H}$ .  $E_s^*(R')$  représente l'énergie de la partition optimale de la région  $R'$ .

1. Une partition d'initialisation, obtenue par une ligne de partage des eaux, peut comprendre des milliers de régions, cela donne une idée du nombre de coupes horizontales dans une hiérarchie.

2. Deux régions d'intérêt peuvent être correctement segmentées dans une hiérarchie avec une intersection vide de leurs intervalles de persistance. Il en résulte qu'aucune coupe horizontale de la hiérarchie ne peut contenir les deux régions à la fois. L'utilisateur devra alors choisir une partition sur-segmentée afin de s'assurer de la présence des sous-structures d'une des deux régions.

La partition optimale obtenue est celle qui minimise l'énergie  $E_s$  sur l'ensemble des partitions que l'on peut extraire. L'ensemble des coupes horizontales de la hiérarchie n'est qu'une partie de cet ensemble comme le montre la figure ??.

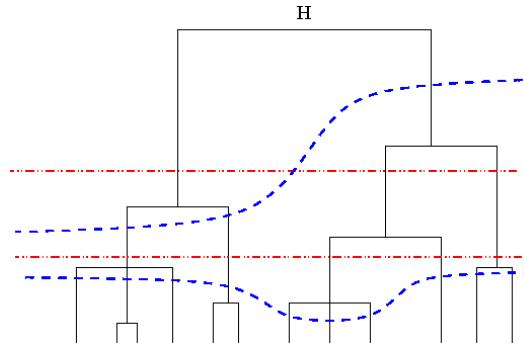


FIGURE 4.1 – Exemples de coupes dans une hiérarchie de régions : les coupes horizontales d'une hiérarchie (en rouge) ne représente qu'une petite partie de l'ensemble des partitions en régions qui appartiennent à la hiérarchie

La complexité de l'algorithme, pour une hiérarchie binaire construite à partir d'une sur-segmentation  $P_0$  de taille  $N_0$ , est une fonction linéaire de  $N_0$  (voir Guigues [2003]). Le temps de calcul de la partition optimale sur une hiérarchie est donc rapide et se fait en une phase ascendante pour le calcul des énergies et une phase descendante pour la collecte des noeuds optimaux.

Si l'on suppose que la hiérarchie a été construite par l'approche ensemble-échelle en utilisant un couple d'énergies  $E = (C, D)$ , il reste à définir une énergie  $E_s$  intégrant l'*a priori* spatial introduit par l'expert (ou fournit par un détecteur) via l'ensemble des blobs.

#### 4.1.3 Energie spatiale

Avant de construire l'énergie spatiale  $E_s$ , nous devons nous interroger sur le comportement que l'on souhaite lui imprégner. La figure ?? montre une hiérarchie de régions dans laquelle on a intégré un blob désignant une région d'intérêt de l'image. Comme le montre cette figure, cette région n'existe pas dans la hiérarchie mais elle est composée de deux sous hiérarchies. La coupe idéale devra donc sélectionner les sous-structures incluses dans le blob, les plus hautes dans la hiérarchie. Le processus de minimisation de l'énergie doit donc permettre d'escalader les régions incluses dans le blob afin d'atteindre la ou les régions d'intérêt incluse(s) dans le blob. Par conséquent, cette énergie devra être sous-additive ( $E_s(x \cup y) \leq E_s(x) + E_s(y)$ ) à l'intérieur du blob. Par contre, cette énergie devra pénaliser la fusion des structures incluses dans le blob avec les régions externes aux blobs ou limitrophes (notamment quand la taille de ces régions est plus grande que celle du blob). L'énergie devra donc changer de comportement et devenir sur-additive. Quand au comportement de l'énergie en dehors du blob, il devra être sous-additif car l'objectif est d'obtenir une partition la plus simple possible qui se focalise sur le(s) région(s) incluse(s) dans le blob mais qui présente les grandes structures de l'image à l'extérieur du blob. L'énergie spatiale devra donc escalader la hiérarchie à l'intérieur et à l'extérieur du blob (sous-additivité) et stopper son ascension aux frontières du blob comme l'illustre la coupe de la hiérarchie dans la figure ??.

L'énergie spatiale que nous allons construire s'inspire largement de celle utilisée dans le chapitre précédent. Cependant, la condition de sous-additivité n'est pas requise. Nous allons présenter notre énergie, sous la forme d'une combinaison de deux critères l'un métrique et l'autre ensembliste.

On définit l'énergie spatiale d'une région relative à un blob  $B$  par :

$$E_s(R, B) = \xi(R, B).d(R, B),$$

tel que  $d(., B)$  représente une distance métrique quelconque entre le centre de gravité de la région et celui du blob.  $\xi(., B)$  représente un critère ensembliste que l'on va expliciter plus en détail.

#### 4.1.3.1 Construction ad-hoc du critère ensembliste

Le critère ensembliste est un ensemble de règles décrivant la relation (au sens de l'inclusion) entre une région et un blob. Le critère applique une pénalisation en fonction de la configuration de la région vis à vis d'un blob. On notera ce critère ensembliste relatif à un blob par  $\xi(., B)$ .

Nous distinguerons le cas où la surface de la région est plus grande que celle du blob du cas où celle-ci lui serait inférieure.

Si  $|R| \leq |B|$  alors :

$$\xi(R, B) = \begin{cases} \xi(R, B) &= \frac{1}{|R|} & si & R \subset B \\ \xi(R, B) &= \frac{|R \cap B|}{|R|} & si & R \cap B \neq \emptyset \\ \xi(R, B) &= 1 & si & R \cap B = \emptyset \end{cases}$$

Si  $|R| > |B|$  alors :

$$\xi(R, B) = \begin{cases} \xi(R, B) &= |R| & si & B \subset R \\ \xi(R, B) &= |R \cap B| & si & R \cap B \neq \emptyset \\ \xi(R, B) &= 1 & si & R \cap B = \emptyset \end{cases}$$

La définition du critère ensembliste se divise en deux parties. Dans le cas où une région a une surface inférieure à celle du blob alors la pénalisation prend ses valeurs dans l'intervalle  $[0, 1]$ . La plus faible pénalisation est attribuée aux régions incluses dans le blob afin de minimiser leur énergie spatiale globale. Dans le cas où la région est externe au blob alors la pénalisation est de 1. En cas d'intersection entre la région et le blob alors on applique une pénalisation intermédiaire qui correspond au ratio de l'inclusion de la région dans le blob.

La deuxième partie du critère  $\xi(., B)$  décrit les règles de pénalisation dans le cas où la taille d'une région est plus grande que celle du blob. Si le blob est inclus dans la région alors la pénalisation est donnée tout simplement par la taille de la région. Cette forte pénalisation par rapport au premier cas de figure ( $|R| \leq |B|$ ) permet d'empêcher la sélection de régions englobant le blob et favorise plutôt la sélection de régions composant le blob. Le blob étant une localisation grossière d'un objet complexe telle qu'une voiture ou un piéton, on préférera obtenir une partition décrivant les sous-structures de la région d'intérêt qu'une partition composée d'une super-structure enveloppant le blob. Une telle super-structure ne contiendra pas les contours de l'objet d'intérêt et ne permettra pas une reconnaissance ou une reconstitution de l'objet *a posteriori*. Dans le cas de figure où la région n'intersecte pas le blob, la pénalisation est la même que celle du cas de figure ( $|R| \leq |B|$ ).

Le choix de garder la même pénalisation permet de sélectionner les régions externes aux blobs les plus élevées dans la hiérarchie, quelque soit leur taille. Pour la dernière situation, celle d'une zone plus grande que le blob mais intersectant ce dernier, nous avons opté pour une pénalisation intermédiaire, à savoir le cardinal de la zone partagée par le blob et la région.

Nous avons établi ces règles de pénalisation de façon heuristique et par une analyse élémentaire des configurations qui peuvent exister entre un blob et une région.

#### 4.1.3.2 Configurations possibles de l'énergie

L'énergie spatiale d'une région relativement à un blob  $B$  est définie par :

$$E_s(R, B) = \xi(R, B).d(R, B),$$

Le choix d'une distance métrique  $d(R, B)$  est très large. On peut opter pour la distance euclidienne, la distance de Mahalanobis, la distance de lien minimum, etc.

Nous pouvons également intégrer le voisinage de la région dans le calcul de l'énergie comme nous l'avons fait dans le chapitre précédent :

$$E_m(R, B) = \sum_{R' \in T(R)} E_s(R', B),$$

tel que  $T(R) = V(R) \cup \{R\}$ .

On peut également définir une énergie spatiale sur le même principe que l'énergie spatiale bi-potentielle :

$$E_b(R, B) = \sum_{R' \in V(R)} E_B(R, R')$$

tel que  $E_B(R, R') = E_s(R, B).E_s(R', B)$ .

#### 4.1.3.3 Discussion

Dans le précédent chapitre, on s'est attaché à construire une énergie spatiale sous-additive afin de respecter le formalisme ensemble-échelle. Un des termes de cette énergie représente une distance métrique qui est convexe et/ou sous-additive. Ici, la situation est presque identique, l'énergie comporte un terme de distance. La sous-additivité de cette distance aura pour effet de préférer la fusion de deux régions que leur disjonction. L'introduction du critère ensembliste permet d'adapter ce comportement en fonction des différentes configurations entre une région et un blob. Un comportement sous-additif est souhaité lorsque deux régions sont incluses dans le blob afin de continuer à escalader la hiérarchie dans le but (idéal) de sélectionner la région d'intérêt ou ses sous-structures les plus élevées (voir la figure ??).

Par contre, lorsqu'une région incluse dans un blob rencontre une région intersectant le blob et de taille plus grande que celui-ci, il est naturel d'interdire la sélection de leur région mère dans la hiérarchie (voir la figure ??). La fusion de ces deux régions donnera naissance à une région intersectant le blob et ayant une surface plus grande que le blob. L'énergie de cette région sera supérieure à la somme des énergies de ses sous-structures car la pénalisation est beaucoup plus

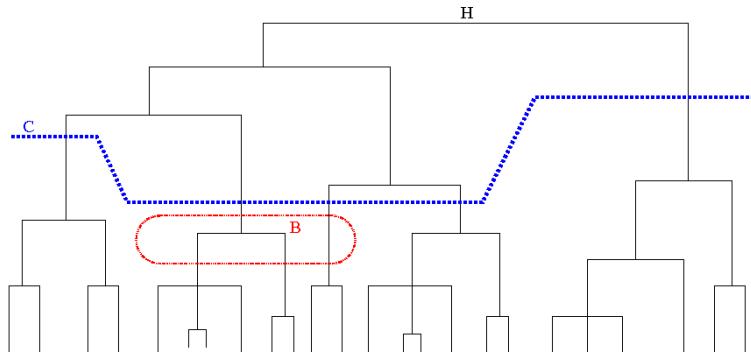


FIGURE 4.2 – La figure montre une hiérarchie de régions dans laquelle une région d'intérêt est entourée par un blob. La coupe idéale traverse les grandes structures de la hiérarchie ainsi que les sous-structures incluses dans le blob. L'énergie doit escalader (sous-additivité) la hiérarchie à l'intérieur et à l'extérieur du blob mais stopper sa montée (sur-additivité) aux abords du blob.

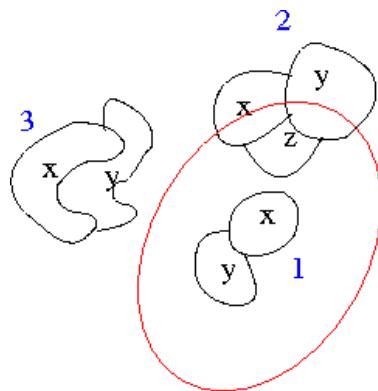


FIGURE 4.3 – La figure montre 3 configurations différentes entre deux régions adjacentes et un blob.  
(1) quand les deux régions ( $x$  et  $y$ ) sont incluses dans le blob alors on sélectionnera la région de la hiérarchie représentant leur fusion ( $x \cup y$ ) ce qui correspond à escalader la hiérarchie. (2) les deux régions sont limitrophes au blob. On continue à escalader la hiérarchie en sélectionnant la région représentant l'union dans la hiérarchie. (3) les deux régions sont en dehors du blob. La pénalisation est de 1 et le terme énergétique dépend uniquement de la distance qui est sous-additive. Là encore la région représentant leur fusion sera sélectionnée.

faible ( $\in [0, 1]$ ) pour la région incluse dans le blob. Il en résulte un inversement de comportement, l'énergie devient sur-additive et la région représentant la fusion des deux régions ne sera pas optimale.

Dans le cas où l'*a priori* spatial comporte plusieurs blobs, l'énergie spatiale sera modélisée par la norme du vecteur des énergies spatiales relatives à chaque blob :

$$E_B(R) = \|\overrightarrow{E_B(R, .)}\|_p^p$$

tel que tel que  $\overrightarrow{E_B(R, .)} = \{E_s(R, B_k)\}_{k=1, \dots, M}$ . et  $p \geq 1$ .

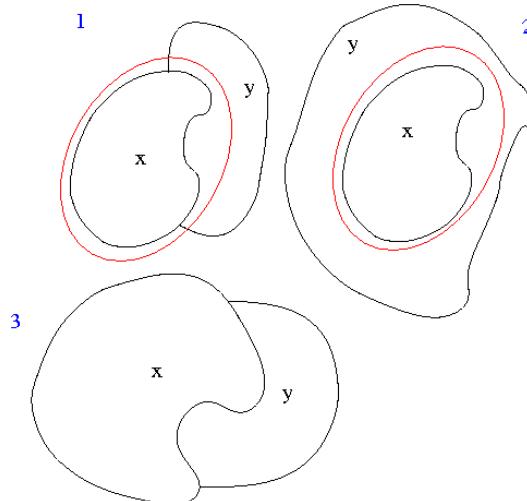


FIGURE 4.4 – La figure montre 3 configurations différentes entre deux régions adjacentes et un blob dans le cas où la fusion des régions  $x$  et  $y$  a une surface plus grande que celle du blob. (1) et (2) illustre la situation où une région ( $x$ ) est incluse dans le blob et l'autre région ( $y$ ) intersecte le blob. La région de la hiérarchie représentant leur fusion ( $x \cup y$ ) ne sera pas optimale. (3) les deux régions sont en dehors du blob. la pénalisation est de 1 et le terme énergétique dépend uniquement de la distance qui sous-additive. Là encore la région représentant leur fusion sera sélectionnée.

## 4.2 RÉSULTATS EXPÉRIMENTAUX

Nous allons mettre en application cette approche heuristique sur différentes images naturelles. Afin d'extraire une coupe optimale au sens de l'énergie  $E_B$ , il faudra dans un premier temps construire une hiérarchie de régions de l'image. On adoptera l'approche ensemble-échelle pour construire la hiérarchie de chaque image. On utilisera l'énergie constante par morceaux de Mumford-Shah et on initialisera le processus par une partition construite avec l'algorithme de partage des eaux.

L'introduction d'un *a priori* spatial, à travers des blobs, a pour but de se focaliser en priorité sur certaines zones de l'image. L'objectif est donc de représenter soit l'objet d'intérêt soit ses sous structures. Quand au reste de l'image, l'énergie devra illustrer des détails grossiers du contenu compte tenu de sa nature sous-additive en dehors des blobs.

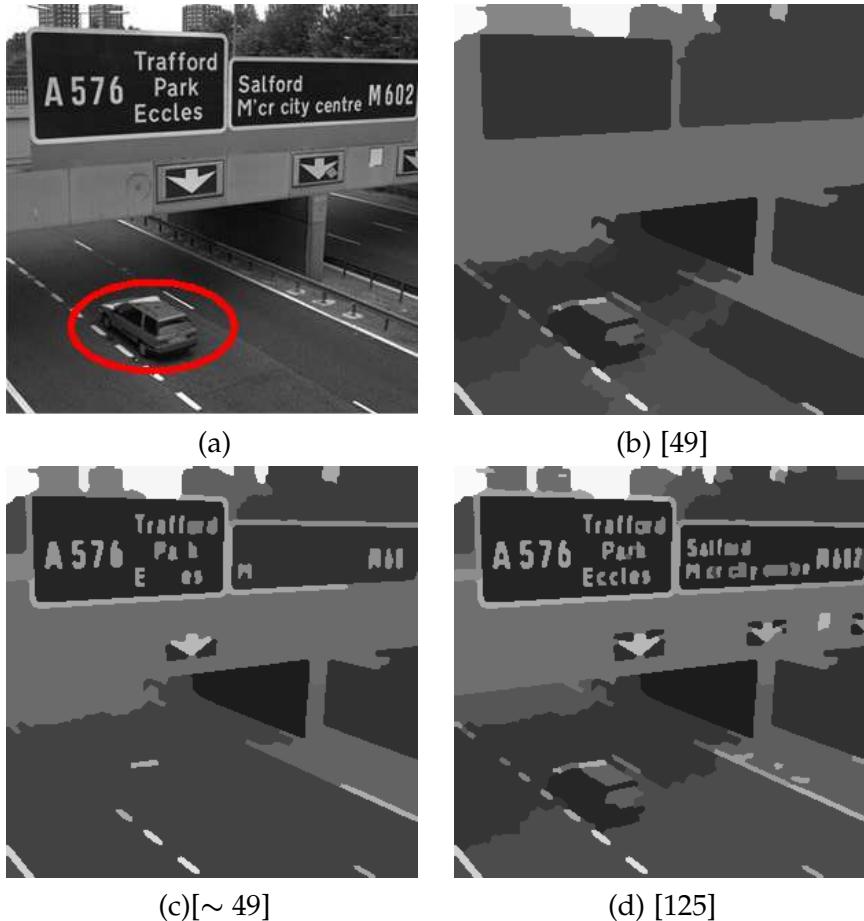


FIGURE 4.5 – La figure (a) représente l'image étudiée comportant un blob. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b). (d) coupe horizontale extraite manuellement et exhibant un niveau de détail de l'objet d'intérêt équivalent à celui de la coupe (b).

### 4.2.1 Cas d'un blob

La figure ??, reprend l'image du trafic périphérique, utilisée dans le précédent chapitre. La partition (b) représente la coupe optimale extraite avec notre énergie spatiale qui combine le critère ensembliste et la distance euclidienne. Elle comporte 49 régions. Comme attendu la partition exhibe très peu de détail à l'extérieur du blob. Pour exemple, le contenu des panneaux est complètement fusionné. On remarquera qu'un nombre non négligeable de régions limitrophes au blob sont décrites. L'agrégation de celles-ci pouvant entraîner la disparition de structures de l'objet d'intérêt, l'énergie spatiale conduit à l'interdiction de ces fusions. La partition (c) représente une coupe horizontale de la hiérarchie comportant un nombre équivalent de régions. Cette coupe correspond à une partition optimale au sens de l'énergie de Mumford-Shah à l'échelle correspondante. On notera qu'à partitions égales en nombre de régions, cette dernière, exhibe les détails concernant le contenu des panneaux et aucun détail correspondant à l'objet d'intérêt. Il est évident que la coupe extraite par notre approche n'est pas horizontale et ne correspond donc à aucune coupe optimale au sens de l'énergie affine utilisée pour la construction de la hiérarchie. La partition (d) représente une coupe horizontale extraite manuellement. C'est la première coupe dont la description de l'objet d'intérêt se rapproche de celle extraite par notre énergie. Cependant elle contient 125 régions.

L'analyse de la figure ?? confirme les observations faites dans la figure précédente. La description de la structure du bus, obtenue à travers notre énergie, est satisfaisante. La coupe extraite comporte 298 régions (sachant que la taille en régions de la partition initiale, obtenue avec la L.P.E est de 11244). En revanche, pour obtenir une description équivalente, il a fallu extraire une coupe horizontale comportant 520 régions. On notera que notre coupe (b) donne très peu de détail sur l'immeuble se trouvant au coin supérieur gauche de l'image alors que les coupes (c) et (d) en donnent une description très riche. Il semble que des zones d'intérêt qui sont assez éloignées de l'objet d'intérêt ont moins de chance d'être présent dans la partition finale. Ces zones ont de fortes chances d'être des sous-hiéronymies indépendantes des sous hiérarchies représentant l'objet d'intérêt. Les zones de l'image correspondant à ces sous-hiéronymies sont donc ignorées pendant la phase de minimisation.

La figure ?? illustre un autre exemple du rôle joué par l'énergie spatiale dans la sélection des régions optimisant cette dernière. A nombre de régions égal, la coupe horizontale (c) (extraite manuellement) ne décrit qu'une partie de l'objet d'intérêt. On remarquera que les sous-structures de notre objet ont été absorbées par la zone représentant la voie à cause d'une forte similarité radiométrique. En revanche, la coupe (b) (extraite par l'énergie spatiale) décrit clairement les contours de l'objet d'intérêt. Celui ci est composé de plusieurs sous-structures car il n'existe pas en tant que structure à part entière dans la hiérarchie. L'énergie spatiale présentera alors ses sous-structures, ce qui démontre le comportement sur-additif de l'énergie dans ce type de situation.

La figure ?? représente les résultats de l'extraction d'une coupe optimale sur un autre type d'image. L'image illustre des piétons à la sortie d'une université. Le blob circonscrit le piéton se trouvant au milieu de l'image. La coupe (b) comporte 82 régions et décrit le piéton de façon satisfaisante. Comme attendu, il est difficile de segmenter un piéton en tant qu'objet unique car celui-ci est composé de régions très hétérogènes. Ainsi, la coupe (b) nous présente les structures le composant. A

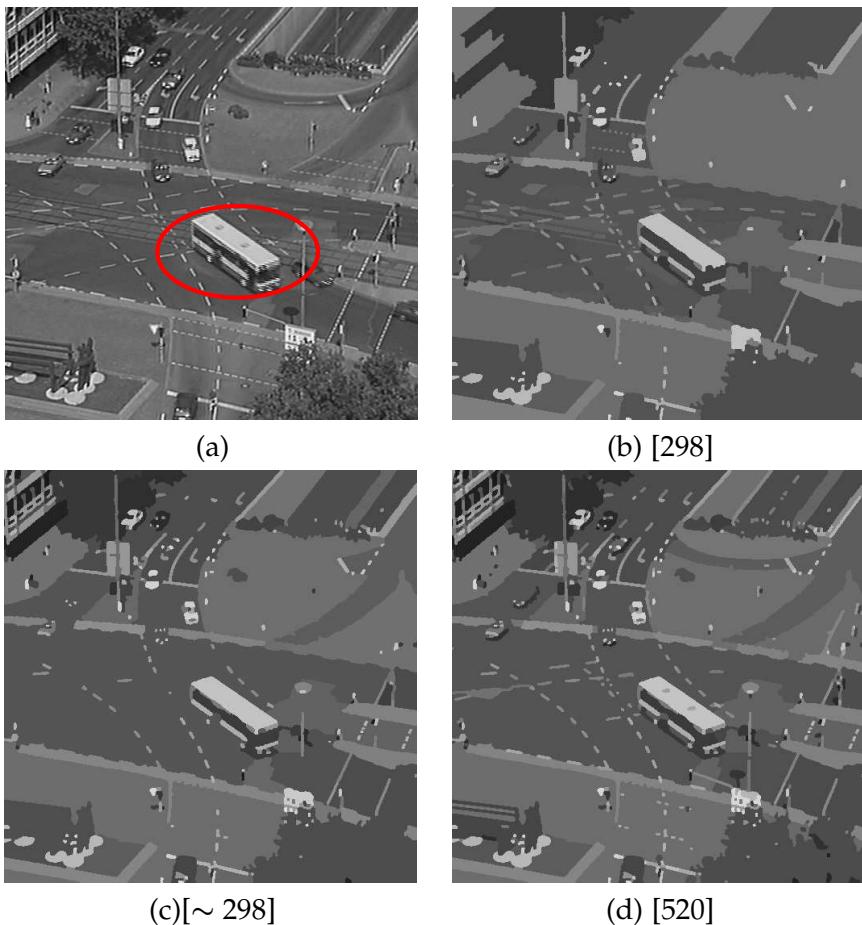


FIGURE 4.6 – La figure (a) représente l'image étudiée comportant un blob. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b). (d) coupe horizontale extraite manuellement et exhibant un niveau de détail de l'objet d'intérêt équivalent à celui de la coupe (b).

l'inverse, la coupe horizontale (c) ne montre qu'une partie des structures du piétons. On relèvera que les parties manquantes ont été absorbées par la chaussée ce qui confirme l'absence dans la hiérarchie d'une région unique décrivant le piéton.

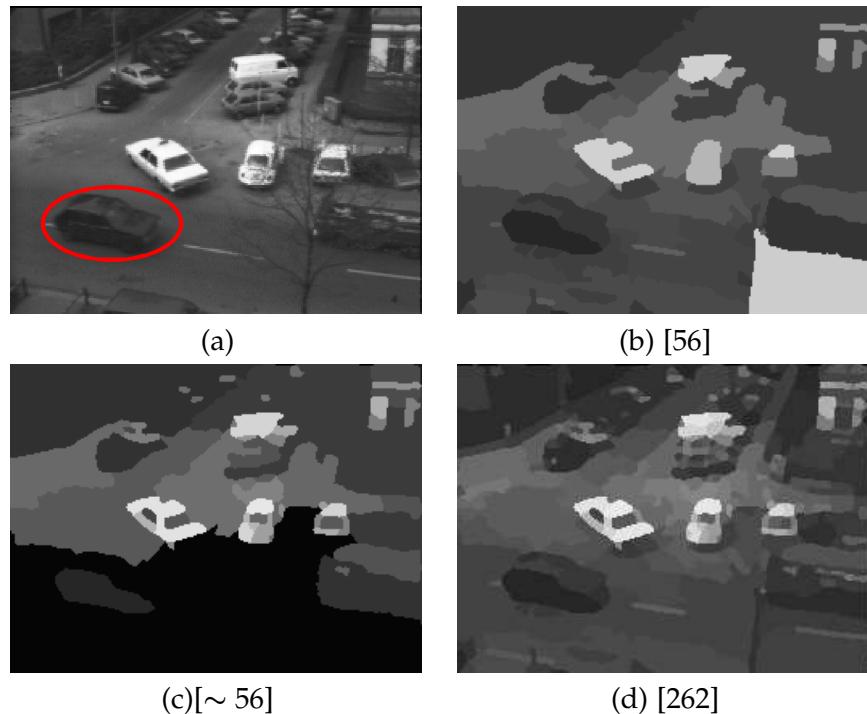


FIGURE 4.7 – La figure (a) représente l'image étudiée comportant un blob. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b). (d) coupe horizontale extraite manuellement et exhibant un niveau de détail de l'objet d'intérêt équivalent à celui de la coupe (b).

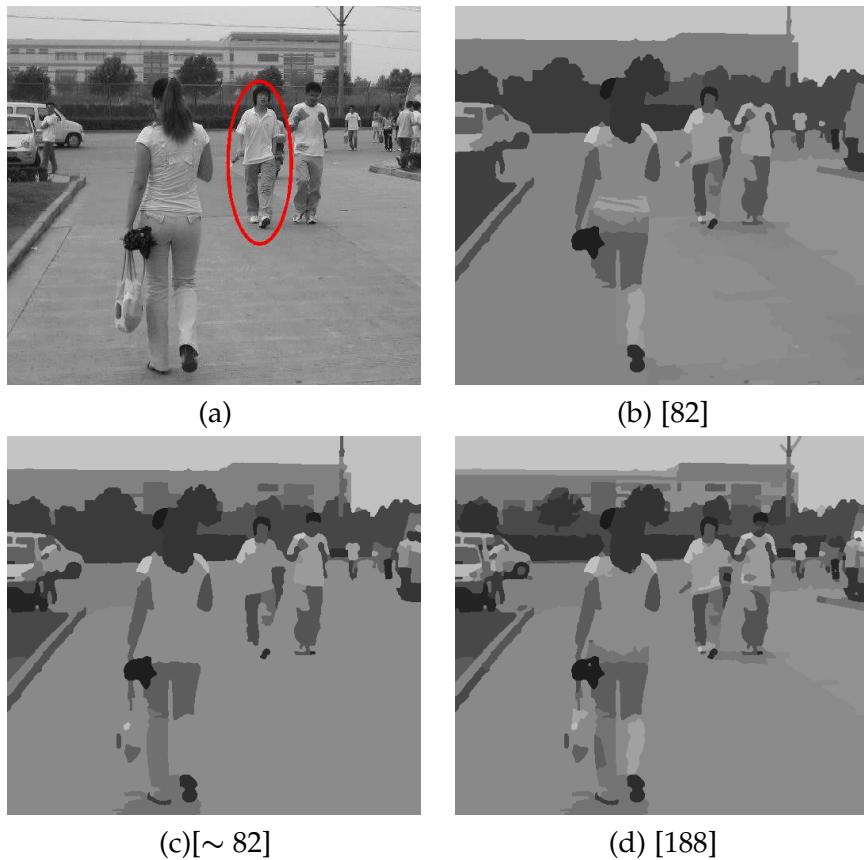


FIGURE 4.8 – La figure (a) représente l'image étudiée comportant un blob. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b). (d) coupe horizontale extraite manuellement et exhibant un niveau de détail de l'objet d'intérêt équivalent à celui de la coupe (b).

#### 4.2.2 Cas de plusieurs blobs

Nous allons maintenant analyser le comportement de notre approche sur des images comportant plusieurs blobs. L'énergie spatiale utilisée combine la distance euclidienne avec le critère ensembliste définie ci-dessus.

La figure ??-(a) montre les blobs considérés dans l'image du trafic routier périphérique. Le premier blob entoure la voiture et le deuxième blob encercle le numéro de l'autoroute indiqué sur le panneau. La coupe optimale (b) contient 116 régions et décrit parfaitement les objets d'intérêt. On remarquera que le deuxième panneau est complètement absorbé. Ce dernier représenté probablement par une sous hiérarchie indépendante des sous-hiéronymies interagissant avec les blobs, est négligé par l'énergie spatiale. L'énergie employée, saute les zones d'intérêt n'interagissant pas directement avec le voisinage des blobs sélectionnés. La coupe horizontale (c), extraite manuellement, montre la description donnée de l'image à nombre de régions équivalent. On remarquera que la description de la voiture est incomplète par rapport à celle de proposée par la coupe (c). En revanche, le numéro de l'autoroute est bien présent dans cette coupe. Cependant, ceci n'est pas surprenant, puisque ces régions sont d'une part homogène et d'autre part radiométriquement différent de leur environnement immédiat. On relèvera aussi que cette partition certains détails sont décrit dans cette coupe horizontale (flèches, contenu du deuxième panneau) tandis que dans notre coupe, ces détails sont ignorés pour les mêmes raisons qu'on a évoqué ci-dessus.

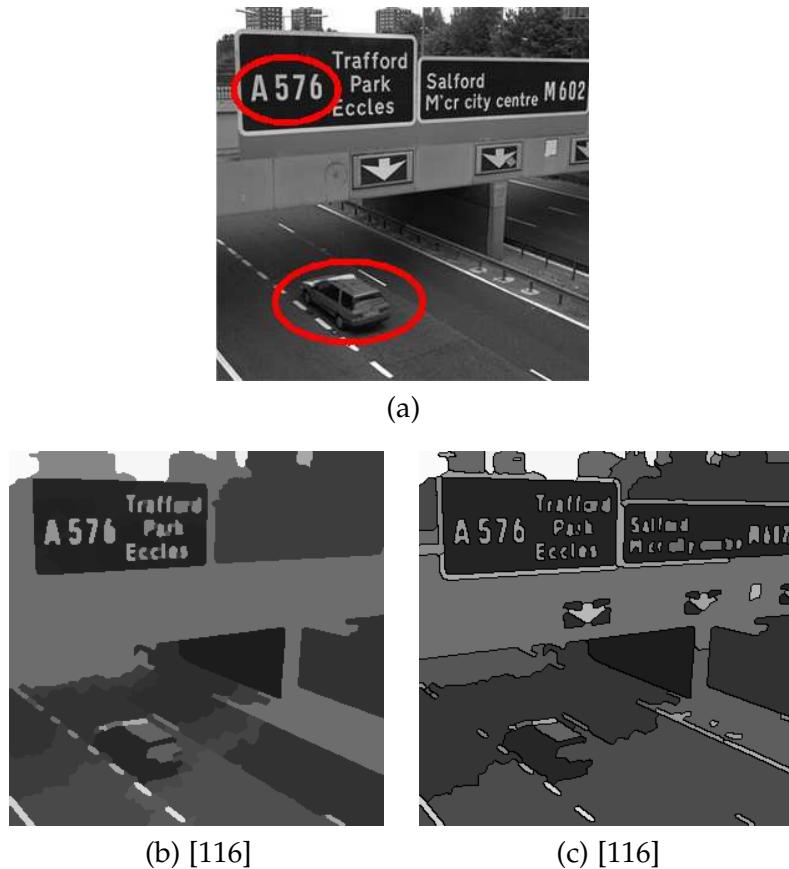


FIGURE 4.9 – (a) image comportant deux blobs. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b).

Dans l'exemple suivant (figure ??), nous avons entouré l'image par trois blobs. L'image représente une vue aérienne d'un aéroport. Les trois blobs délimite grossièrement trois avions au sol. La coupe optimale (b) est un exemple saisissant du comportement de l'énergie. Avec un nombre équivalent de régions par rapport à la coupe horizontale (c), notre coupe décrit le contenu des objets d'intérêt ainsi que les régions les avoisinant mais elle ignore complètement les structures composant la moitié gauche de l'image. La coupe horizontale (c) décrit également les régions d'intérêt, ce qui n'est pas surprenant, puisque la dissimilarité radiométrique de ces objets dans leurs voisinages est forte (régions blanches sur fond gris). Néanmoins, elle exhibe aussi des détails sur le contenu de la partie gauche de l'image. Est-ce un inconvénient ou un avantage pour l'utilisateur, ceci dépend de l'intérêt qu'il porte au zones éloignées des blobs sélectionné. Si l'utilisateur se focalise sur des régions se trouvant dans une partie bien précise de l'image, on peut supposer que le contenu du reste de l'image ne l'intéresse pas. Du moins, que cette partie ne lui sera pas d'une grande utilité dans l'interprétation et l'analyse du résultat. Le fait que l'énergie ignore certaines zones éloignées des blobs dans l'image montre également que ces zones ne participent pas à l'interprétation des zones étudiées et peuvent donc être ignorées.

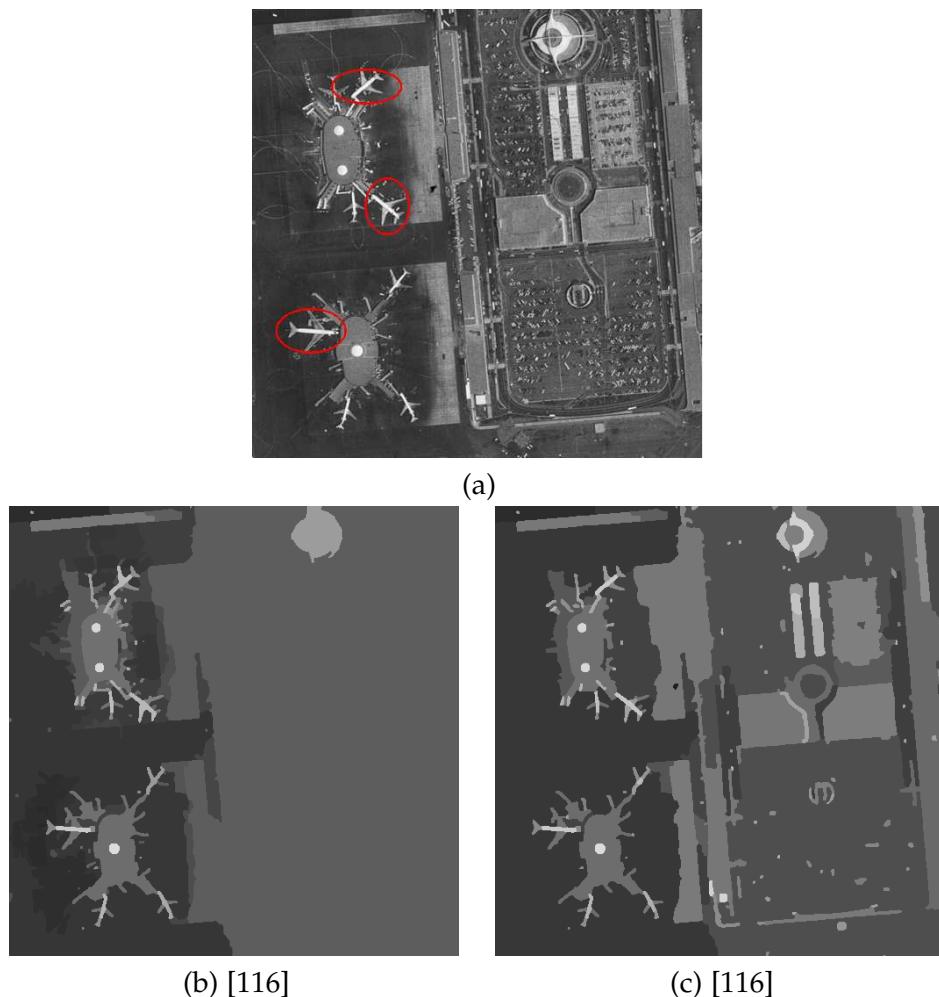
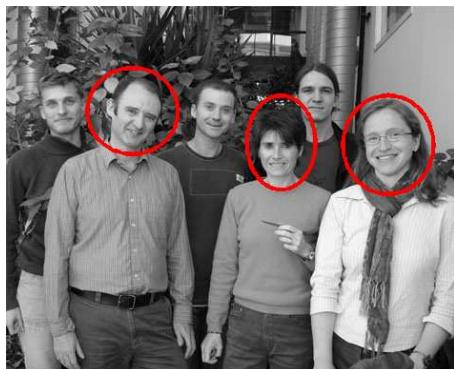


FIGURE 4.10 – (a) image comportant trois blobs. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b).

L'image (a) de la figure ?? illustre un groupe d'individus sur lequel nous avons délimité trois visages. La coupe horizontale (c) montre la partition extraite manuellement et comportant un nombre équivalent de régions que celui de la coupe optimale (b) obtenue avec l'énergie spatiale. La comparaison entre ces deux coupes en terme de description des objets d'intérêt est presque identique. On notera tout de même que certain détails inclus dans les blobs, tels que le nez du visage à droite de l'image ou les yeux et le front du visage à gauche de l'image, sont présents dans la coupe horizontale (c) mais absent de la coupe optimale (b). Ce résultat n'est pas surprenant, puisque ces régions, incluses dans le blob, ont été absorbées par une région plus grande et incluse dans le blob à cause de la sous-additivité de l'énergie dans ce type de configurations. On remarquera que les visages en arrière plan sont aussi présent dans la coupe horizontale tandis que la coupe optimale (b) les ignorent complètement ou partiellement.



(a)



(b) [142]



(c) [142]

FIGURE 4.11 – (a) image comportant trois blobs. (b) coupe optimale au sens de l'énergie spatiale, extraite de la hiérarchie. (c) coupe horizontale de la hiérarchie comportant un nombre équivalent de régions que la figure (b).

### 4.2.3 Energie spatiale multi-potentielle

Nous allons analyser le comportement de notre énergie spatiale dans le cas où le voisinage de la région est pris en compte dans le calcul de l'énergie relative à un blob. Nous utiliserons donc l'énergie qu'on a noté  $E_M$  et qui s'apparente à ce qu'on avait désigné dans le chapitre précédent par énergie multi-potentielle.

La coupe optimale (c) de la figure ?? illustre la partition obtenue en incluant le voisinage de la région dans le calcul de l'énergie spatiale. Elle est de taille plus importante que celle obtenue avec la précédente énergie. Cependant, elle exhibe plus de détails sur le contenu de l'objet d'intérêt par rapport à la coupe (b). On peut distinguer clairement des structures telles que les fenêtres, les roues, etc. Ces détails sont absorbés et donc absents dans la partition (b). Par ailleurs, la coupe optimale (c) ignore également le contenu les zones d'intérêt éloignées du blobs tels que les panneaux. Il semble donc que l'introduction du voisinage de la région dans le calcul de l'énergie pénalise la fusion de certaines structures incluses dans le blob, à l'inverse de ce qui se produisait quand l'énergie était mono-potentielle. En tenant compte du voisinage d'une région incluse dans le blob, celle-ci peut être au contact de régions intersectant le blob, or la pénalisation est plus forte pour ce type de régions, et puisque leurs énergies sont comptabilisées dans l'énergie totale de la région, il en résulte un comportement sur-additif de l'énergie.



(a)

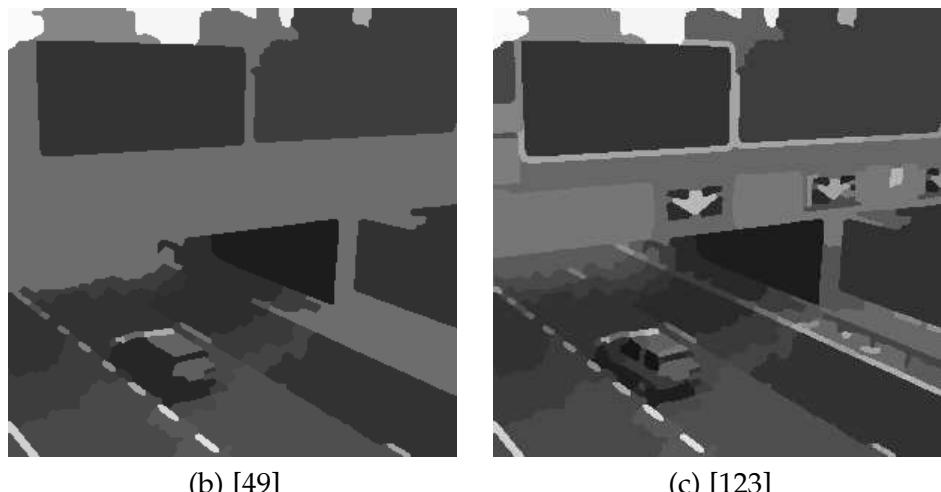


FIGURE 4.12 – (a) image comportant un blob. (b) coupe optimale au sens de l'énergie spatiale mono-potentielle, extraite de la hiérarchie. (c) coupe optimale au sens de l'énergie spatiale multi-potentielle.

Les figures ?? et ?? illustrent également le même phénomènes. Les coupes optimales obtenues avec l'énergie multi-potentielle sont certes de plus grandes tailles mais elles exhibent plus de détails sur les blobs et les zones limitrophes à ces derniers.



(a)



(b) [82]



(c) [203]

FIGURE 4.13 – (a) image comportant un blob. (b) coupe optimale au sens de l'énergie spatiale mono-potentielle, extraite de la hiérarchie. (c) coupe optimale au sens de l'énergie spatiale multi-potentielle.

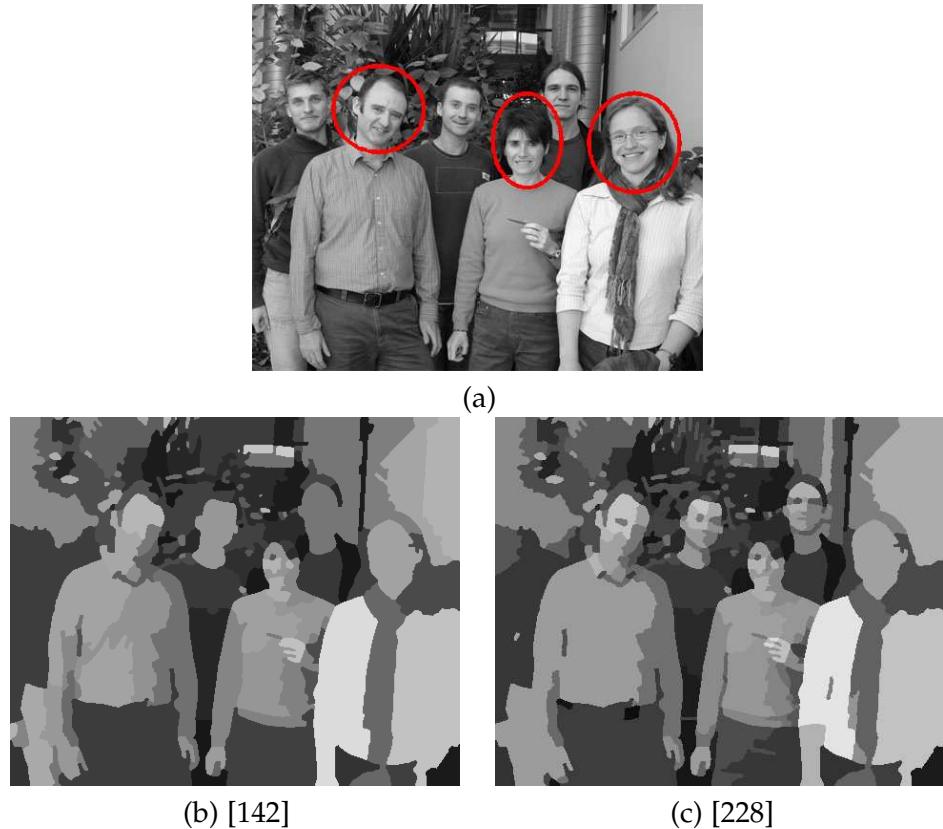


FIGURE 4.14 – (a) image comportant trois blobs. (b) coupe optimale au sens de l'énergie spatiale mono potentielle, extraite de la hiérarchie. (c) coupe optimale au sens de l'énergie spatiale multi potentielle.

#### 4.2.4 Discussion

En dépit de sa construction ad-hoc, l'énergie spatiale utilisée, pour extraire une partition optimale en intégrant un *a priori* spatial, remplit parfaitement son rôle. Elle permet à l'utilisateur de se focaliser sur certaines zones de l'image et d'ignorer des zones extérieures aux blobs. Par « ignorer », nous entendons que la minimisation de l'énergie aboutit à une description grossière des objets externes aux blobs et à une description plus détaillée des régions internes et limitrophes aux blobs. Par ailleurs, la partition obtenue est optimale, au sens de cette énergie, sur l'ensemble des coupes que l'on peut extraire de la hiérarchie qu'elles soient horizontales ou non. Par conséquent, la coupe sélectionnée est rarement horizontale, ce qui est impossible d'obtenir en explorant manuellement la hiérarchie à travers ses coupes optimales, au sens de l'énergie affine utilisée pour la construction de la hiérarchie.

Les règles établies pour la construction de notre critère ensembliste ont permis d'adapter le comportement de l'énergie en fonction de différentes situations. L'objectif était de donner à l'énergie la propriété de sous-additivité dans certaines configurations et la propriété de sur-additivité dans d'autres cas. La condition sur la taille de la région relativement à celle du blob nous semble nécessaire malgré que l'utilisation de ce type de condition -comparaison entre surfaces de régions- fasse craindre des comportements imprévisibles. Néanmoins, cette condition est un choix de bon sens, la fusion doit être pénalisée si la région interne au blob

se voit proposer une fusion avec une région dont la surface résultante serait plus grande que la taille du blob. Cette condition nous paraît donc inévitable mais on aurait pu l'assouplir en faisant appel au ratio « surface région/surface blob », avec un seuil  $\alpha$ .

La coupe optimale extraite par l'énergie spatiale dépend fortement du contenu de la hiérarchie. Ce contenu est intimement lié aux énergies d'attaches aux données et de régularisation employées. L'approche proposée ne crée pas d'informations supplémentaires, que ce soit des frontières ou des régions (l'un va avec l'autre). Par contre cette énergie permet de restructurer l'information recherchée par l'utilisateur en fonction de l'*a priori* injecté.

## CONCLUSION DU CHAPITRE

Ce court chapitre a proposé une approche de sélection d'une partition optimale en fonction de l'*a priori* spatial injecté par l'utilisateur. Cette méthode rentre dans le cadre de la segmentation semi-interactive d'une image. La manipulation d'une hiérarchie de régions n'est pas une chose aisée, elle offre à l'utilisateur un volume de coupes riches en partitions allant du détail le plus grossier au détail le plus fin. Elle permet donc une interaction entre l'utilisateur et la structure qui a pour objectif, généralement la sélection d'une coupe satisfaisante. Au lieu de parcourir manuellement la hiérarchie à la recherche d'une coupe « idéale », nous avons proposé de reporter cette interactivité sur l'image elle même en lieu et place de la structure. Dans le contexte où l'utilisateur possède une idée grossière sur la (les) région(s) qu'il recherche, il est préférable de lui offrir un outil par lequel il peut introduire des localisations grossières des zones d'intérêt. Cet outil, qui peut prendre la forme d'un pinceau, est de notre point de vue plus intuitif et adapté, que le parcourt d'une hiérarchie en déplaçant la ligne de coupe de haut en bas ou de bas en haut. La sélection des zones spatiales de l'image peut ne pas se cantonner à encercler une région par une ellipse. On peut facilement envisager que l'utilisateur, réalise des croix ou des courbes non régulières afin de rendre le processus moins usant et plus rapide. Des méthodes d'estimation des paramètres d'une ellipse, à partir d'un certain nombre de point, existent et peuvent être intégrées dans le processus. Une fois l'*a priori* spatial construit, l'énergie spatiale décrite dans ce chapitre permettra d'extraire rapidement une coupe optimale correspondant à l'*a priori* spatial sélectionné. On notera que le résultat de cette segmentation semi-interactive dépend du contenu de la hiérarchie préalablement construite par l'intermédiaire d'un algorithme quelconque. Néanmoins la richesse de ce contenu, a permis de constater expérimentalement que les coupes sélectionnées répondent aux exigences préalablement établies par l'utilisateur.



# CONCLUSION GÉNÉRALE ET PERSPECTIVES

Nous allons, pour terminer, dresser nos conclusions sur le travail accompli et examiner les différentes perspectives qui s'offrent à nous. La *conclusion* générale se divise en deux parties. La première partie fait le bilan de l'approche contextuelle agrégative proposée dans le chapitre *II* alors que la deuxième réalise une synthèse de l'approche ensemble-échelle par injection d'un *a priori* spatial, introduite dans le chapitre *III*. La partie *perspectives* respectera quasiment le même schéma. On abordera les différentes voies de recherches, à la fois théoriques et expérimentales, concernant les deux principales approches proposées dans cette thèse. Mais on évoquera également des perspectives plus générales portant sur la représentation hiérarchique des images qu'elle soit construite avec *a priori* ou sans.

## 5.1 CONCLUSIONS

L'approche ensemble-échelle offre une analyse multi-échelle bas-niveau de l'image sous la forme d'une hiérarchie de régions. Cette description est plus riche mais également plus volumineuse qu'une partition plate. Le formalisme de cette approche repose sur la combinaison de deux énergies séparables et antagonistes. La première est généralement sur-additive et traduit la fidélité aux données. La deuxième est sous-additive et mesure la complexité de la partition.

Cette description bas-niveau relègue le choix de(s) l'échelle(s) d'analyse à une phase ultérieure dépendant généralement de l'application visée. Ce choix est naturellement guidé par les connaissances haut-niveau liées aux spécificités de l'application. Cette démarche classique correspond au paradigme de vision énoncé par David Marr à la fin des années 70, celui du traitement séquentiel des deux analyses : bas et haut-niveau et dans cet ordre. Néanmoins, un nombre croissant de travaux ont commencé à coupler ces deux opérations à savoir localisation et reconnaissance afin d'accroître la performance des systèmes. Nous avons pris le partie d'explorer cette deuxième voie dans notre travail de recherche. Le principe repose sur l'injection d'*a priori* dans la construction d'une hiérarchie de régions afin d'améliorer son contenu, de répondre au mieux aux besoins de l'application ou de faciliter les étapes qui suivront dans un système de vision. Le choix d'intégrer des connaissances *a priori* a été principalement motivé par les difficultés éprouvées par l'approche ensemble-échelle dans la segmentation de certains types d'images complexes.

### 5.1.1 Construction hiérarchique supervisée

La supervision est une approche souvent adoptée pour labelliser les différentes régions de l'image, à condition que le contenu de l'image soit connu et qu'il puisse être représenté par un nombre raisonnable de classes. L'approche que nous avons adopté couple un superviseur sous forme d'un coût contextuel avec deux critères agrégatifs bas-niveau et antagonistes. Cependant, l'objectif n'est pas d'étiqueter chaque région de la hiérarchie par un label. Il s'agit de vérifier si deux régions ont une forte probabilité d'appartenir à une même classe, auquel cas leur fusion sera favorisée. Dans le cas contraire, la fusion sera pénalisée. La région est caractérisée par un vecteur d'attributs sélectionnés pour décrire au mieux les différentes classes (apprentissage). La région porte également un vecteur des probabilités d'appartenance aux différentes classes étudiées. Ainsi, le coût contextuel mesure la dissimilarité entre les vecteurs de probabilité respectifs de deux régions adjacentes. De ce fait, nous influençons la décision locale de fusion dans le but de réduire les fusions aberrantes. L'approche proposée découle indirectement de l'approche ensemble-échelle mais ne respecte plus la formulation énergétique initiale. Néanmoins, cette formulation agrégative supervisée que nous avons proposé a montré qu'elle était plus performante. La description hiérarchique que l'on obtient ne peut plus être qualifiée d'analyse multi-échelle au sens définit par l'approche ensemble-échelle. Elle représente une analyse multi-niveau de l'image dans laquelle on peut explorer les structures de l'image du détail le plus grossier au détail le plus fin.

Cette approche nous paraît bien adaptée dans le cas où le nombre de classes est faible et le nombre d'attributs pas trop élevé. La performance est liée à celle du superviseur mais également à la séparabilité entre les classes étudiées. Nous n'avons pas pu tester la performance de l'approche sur un plus grand nombre de classes car nous ne disposons pas d'une base d'images adéquates. Néanmoins, nous pensons que plus le nombre de classes augmentent et que la séparabilité entre ces classes d'estampent et plus la performance diminue. L'utilisation d'un superviseur linéaire est également un handicap puisqu'il décrit des classes de formes elliptiques. Le choix de la pondération entre le coût contextuel et les deux coûts bas-niveaux est également un problème épique, commun à plusieurs approches de segmentation que l'on a déterminé de manière expérimentale. L'extension au cas 3D ne pose *a priori* aucune difficulté, au contraire elle devrait améliorer les résultats puisque les attributs géométriques seront plus adaptés que dans le cas 2D où les formes des régions d'intérêt peuvent être très variables, empêchant leur description par des attributs de forme. Pour terminer, il est évident que cette approche supervisée est préconisée uniquement quand la classe d'images étudiée présente le même contenu.

### 5.1.2 Construction d'un ensemble-échelle via un *a priori* spatial

Nous avons exploré la possibilité d'introduire un *a priori* spatial dans la construction hiérarchique afin de se focaliser sur certaines zones de la scène. L'idée du départ était simple : conférer un degré de pertinence plus élevé à certaines zones de l'image afin de préserver les structures composant ces zones. Le travail de recherche s'est concentré sur la construction d'une énergie sous-additive spatiale. Nous avons aboutit à différentes formulations de cette énergie en se basant sur la définition de la notion de potentiel spatial. Une énergie spatiale combine à la fois une information métrique exprimée par une distance et une information ensembliste portant sur la région et l'ensemble des blobs.

La construction d'un ensemble-échelle via un *a priori* spatial a montré que les zones d'intérêt, encerclées par des blobs, sont mieux préservées. Elles résistent plus longtemps au processus de fusion, ce qui a pour effet d'augmenter leurs échelles de disparition. Cependant, le choix de l'énergie spatiale adéquate demeure flou. Plusieurs modes et combinaisons donnent des résultats sensiblement similaires. Nous avons relevé sans surprise que l'absence d'une information sur le contour dans l'énergie globale a pour effet de réduire la qualité des frontières des régions. Nous avons proposé ainsi de regrouper l'information contour et l'information spatiale sous la forme d'une combinaison convexe de deux énergies. Le choix de la pondération a été déterminé expérimentalement et permet de concilier à la fois la régularité des contours et l'*a priori* spatial injecté.

## 5.2 PERSPECTIVES

### 5.2.1 Injection d'*a priori* radiométrique et/ou géométrique

La construction hiérarchique supervisée que nous avons proposé utilise un superviseur euclidien. Le choix s'est porté sur ce dernier pour sa simplicité et sa complexité (temps de calcul) peu élevée. D'autres approches de classification plus performantes existent. Il serait intéressant d'explorer ces autres approches ou de s'en inspirer afin de construire différents coûts contextuels. Le classifieur euclidien nous permet d'estimer un vecteur des probabilités *a posteriori* pour chaque région. Le coût contextuel prend alors la forme d'une mesure de dissimilarité entre deux vecteurs de probabilité. Les autres méthodes de classification ne fournissent pas forcément un vecteur des probabilités *a posteriori* d'où la nécessité d'une étude plus approfondie afin de les adapter à notre approche. L'utilisation d'un coût contextuel issu d'un superviseur plus performant peut permettre de traiter un nombre de classes plus conséquent et faire face au chevauchement des classes grâce à une modélisation plus fidèle des formes des classes. Rappelons que celles-ci son supposées gaussiennes dans le cas d'un superviseur euclidien.

Le problème de la pondération, évoqué ci-dessus, est également une question ouverte. Nous introduisons une supervision dès le début du processus de fusion avec un poids fixe. Or, on peut raisonnablement penser que ce poids doit être dynamique et croissant. Néanmoins, modifier le poids durant le processus de fusion peut être source d'instabilité engendrant des phénomènes d'inversion imprévisibles dans la construction de la hiérarchie. La question d'une pondération variable le long du processus de fusion paraît être un problème très difficile.

### 5.2.2 Injection d'*a priori* spatial

La construction d'une énergie spatiale a aboutit sur différentes formulations offrant un nombre de combinaison assez important. Néanmoins, cela s'est fait au détriment de l'information sur les contours. Nous avons résolu partiellement ce problème en proposant une combinaison convexe entre deux termes. On devrait explorer d'autres manières plus directes d'intégrer une information sur le contour (périmètre, courbure, convexité, ...) dans le terme énergétique afin de s'affranchir du choix de la pondération. Si ce problème a une solution, il se peut que cela se fasse au détriment de la condition de sous-additivité de l'énergie, condition nécessaire pour respecter le formalisme ensemble-échelle. Néanmoins, une énergie non

sous-additive telle que la convexité a montré le rôle positive et pertinent qu'elle joue dans la construction hiérarchique.

L'énergie spatiale présentée dans le chapitre *III* prend la forme d'une norme d'un vecteur. Elle représente l'intensité du vecteur. Chaque composante du vecteur représente l'énergie spatiale relativement à un blob de l'ensemble des blobs. Aucune information sur la direction de ce vecteur n'est incorporée. Ainsi, la comparaison des énergies spatiales entre deux régions adjacentes se résume à la seule comparaison entre les intensités de leurs vecteurs d'énergie respectifs. L'intégration de la direction dans le terme énergétique permettra d'exprimer de manière plus précise l'interaction entre les deux régions vis à vis de l'ensemble des blobs.

Nous avons envisagé l'énergie spatiale comme étant sous-additive pour les raisons que l'on a exposé dans le chapitre *III*. Néanmoins, des investigations pourront être menées afin de construire une énergie spatiale sur-additive. On pourra ainsi préserver le terme de régularisation mais on devra soit combiner l'énergie d'attache aux données avec l'énergie spatiale soit construire une énergie sur-additive d'attache aux données intégrant implicitement *l'a priori* spatial.

Dans de nombreux problèmes de vision, l'information spatiale prend une place importante. La détection d'objets tels que les piétons, les voitures, les panneaux de signalisation en sont des exemples. Cette information spatiale est devenu accessible grâce à l'apparition de détecteurs performants et robustes souvent couplés avec les systèmes d'acquisitions d'images. Ainsi, l'approche proposée ici trouve naturellement sa place dans ce type de problème. Un système de détection comporte une étape de description de l'image qui prend généralement la forme d'une carte de contour ou d'une partition en régions de l'image. Cette description peut être remplacée par une représentation hiérarchique en régions de l'image ou sous la forme d'une carte de contours valués de l'image. Cette dernière est une version orientée frontière de l'ensemble-échelle construit. De nombreux travaux dans ce domaine reposent sur une représentation de l'image sous la forme d'une carte de contour binaire (issue d'un détecteur de contour classique). L'objectif est de rechercher, par des mises en correspondance, les objets d'intérêt en utilisant des modèles de formes. Les problèmes liés au choix du seuil pour construire une carte contour et celui de la non fermeture des contours affectent la performance des systèmes de détection. Introduire une carte de contours valués paraît être plus approprié car d'une part, le problème de non fermeture des contours est absent. D'autre part, on peut exploiter la description multi-échelle de l'image afin d'affiner la recherche. Il serait donc intéressant d'intégrer cette approche dans un tel système et d'évaluer sa performance par rapport à une représentation classique de l'image.

### 5.2.3 Perspectives générales

Au delà du sujet traité dans cette thèse, nous souhaitons dresser de manière générale les différents axes de recherches concernant *la description hiérarchique d'une image*, qui restent à explorer.

La description d'une image par une hiérarchie de régions -qu'elles soit issue d'une analyse multi-échelle ou multi-niveau, sans ou avec *a priori*- est plus riche qu'une description de l'image par une simple partition. Cependant, cette description est plus volumineuse (suite de partitions monotone). Utiliser une hiérarchie de régions en tant que représentation est plus coûteux mais permet de préserver une information contextuelle plus importante. Outre cet aspect de coût élevé,

l'utilisation de la représentation hiérarchique en tant que description soulève de nombreux problèmes encore inexplorés.

La première difficulté est la *comparaison* entre deux hiérarchies de régions construites à partir de deux images ayant un même contenu. L'expérimentation a montré que les hiérarchies obtenues sont topologiquement (au sens de la structure de leurs graphes) différentes. L'ordre d'apparition et de disparition des régions d'intérêt, d'une hiérarchie à une autre, est différent. Une même région d'intérêt (pensez à un os du métacarpe) peut fusionner avec une région dans une hiérarchie qui est totalement différente dans une autre hiérarchie. On observe que le chemin des fusions successives menant d'une région d'intérêt au sommet de la hiérarchie est différent dans une autre hiérarchie d'une autre image ayant le même contenu sémantique. Cette différence de topologie rend l'utilisation de techniques de *Graph ou Tree Matching* impossible. De plus, étant donné la taille des hiérarchies des images le temps d'exécution devient rédhibitoire. La comparaison de deux hiérarchies doit se concentrer sur ces zones remarquables. Il faut alors dans un premier temps définir des mesures qui permettent de caractériser et d'extraire ces zones. La persistance et les échelles d'apparition et de disparition peuvent aider à cette tâche mais ne sont pas suffisantes. Elles dépendent de la radiométrie de l'image et peuvent par conséquent varier significativement selon la distribution des niveaux de gris de l'image. On peut également s'appuyer sur l'application visée pour construire des mesures qui permettent de rechercher et d'extraire les zones d'intérêt.

L'extraction automatique d'une coupe dans une hiérarchie est également un problème difficile qui requiert l'utilisation de connaissances haut-niveau issues de l'application visée. Notons que la recherche de l'échelle d'analyse adéquate pour une application est un problème plus contraint que ce dernier puisque l'on ne s'intéresse qu'aux coupes horizontales de la hiérarchie. Dans le cadre d'une application spécifique, la première approche consiste à minimiser une énergie sur l'ensemble des partitions issues de la hiérarchie. Cette énergie devra intégrer des connaissances *a priori* comme cela a été proposé de manière heuristique dans le chapitre IV. Cependant, la partition optimale ne coïncidera pas obligatoirement avec une coupe horizontale de la hiérarchie (Rappelons qu'une coupe horizontale représente la partition optimale de l'image à une certaine échelle au sens de l'énergie affine). La partition optimale ne représentera en aucun cas une description de l'image à une certaine échelle (définie au sens du formalisme ensemble-échelle) mais correspondra à une partition comportant des zones caractériser par des échelles d'analyse différentes. L'extraction d'une coupe horizontale correspond à la recherche d'une échelle d'analyse, celle-ci est fortement liée à la radiométrie de l'image. Il paraît donc difficile de trouver une seule échelle d'analyse pour une application donnée mais plutôt un intervalle d'échelles d'analyse. Nous pensons que cet intervalle doit caractériser certains objets de l'image et que l'estimation des échelles de ces objets devrait permettre de délimiter cet intervalle.

L'approche ensemble-échelle tente de respecter un concept fondamental selon lequel les objets dans l'univers sont hiérarchisés. Une simple partition ne peut représenter cette notion. Décrire une image par une hiérarchie semble être plus naturel, cependant cette description est si volumineuse qu'elle devient un frein à son utilisation. Il est important de mener une réflexion sur la question de la réduction de l'information contenue dans une hiérarchie. Actuellement, une hiérarchie, même si elle segmente correctement les objets d'intérêt, comporte un nombre important de régions intermédiaires, nécessaires au processus de fusion pour

construire les prochaines régions d'intérêt. La sur-segmentation par la **LPE** produit également un nombre de régions très élevé mais nécessaire pour la construction de régions d'intérêt. Est-il possible de créer des mécanismes de réductions qui permettent de construire une hiérarchie plus simple contenant les zones remarquables de l'image et reflétant la notion de hiérarchisation des objets. Le problème reste posé.

Nous terminerons cette conclusion en soulignant l'important travail d'analyse effectué durant cette thèse. L'observation et la manipulation des hiérarchies est une tâche longue et exige un examen pointilleux.

## PARTIE ANNEXES



# DESCRIPTEURS DE RÉGIONS

A

Cette annexe présente de manière très succincte les définitions des différents attributs de région, utilisés dans le chapitre II, lors de la phase d'apprentissage.

Soit  $R$  une région de l'image  $I$ . Le couple  $(x, y)$  désigne les coordonnées d'un pixel  $p$  dans une image.  $[0, n]$  désigne l'intervalle discret du niveau de gris de l'image.

## ENTROPIE

Soit  $P(R) = (p_i)_{i=1,\dots,n}$  la distribution de niveau de gris de la région  $R$ , estimée à partir de l'histogramme de la région :

$$H(R) = - \sum_{i=1}^n p_i \log_2(p_i)$$

L'entropie permet de caractériser le niveau de désordre. Une région homogène induit une entropie faible.

## CONTRASTE GLOBAL

$$C(R) = \max_{p,q \in R^2} \{I(p) - I(q)\}$$

## COMPACITÉ

Soit  $\mathcal{L}(R)$  le périmètre de la région  $R$  et  $|R|$  sa surface. La compacité de la région  $R$  est définie par :

$$C_p(R) = \frac{\mathcal{L}(R)}{4\pi|R|}$$

Cet attribut atteint son minimum 1 quand la région  $R$  est circulaire. Il prend de plus grandes valeurs pour des régions longilignes. Par ailleurs, la copacité est invariant par rotation, réflexion et changement d'échelle.

## MOMMENTS D'INERTIE

Soient  $\bar{x}$  et  $\bar{y}$  les coordonnées du barycentre de la région  $R$ , i.e

$$\bar{x} = \sum_{(x,y) \in R}^{|R|} xf(x,y),$$

and

$$\bar{y} = \sum_{(x,y) \in R}^{|R|} y f(x,y).$$

tel que  $f(x,y) = 1$  si  $(x,y) \in R$  et 0 sinon. Les moments d'inertie de la région  $R$  sont définies par :

$$\begin{aligned}\mu_{20} &= \sum_{(x,y) \in R}^{|R|} (x - \bar{x})^2 \cdot f(x,y), \\ \mu_{02} &= \sum_{(x,y) \in R}^{|R|} (y - \bar{y})^2 \cdot f(x,y), \\ \mu_{11} &= \sum_{(x,y) \in R}^{|R|} (x - \bar{x})(y - \bar{y}) \cdot f(x,y).\end{aligned}$$

Ces moments caractérisent la distribution des pixels de la région autour de son centre de gravité. Ils sont également invariant par rotation, translation et par homothétie.

## ORIENTATION PRINCIPALE

Elle représente la direction de l'axe majeur de l'inertie, i.e l'orientation de l'ellipse englobant la région. Elle est définie par le vecteur propre associé avec la plus grande valeur propre de la matrice d'inertie :

$$\mathbf{M} = \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix}$$

L'orientation principale d'un objet est définie par l'angle  $\phi$  entre l'axe des abscisses et l'axe majeur de l'ellipse englobant. Cette dernière possède les mêmes seconds moments centrés normalisés que ceux de la région.

$$\phi = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \quad -\frac{\pi}{4} \leq \phi \geq \frac{\pi}{4}.$$

## ELONGATION ELLIPTIQUE

L'élongation elliptique,  $E_{ell}$ , est le rapport entre les longueurs des axes majeur et mineur de l'ellipse englobante :

$$E_{ell} = \frac{a}{b}.$$

tel que

$$a = \left( \frac{2[\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2}]}{\mu_{00}} \right)^{\frac{1}{2}},$$

et

$$b = \left( \frac{2[\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2}]}{\mu_{00}} \right)^{\frac{1}{2}}.$$

## INTENSITÉ ELLIPTIQUE

Elle représente le ratio entre la surface de la région et la surface de ellipse englobante :

$$I_{ell} = \frac{\mu_{00}}{\pi \cdot a \cdot b}$$

## BOÎTE ENGLOBANTE

C'est le plus petit rectangle contenant la région  $R$ , centré à son barycentre et ayant pour axes de référence les axes majeur et mineur de l'ellipse englobante. Les coordonnées des deux points extrêmes du rectangle englobant notés,  $p_m = (x_m, y_m)$  et  $p_M = (x_M, y_M)$ , sont définis par :

$$x_m = \min_{x \in \delta R} x \text{ and } y_m = \min_{y \in \delta R} y$$

$$x_M = \max_{x \in \delta R} x \text{ and } y_M = \max_{y \in \delta R} y$$

## ELONGATION RECTANGULAIRE

C'est le ratio entre la longueur et la largeur des axes du rectangle englobant.

## LES MOMENTS DE HU

Les moments de HU sont des mesures de formes très populaires. Ils sont invariants par translation, rotation and changement d'échelle. Ils sont définies par les formules suivantes :

$$I_1 = \frac{1}{\mu_{00}^2} (\mu_{20} + \mu_{02});$$

$$I_2 = \frac{1}{\mu_{00}^4} \{ (\mu_{20} - \mu_{02})^2 + 4 \cdot \mu_{11}^2 \};$$

$$I_3 = \frac{1}{\mu_{00}^5} \{ (\mu_{30} - 3\mu_{12})^2 + (\mu_{03} - 3\mu_{21})^2 \};$$

$$I_4 = \frac{1}{\mu_{00}^5} \{ (\mu_{30} + \mu_{12})^2 + (\mu_{03} + \mu_{21})^2 \};$$

$$I_5 = \frac{1}{\mu_{00}^{10}} \{ (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}) \cdot [(\mu_{30} + \mu_{12})^2 - 3(\mu_{03} + \mu_{21})^2] \\ + (3\mu_{21} - \mu_{03})(\mu_{03} + \mu_{21}) \cdot [3(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] \};$$

$$I_6 = \frac{1}{\mu_{00}^7} \left\{ (\mu_{20} - \mu_{02}) \cdot [(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{03} + \mu_{21}) \right\};$$

$$I_7 = \frac{1}{\mu_{00}^{10}} \left\{ (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12}) \cdot [(\mu_{30} + \mu_{12})^2 - 3(\mu_{03} + \mu_{21})^2] \right. \\ \left. + (3\mu_{12} - \mu_{30})(\mu_{03} + \mu_{21}) \cdot [3(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] \right\}$$

tel que  $\mu_{pq}$  est le moment centré normalisé d'ordre  $(p + q)$

$$\mu_{pq} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q \cdot f(x, y)$$

tel que

$$f(x, y) = \begin{cases} 1 & \text{si } (x, y) \in R \\ 0 & \text{sinon} \end{cases}$$

## LES MASQUES DE LAW

C'est un ensemble de masques de taille  $3 \times 3$  convolués avec l'image :

$$\begin{aligned} \mathbf{ML}_1 &= \frac{1}{36} \begin{bmatrix} +1 & +2 & +1 \\ +2 & +4 & +2 \\ +1 & +2 & +1 \end{bmatrix} & \mathbf{ML}_2 &= \frac{1}{12} \begin{bmatrix} +1 & +0 & -1 \\ +2 & +0 & -2 \\ +1 & +0 & -1 \end{bmatrix} & \mathbf{ML}_3 &= \frac{1}{12} \begin{bmatrix} -1 & +2 & -1 \\ -2 & +4 & -2 \\ -1 & +2 & -1 \end{bmatrix} \\ \mathbf{ML}_4 &= \frac{1}{12} \begin{bmatrix} -1 & -2 & -1 \\ +0 & +0 & +0 \\ +1 & +2 & +1 \end{bmatrix} & \mathbf{ML}_5 &= \frac{1}{4} \begin{bmatrix} -1 & +0 & +1 \\ +0 & +0 & +0 \\ -1 & +0 & +1 \end{bmatrix} & \mathbf{ML}_6 &= \frac{1}{4} \begin{bmatrix} -1 & +2 & -1 \\ -1 & +2 & -1 \\ +0 & +0 & +0 \end{bmatrix} \\ \mathbf{ML}_7 &= \frac{1}{12} \begin{bmatrix} -1 & -2 & -1 \\ +2 & +4 & +2 \\ -1 & -2 & -1 \end{bmatrix} & \mathbf{ML}_8 &= \frac{1}{4} \begin{bmatrix} +2 & +0 & -2 \\ -1 & +0 & +1 \end{bmatrix} & \mathbf{ML}_9 &= \frac{1}{4} \begin{bmatrix} +1 & -2 & +1 \\ +1 & -2 & +1 \\ -2 & +4 & -2 \\ +1 & -2 & +1 \end{bmatrix} \end{aligned}$$

L'objectif est de caractériser la textures par la mesure d'énergie de région après convolution. Dans notre cas, l'image a été dans un premier temps convolué avec chaque masque afin d'obtenir neuf images filtrées, notées  $\{I_k\}_{k=1,\dots,9}$ . Ensuite, pour toute région extraite pour l'apprentissage, on calcule les différents énergies de texture :

$$\forall R \in P(\mathcal{D}), \quad T_k(R) = \frac{1}{|R|} \sum_{(x,y) \in R} I_k^2(x, y)$$

L'avantage d'utiliser ces attributs est la relation de récurrence qui existe entre une structure et ses sous structures :

$$\forall R_1, R_2 \in \mathcal{P}(\mathcal{D}), \quad R_1 \cap R_2 = \emptyset$$

$$T_k(R_1 \cup R_2) = \frac{|R_1| \cdot T_k(R_1) + |R_2| \cdot T_k(R_2)}{|R_1| + |R_2|}$$

Elle permet ainsi une rapide mise à jour des mesures de textures après la fusion.

# SUR-SEGMENTATION PAR LIGNE DE PARTAGE DES EAUX

B

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 interprétant 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 une infime perturbation, à son écoulement dans plusieurs bassins différents<sup>1</sup>. 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  $\mathcal{R}^n$  dans  $\mathcal{R}$  (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<sup>2</sup>.

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),

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 (Prêteux 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.

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]).

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, Stamen, 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 :

$$\forall p \in P \quad \forall b \in V(p), \quad (p; b) \in B \Leftrightarrow I(b) = \min_{v \in V(p)} I(v) \quad \text{et} \quad I(b) < I(p).$$

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  $(x; y) \in B^s$ ,  $(x; y) \in B$  ou  $(y; x) \in B$ .

Les composantes connexes de  $S$  définissent les bassins versants du relief.

L'algorithme de Marc Pierrot-Deseilligny est très efficace : Un balayage de l'image (celle du module du gradient 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).

# BIBLIOGRAPHIE

- R. Adams et L. Bishof. Seeded region growing. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 16 :641–647, 1994. (Cité page 12.)
- M. N. Ahmed, S. M. Yamany, N. A. Mohamed, et A. A. Farag. A modified fuzzy c-means algorithm for MRI bias estimation and adaptive segmentation. Dans *Proc. Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 1999. (Cité page 12.)
- F. Aldershoff et A. Salden. Multi-scale audio-video analysis and processing : Segmentations and arrangements. *Proceedings of the SPIE - The International Society for Optical Engineering*, 4519(9) :20–31. (Cité page 17.)
- L. Alvarez, F. Guichard, P.-L. Lions, et J.-M. Morel. Axioms and Fundamental Equations of Image Processing. *Arch. Rational Mech. Anal.*, 123 :199–257, 1993. (Cité page 16.)
- L. Alvarez, P. Lions, et J. Morel. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(3) :845–866, 1992. (Cité page 14.)
- L. Alvarez et J. Morel. Morphological approach to multiscale analysis, geometry-driven diffusion in computer vision. Chapitre 8, pages 199–257. Kluwer Academic Publishers, 1994. (Cité page 14.)
- A. Amini, T. Weymouth, et R. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9) :855–867, 1990. (Cité page 11.)
- G. Aubert et L. Blanc-Féraud. Some remarks on the equivalence between 2D and 3D classical snakes and geodesic active contours. *Intern. J. Computer Vision*, 34 (1) :19–28, February 1999. (Cité page 11.)
- Gilles Aubert et Pierre Kornprobst. *Mathematical Problems in Image Processing Partial Differential Equations and the Calculus of Variations*, volume 147. Paris, springer édition, 2006. (Cité page 11.)
- C. Ballester, V. Caselles, et M. González. Affine Invariant Segmentation by Variational Method. Dans *Proc. of 9th RFIA, Paris*, pages 379–390, January 1994. (Cité pages 22 et 23.)
- A. Basu, I. R. Harris, et S. Basu. Minimum distance estimation : The approach using density-based distances. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics*, North-Holland, 15 :21–48, 1997. (Cité page 54.)
- G. Bertrand, J.-C. Everat, et M. Couprise. Image segmentation through operators based upon topology. *Journal of Electronic Imaging*, 6(4) :395–405, 1997. (Cité page 178.)

- P. Besl et R. Jain. Segmentation through variable-order surface fitting. *IEEE Trans. Pattern Anal. Machine Intell.*, 10 :167–192, March 1988. (Cité page 13.)
- S. Beucher et C. Lantuejoul. Use of watersheds in contour detection. Dans *In Proceedings of Int. Workshop on Image Processing, Real-time Edge and Motion Detection/Estimation*, 1979. (Cité page 177.)
- S. Beucher et F. Meyer. The morphological approach to segmentation : the watershed transformation. Dans Dougherty, éditeur, *Mathematical Morphology in Image Processing*, Chapitre 12, pages 433–481. Marcel Dekker, 1993. (Cité page 29.)
- W. F. Bischof et T. Caelli. Parsing scale-space and spatial stability analysis. *Computer Vision, Graphics and Image Processing*, 42 :192–205, 1988. (Cité page 17.)
- M. Black, G. Sapiro, D. Marimont, et D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3) :421–432, 1998. (Cité page 14.)
- A. Blake et Michael Isard. *Active Contours*. London, springer verlag édition, 1998. (Cité page 11.)
- A. Blake et A. Zisserman. *Visual reconstruction*. MIT Press édition, 1987. (Cité page 11.)
- I. M. Bomze, M. Budinich, P. M. Pardalos, et M. Pelillo. The maximum clique problem. volume A de *Handbook of Combinatorial Optimization*, page 174. Kluwer, Boston, MA, in d.-z. du and p. m. pardalos, editors édition, 1999. (Cité page 39.)
- T. Brox, D. Farin, et P.H.N. de With. Multi-stage region merging for image segmentation. Dans *In Proceedings of the 22nd Symposium on Information Theory in the Benelux*, 2001. (Cité page 20.)
- John Canny. A computational approach to edge detection. *IEEE trans on Pattern Analysis and Machine Intelligence*, 8(6) :679–698, Nov 1986. (Cité page 10.)
- V. Caselles, R. Kimmel, et Sapiro G. On geodesic active contours. *International Journal of Computer Vision*, 22(1) :61–79, February 1997. (Cité page 11.)
- M. Ceccarelli, F. Musacchia, et A. Petrosino. A fuzzy scale-space approach to feature-based image representation and retrieval. Dans *Brain, Vision, and Artificial Intelligence. First International Symposium, BVAI 2005. Proceedings (Lecture Notes in Computer Science)*, volume 3704, pages 377–85, 2005. (Cité page 17.)
- Y. Chang et X. Li. Adaptive image region growing. *IEEE Trans on Image Proc.*, 3 : 868–872, 1994. (Cité page 12.)
- N. Chehata, F. Jung, M. Pierrot-Deseilligny, et G. A. Stamou. Region-based matching approach for 3D-roof reconstruction from HR satellite stereo pairs. Dans *DICTA03 , Digital Image Computing Techniques and Applications*, volume II, pages 889–898, décembre 2003. (Cité page 36.)
- A. Chehikian. Recursive filters for gradient estimation and contour extraction using B-Spline interpolation. *Traitement du Signal*, 14(1) :29, 1997. (Cité page 10.)
- Yunmei Chen, Hemant D. Tagare, Sheshadri Thiruvenkadam, Feng Huang, David Wilson, Kaundinya S. Gopinath, Richard W. Briggs, et Edward A. Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3) :315–328, 2002. (Cité page 43.)

- Yunmei Chen, S. Thiruvenkadam, H.D. Tagare, Feng Huang, D. Wilson, et E.A. Geiser. On the incorporation of shape priors into geometric active contours. Dans *Variational and Level Set Methods in Computer Vision, 2001. Proceedings, IEEE Workshop on Volume*, 2001. (Cité page 43.)
- J.Pierre Cocquerez et Sylvie Philipp. *Analyse d'images : filtrage et segmentation*. Paris, masson édition, 1995. (Cité pages 10, 11 et 18.)
- L. D. Cohen et I. Cohen. Finite element methods for active contour models and ballons for 2D and 3D images. *IEEE Trans on PAMI*, 15(11) :1131–1147, 1993. (Cité page 11.)
- T.M. Cover et J.A. Thomas. *Elements of Information Theory*. New York, NY, wiley édition, 1991. (Cité page 53.)
- R. Cutler et L. S. Davis. Robust real-time periodic motion detection, analysis and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) : 781796, 2000. (Cité page 39.)
- Z. Dengsheng et L. Guojun. A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval. *Journal of Visual Communication and Image Representation*, 14(1) :41–60, March 2003. (Cité page 17.)
- Rachid Deriche. Using canny's criteria to derive recursively implemented optimal edge detector. *Int. J. Computer Vision*, 1(2) :167–187, 1987. (Cité page 10.)
- J. M. B. Dias et J. M. N. Leitao. Wall position and thickness estimation from sequences of echocardiographic images. *IEEE Tran on Medical Imaging*, 15(1) :25–38, 1996. (Cité page 11.)
- R. Dubes et A. Jain. Random field models in image analysis. *J. Appl. Stat.*, 16 : 131–164, 1989. (Cité page 13.)
- H. Elzein, S. Lakshmanan, et P. Watta. A motion and shape-based pedestrian detection algorithm. *IEEE Intelligent Vehicles Symposium*, page 500504, 2003. (Cité page 39.)
- L. Florack, A. Salden, et B. ter Haar Romeny. Nonlinear scale-space, image and vision computing. *Kluwer Academic Publishers*, 13(4) :279–294, 1995. (Cité page 14.)
- L. M. J. Florack. Image structure. *Kluwer Academic Publishers*, 13(4) :279–294, 1997. (Cité page 14.)
- P. Fua et Y. G. Leclerc. Model driven edge detection. *Machine Vision and Applications*, 3(1) :45–56, 1988. (Cité page 11.)
- F. Fuchs et H. Le Men. Detecting planar patches in urban scenes. Dans *Proceedings of SPIE, Visual Information Processing VIII, Orlando, USA*, volume 3716, pages 167–176, apr 1999. (Cité page 22.)
- Frank Fuchs. *Contribution à la reconstruction du bâti en milieu urbain, à l'aide d'images aériennes stéréoscopiques à grande échelle. Etude d'une approche structurelle*. PhD thesis, Université Paris 5 - René Descartes, 2001. (Cité page 23.)
- A. Gagalowicz et O. Monga. A new approach for image segmentation. Dans *In 8th ICPR*, 1986. (Cité page 19.)

- S. Geisser. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70 :320–328, 1975. (Cité page 57.)
- D. Geman. Random fields and inverse problems in imaging. In *Lecture notes in mathematics, Ecole d'été de probabilités de Saint Flour XVIII-1988* (P.L. Hennequin Ed.). Berlin : Springer-Verlag, 6(1) :117–193, 1990. (Cité pages 11 et 13.)
- S. Geman et D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984. (Cité pages 13 et 29.)
- G. Giraudon. Chaînage efficace de contours. *Rapport technique 605*, INRIA, 1987. (Cité page 10.)
- F. Guichard et J. M. Morel. *Image iterating smoothing and P.D.E.'s*. Cours E.N.S Cachan, 2002. (Cité page 16.)
- Laurent Guigues. *Modèles Multi-échelles pour la Segmentation d'Images (Multi-Scale Models for Image Segmentation)*. PhD thesis, Université de Cergy-Pontoise, France, Dec. 2003. (Cité pages 1, 26, 30, 36, 67 et 73.)
- Laurent Guigues, Jean Pierre Cocquerez, et Hervé Le Men. Scale-Sets Image Analysis. *International Journal of Computer Vision*, 68(1), April 2006. (Cité pages 1, 23, 25 et 29.)
- E. Haber et J. Modersitzki. A scale space method for volume preserving image registration. Dans *Scale Space and PDE Methods in Computer Vision. 5th International Conference, Scale-Space 2005. Proceedings (Lecture Notes in Computer Science)*, volume 3459, pages 561–72, 2005. (Cité page 17.)
- R. M. Haralick. Digital step edges from zero-crossings of second directional derivative. *IEEE Trans. on PAMI*, 6(1) :58–68, 1984. (Cité page 10.)
- K. Haris, S. N. Estradiadis, N. Maglaveras, et A. K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. on Image Processing*, 7(12) :1684–1699, 1998. (Cité pages 20 et 29.)
- S. Henn et K. Witsch. A variational image registration approach based on curvature scale space. Dans *Scale Space and PDE Methods in Computer Vision. 5th International Conference, Scale-Space 2005. Proceedings (Lecture Notes in Computer Science)*, volume 3459, pages 143–54, 2005. (Cité page 17.)
- A. Hernandez. *Fusion d'images radiographique et ultrasonores, visualisation stéréoscopique : Application aux pathologies du sein*. PhD thesis, Institut National des Sciences Appliquées de Lyon, France, 1995. (Cité page 12.)
- S. L. Horowitz et T. Pavlidis. Picture segmentation by a tree traversal algorithm. *J. of the Assoc. for Comp. Mach.*, 23(2) :36–388, 1976. (Cité pages 18, 19 et 21.)
- M. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, IT-8 :179–187, 1962. (Cité page 57.)
- M. F. Hueckel. An operator which locates edges in digitized pictures. *J. Asso. Comput. Mach*, 18(1) :113–125, 1971. (Cité page 10.)

- E. Izquierdo. Linear and nonlinear scale-spaces for video indexing and retrieval. Dans *IEE Seminar on Time-Scale and Time-Frequency Analysis and Applications* (Ref. No.00/019), pages 19/1–5, 2000. (Cité page 17.)
- Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, Englewood Cliffs édition, 1989. (Cité page 19.)
- K. Jang Ben. Gaussian and morphological scale space for shape analysis of medical images. Dans *IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 2, pages 1327–1331, 1994. (Cité page 17.)
- D. H. Johnson et S. Sinanovic. Symmetrizing the Kullback-Leibler distance. *IEEE Transactions on Information Theory*, 2001. (Cité page 53.)
- J.-M. Jolion et A. Montanvert. The adaptive pyramid, a framework for 2D image analysis. *CVGIP : Image Understanding*, 55(3) :339–348, 1982. (Cité page 21.)
- T. Kanungo, B. Dom, D. Niblack, W. Steele, et J. Sheinvald. MDL-based multi-band image segmentation using a fast region merging scheme. *Technical Report RJ 9960 (87919)*, IBM Research Division, 1995. (Cité pages 13 et 22.)
- M. Kass, A. Witkin, et D. Terzopoulos. Active contour models. *Int. J. Comp. Vision*, 1 :321–331, 1988. (Cité page 11.)
- A. Khashman et K.M. Curtis. Scale space analysis applied to medical image processing. Dans *Proceedings of the Fourth IEEE International Conference on Electronics, Circuits and Systems*, volume 1, pages 452–5, 1997. (Cité page 17.)
- 29
- Wolfgang Köhler. *Gestalt Psychology*. New York, Liveright édition, 1929. Trad. Franç. (1964) Psychologie de la forme. Paris : Gallimard. (Cité page 8.)
- I. Kitahara et Y. Ohta. Recording multiple videos and 3D video display in a large-scale space. Dans *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems MFI2003* (Cat. No.03TH8702), pages 137–40, 2003. (Cité page 17.)
- J. Kittler, J. Illingworth, et J. Foglein. Threshold selection based on a simple image statistic. *Computer Vision Graphics and Image Processing*, 30(2) :125–147, 1985. (Cité page 12.)
- J. Koenderink. The structure of images. *Biol. Cybern.*, 50 :363–370, 1984. (Cité pages 2 et 14.)
- G. Koepfle, C. Lopez, et J. M. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM Journal on Numerical Analysis*, 31(1) :282–299, 1994. URL [citeseer.nj.nec.com/koepfle94multiscale.html](http://citeseer.nj.nec.com/koepfle94multiscale.html). (Cité pages 22 et 23.)
- Kurt Koffka. *Principles of Gestalt Psychology*. New York, Harcourt Brace édition, 1935. (Cité page 8.)
- S. Kullback et R.A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22 : 79–86, 1951. (Cité page 53.)

- Guigues Laurent. *Modèles Multi-échelles pour la Segmentation d'Images (Multi-Scale Models for Image Segmentation)*. PhD thesis, Université de Cergy-Pontoise, France, Dec. 2003. (Cité page 3.)
- K.I. Laws. *Textured image segmentation*. PhD thesis, Dept. Electrical Engineering, University of Southern California, January 1980b. (Cité page 57.)
- Y. G. Leclerc. Simplicity of Description as the Basis for Visual Interpretation. Rapport technique, AI Center, SRI International, 1989a. (Cité page 13.)
- Yvan G. Leclerc. Constructing Simple Stable Descriptions for Image Partitioning. *International Journal of Computer Vision*, 3(1) :73–102, 1989b. (Cité pages 29 et 32.)
- T. C. M. Lee. Segmenting images corrupted by correlated noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 :481–492, 1998. (Cité page 22.)
- T. C. M. Lee. 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, 2000. (Cité pages 13 et 22.)
- Lee Lillian. On the effectiveness of the skew divergence for statistical language analysis. *Artificial Intelligence and Statistics*, pages 65–72, 2001. (Cité page 53.)
- T. Lindeberg. *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*. PhD thesis, Department of Numerical Analysis and Computer Science, KTH, 1991. (Cité page 17.)
- T. Lindeberg. Scale-space theory in computer vision. Kluwer Academic Publishers, 1994. (Cité page 14.)
- T. Lindeberg. Edge detection and ridge detection with automatic scale selection. Dans *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pages 465–470, June 1996. (Cité page 17.)
- T. Lindeberg. Feature detection with automatic scale selection. *Int. J. of Computer Vision*, 30(2), 1998. (Cité pages 16 et 17.)
- T. Lindeberg et J.-O Eklundh. Scale-space primal sketch : Construction and experiments. *Image Vision Comp.*, 10 :3–18, 1992. (Cité page 17.)
- T. Lindeberg et J.-O Eklundh. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2) :117–154, 1998. (Cité page 17.)
- T. Lindeberg et P. E. Roland. Analysis of brain activation patterns using a 3-D scale-space primal sketch. Rapport Technique 3, Human Brain Mapping, 1999. (Cité page 17.)
- C. Lorenz, I.-C. Carlsen, T.M. Buzug, C. Fassnacht, et J. Weese. A multi-scale line filter with automatic scale selection based on the hessian matrix for medical image segmentation. Dans *Scale-Space Theory in Computer Vision. First International Conference, Scale-Space '97. Proceedings*, pages 152–63, 1997. (Cité page 17.)
- J. Marroquin, S. Mitter, et T. Poggio. Probabilistic solution of illposed problems in computational vision. *J. Amer. Stat. Assoc.*, 82 :76–89, March 1987. (Cité page 13.)

- A. Martelli. Edge detection using heuristic search methods. *Comp. Graph. and Image Proc*, 1 :169182, 1972. (Cité page 10.)
- P. Meer. Stochastic image pyramids. *Computer Vision, Graphics and Image Processing*, 45 :269–294, 1989. (Cité page 21.)
- F. Meyer. Graph based morphological segmentation. Dans *In Proc. 2nd IAPR TC-15 workshop on Graph based Representations*, 1999a. (Cité page 23.)
- F. Meyer. Morphological multiscale and interactive segmentation. Dans *In Proc. of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, 1999b. (Cité page 23.)
- F. Meyer. An overview of morphological segmentation. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 15(7) :1089–1118, 2001. (Cité page 23.)
- F. Mokhtarian, S. Abbasi, et J. Kittler. Indexing an image database by shape content using curvature scale space. Dans *IEE Colloquium on Intelligent Image Databases (Ref. No.1996/119)*, pages 4/1–6, 1996. (Cité page 17.)
- U. Montanari. On the optimal detection of curves in noisy images. *Com. ACM*, 14 :335345, 1971. (Cité page 10.)
- A. Montanvert, P. Meer, et A. Rosenfeld. Hierarchical image analysis using irregular tesselations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13 (4) :307–316, 1991. (Cité page 21.)
- D. Mumford et J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. on Pure and Applied Math.*, 17(4) :577–685, 1989. (Cité pages 2, 13, 22 et 29.)
- R. Myung-Cheol, B. Christmas, J. Kittler, et L. Seong-Whan. Gesture spotting in low-quality video with features based on curvature scale space. Dans *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 375–80, 2006. (Cité page 17.)
- D. Navneet et Triggs B. Histograms of oriented gradients for human detection. In *Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, International Conference on Computer Vision and Pattern Recognition*, 2 :886–893, June 2005. (Cité page 39.)
- K. Nevatia, R. et Babu. Linear feature extraction and description. *Comp. Graphics and Image Proc*, 13(3) :257269, 1980. (Cité page 10.)
- S. Ogawa, K. Itoh, K. Omoto, X. Cheng, A. Ohya, et I. A. Akiyama. Three dimensional ultrasonic imaging for diagnosis of breast tumor. Dans *Proc. IEEE Ultrasonics Symposium*, October 1998. (Cité page 12.)
- S. Osher et R. Fedkiw. Level sets methods : An overview and some recent results. *J. Comput. Physics*, 169 :463–502, 2001. (Cité page 11.)
- S. Osher et J. Sethian. Front propagating with curvature dependant speed : Algorithm based on the Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79 :12–49, 1988. (Cité page 11.)
- J. Paillart. Espace visuel et programmation motrice. *Cahiers de Psychologie*, 19 :171–180, 1976. (Cité page 8.)

- T. N. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Trans. on Signal Processing*, 40(4) :901–914, 1992. (Cité page 12.)
- N. Paragios. *Geodesic Active Regions and Level Sets Methods*. PhD thesis, Ph. D. thesis, Université Nice-Sophia Antipolis., 2000. (Cité page 11.)
- N. Paragios. Level set methods and visual grouping. in Mathematical Models of Computer Vision : The Handbook (Paragios, Chen and Faugeras). Springer Verlag, Boston, MA, 2005. (Cité page 11.)
- T. Pavlidis et Y. Liow. Integrating region growing and edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 12 :225–233, March 1990. (Cité page 13.)
- P. Perona et J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE trans. on PAMI*, 12(1) :629–639, 1990. (Cité pages 14 et 15.)
- M. Pierrot Deseilligny, G. Stamon, et C. Y. Suen. Veinerization : A new shape description for flexible skeletonization. *IEEE Trans. on PAMI*, 20(5) :505–521, 1998. (Cité page 178.)
- C. Revol et M. Jourlin. A new minimum variance region growing algorithm for image segmentation. *Pattern Recognition Letters*, 18 :249–258, 1997. (Cité page 12.)
- L. Roberts. Machine perception of 3-D solids. *Optical and Electro-optical Information Processing*, MIT Press, 1965. (Cité page 10.)
- A. Rosenfeld et A. C. Kak. *Digital Picture Processing*, volume I and II, 2nd edn. San Diego, academic press édition, 1982. (Cité page 19.)
- M. Rousson et N. Paragios. Shape priors for level set representations. In *European Conference in Computer Vision (ECCV)*, II(3) :78–93, 2002. (Cité page 43.)
- M. Rousson et N. Paragios. Prior knowledge, level set representations and visual grouping. *International Journal of Computer Vision*, 76(3) :231–243, 2007. (Cité page 43.)
- P. Salembier, A. Oliveras, et L. Garrido. Anti-extensive Connected Operators for Image and Sequence Processing. *IEEE Trans. on Image Processing*, 7(4) :555–570, apr 1998. (Cité page 23.)
- P. Salembier et M. Pardàs. Hierarchical Morphological Segmentation for Image Coding. *IEEE Trans. on Image Processing*, 38(3) :359–386, 1994. (Cité page 23.)
- P. Salembier et J. Serra. Flat zones filtering, connected operators, and filters by reconstruction. *IEEE trans. on Image Processing*, 4(8) :1153–1160, 1995. (Cité page 23.)
- Jean Serra et Philippe Salembier. Connected operators and pyramids. Dans *SPIE conf. on Image Algebra and Mathematical Morphology, San Diego, CA*, volume 2030, pages 65–76, 1993. (Cité page 23.)
- J. A. Sethian. *Level-set methods*. Cambridge university press édition, 1996. (Cité page 11.)
- J. Shen et S. Castan. An optimal linear operator for edge detection. In *Proceeding CVPR, Miami*, 1986. (Cité page 10.)

- J. Shi et J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :888–905, 2000. (Cité pages 12 et 68.)
- J. Sporring, C.I. Colios, et P. E. Trahanias. Generalized scale-selection. Dans *International Conference on Image Processing (ICIP 2000)*, pages 1535–1538, 2000. (Cité page 17.)
- F. Taillandier, L. Guigues, et R. Deriche. A framework for constrained multi-scale range image segmentation. Dans *International Conference on Image Processing (ICIP 2003)*, pages 1535–1538, septembre 2003. (Cité page 36.)
- T. Tamaki, T. Yamamura, et N. Ohnishi. Image segmentation and object extraction based on geometric features of regions. Dans *Proc. of VICIP'99 (Visual Communications and Image Processing) : Electronic Imaging '99, IS T and SPIE*, volume II, 1999. (Cité page 21.)
- V. Torre et T. Poggio. On edge detection. *IEEE, Pattern Analysis and Machine Intelligence*, 8(2) :147–153, 1997. (Cité page 10.)
- R. Trias-Sanz. An edge-based method for registering a graph onto an image with application to cadastre registration. Dans *Proc. of the 2004 Advanced Concepts for Intelligent Vision Systems conference (ACIVS 2004)*, pages 333–340, septembre 2004. (Cité page 35.)
- R. Trias-Sanz et M. Pierrot-Deseilligny. A region-based method for graph to image registration with application to cadastre data. Dans *Proc. of the 2004 International Conference on Image Processing (ICIP 2004)*, pages 333–340, octobre 2004. (Cité page 35.)
- Undeman et T. Lindeberg. Fully automatic segmentation of mri brain images using probabilistic anisotropic diffusion and multi-scale watersheds. Dans *Proc. Scale-Space'03. Springer LNCS*, volume 2695, pages 641–656, June 2002. (Cité page 17.)
- B. C. Vemuri, S. Rahman, et J. Li. Multiresolution adaptive k-means algorithm for segmentation of brain MRI. Dans *Proc. Intl. Comp. Sci. Conf. on Image Analysis and Computer Graphics*, December 1995. (Cité page 12.)
- L. Vincent. *Algorithmes morphologiques à base de files d'attente et de lacets : Extension aux graphes*. PhD thesis, Ph. D. thesis, Ecole des mines de Paris, France, 1990. (Cité page 178.)
- P. Viola, M. J. Jones, et D. Snow. Detecting pedestrians using patterns of motion and appearance. *IEEE International Conference on Computer Vision*, 2 :734741, 2003. (Cité page 39.)
- J. Ward. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.*, 58 :236–244, 1963. (Cité pages 20 et 58.)
- J. Weickert. Anisotropic diffusion in image processing. Teubner, Stuttgart, 1998. (Cité page 16.)
- Max Wertheimer. Gestalt theory. *Social Search*, 11 :78–99, 1944. (Cité page 8.)
- J. Weszka. A survey of threshold selection techniques. *Computer Vision Graphics and Image Processing*, 7(2) :259–265, 1978. (Cité page 12.)

- A. Witkin. Scale-space filtering. Dans *Proc. of 8th Int. Joint Conf. on Artificial Intell., Karlsruhe, West Germany*, pages 1019–1021, 1983. (Cité pages 2 et 14.)
- A. L. Yuille et T. A. Poggio. Scaling theorems for zero-crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 :15–25, 1986. (Cité page 14.)
- F. Zanoguera, B. Marcotegui, et F. Meyer. A toolbox for interactive segmentation based on nested partitions. Dans *In proc. of ICIP'99*, 1999. (Cité page 23.)
- S. Zhu et A. Yuille. Region competition : Unifying snakes, region growing and Bayes/MDL for multiband image segmentation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 18 :884–900, 1996. (Cité page 13.)