

N° d'ordre : 1548

École Doctorale Mathématiques, Sciences de
l'Information et de l'Ingénieur

UdS – INSA – ENGEES

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg
Discipline : Informatique

par

M. Camille KURTZ

Approche collaborative segmentation–classification pour l'analyse descendante d'images multirésolutions

Soutenue publiquement le 11 septembre 2012

Membres du jury

- Directeur de thèse* – M. Pierre GANÇARSKI, Professeur, Université de Strasbourg
- Rapporteur* – M. Jean-Marc OGIER, Professeur, Université de La Rochelle
- Rapporteur* – M. Pierre SOILLE, Principal Researcher, JRC, European Commission, Ispra (Italie)
- Examineur* – M. Antoine CORNUÉJOLS, Professeur, AgroParisTech (INA–PG)
- Examineur* – M. Nicolas PASSAT, Maître de Conférences, Université de Strasbourg
- Examineur* – Mme Anne PUISSANT, Maître de Conférences, Université de Strasbourg

Remerciements

En premier lieu, je tiens à remercier vivement Pierre Gançarski qui a dirigé cette thèse de manière exemplaire et qui a su me guider et m'orienter dans mon travail de recherche. Il a su également se battre pour me fournir des conditions de travail plus que satisfaisantes aussi bien au niveau matériel, qu'au niveau intellectuel et humoristique! Je le considère ainsi comme mon « papa » de la recherche. Durant toute la période de cette thèse, la porte de son bureau est toujours restée ouverte, laissant le chemin libre à toutes mes questions durant ces quatre années passées à travailler à ses côtés.

Je remercie également très sincèrement Nicolas Passat (mon autre « papa » de la recherche) qui m'a accompagné de façon exceptionnelle durant ces quelques années. Sa porte m'a elle aussi été toujours ouverte! Sans lui, je n'aurais sans doute pas pu mener ces travaux de recherche avec autant de rigueur et de pragmatisme! Je tiens à le remercier particulièrement pour sa disponibilité, son investissement personnel et le courage de son estomac pour avoir accepté d'ingurgiter notre café sordide pendant ces quatre années! Il m'a offert la motivation nécessaire pour surmonter les jours de doute, et je lui en suis extrêmement reconnaissant.

Un grand merci également à Anne Puissant (« l'experte » de cette thèse!) pour ses contributions importantes dans l'encadrement de cette thèse, et son expertise géographique en analyse d'environnements urbains. Ses conseils et ses remarques m'ont permis de découvrir les besoins opérationnels des utilisateurs « réels » d'images satellites.

Merci à Jean-Marc Ogier, Pierre Soille et à Antoine Cornuéjols pour l'intérêt qu'ils ont porté à ces recherches et pour avoir accepté d'évaluer ces travaux. La relecture d'un manuscrit de deux cents pages durant une période estivale nécessite des sacrifices que je n'oublierai pas.

Merci également à Cédric Wemmert pour m'avoir légué son savoir relatif à la multirésolution. Merci pour sa sympathie, pour son entrain à la recherche et pour l'aide qu'il a pu m'apporter. En particulier, il a su me défendre admirablement pour l'obtention de mon post-doc aux USA et je lui en suis particulièrement reconnaissant!

Merci à François Petitjean pour toutes ces années passées ensemble. Notre vision très similaire des choses m'a toujours impressionné. Depuis la première année universitaire, le temps est passé vite! Merci à ces grandes conversations philosophiques, ces après-midi de programmation, ces midis de détente sur JURSE. Merci aussi à lui pour ses nombreux conseils qui m'ont souvent sorti d'affaire quand je coïnçais sur un problème! Nos routes vont maintenant se séparer mais j'espère que nous garderons le contact ainsi que de bons souvenirs dernière nous!

Merci à Sébastien Derivaux, Germain Forestier, Jonathan Weber pour la qualité de la vie commune passée avec eux au bureau. Merci à eux pour leurs conseils avisés (et expérimentés) et pour m'avoir fait découvrir la réalité des choses dans le monde, parfois obscur, de la recherche! *Muskatnuss!*

Merci à Bruno Belarte pour cette dernière année de fraîcheur scientifique. L'enthousiasme d'un jeune chercheur nous reconforte toujours dans notre idée qu'il existe toujours, quelque part, un doctorant n'ayant pas encore soutenu sa thèse et à qui il reste à parcourir un bout de chemin long et difficile!

Merci aux autres doctorants, « du fond du couloir », Ogier Maître, Matthieu Herrmann et Emmanuel Benard pour les bons moments que nous avons partagé et nos discussions souvent d'un niveau scientifique *next to null!*

Je tiens également à remercier ma famille qui a financé mes études, m'a poussé vers l'avant et m'a toujours soutenu au cours de mes années de formation universitaire. Merci à eux de m'avoir appris à toujours persévérer pour aller jusqu'au bout des choses!

Enfin, le plus grand « Merci » est donné à ma compagne Marie pour m'avoir supporté, soutenu et encouragé durant ces longues années nécessaires à l'accomplissement de ce travail. Sa présence m'a été plus que bénéfique et a contribué grandement à l'accomplissement de ces travaux dans d'excellentes conditions humaines! Merci encore pour son incroyable patience depuis la fin de nos diplômes respectifs de Master. J'espère maintenant que nous nous dirigerons vers de nouvelles aventures outre-Alsace!

Table des matières

Introduction	3
Contexte des travaux	3
Problématique	3
Axes de recherche	5
Structure du document	6
Partie I	
Approches basées régions pour l'analyse d'images : segmenter pour classifier	7
1 La fouille de données pour l'analyse d'images optiques	9
1.1 Analyse de masses de données image	9
1.2 Analyse automatique d'images	11
1.3 Les approches basées pixels pour l'analyse d'images	15
2 Les approches basées régions pour l'analyse d'images	19
2.1 Principe des approches basées régions	19
2.2 L'apport de la segmentation d'images pour la classification	26
2.3 Limites des approches basées régions	28
3 L'intérêt des images multisources	39
3.1 Images multisources	39
3.2 Analyser des images multisources	43
3.3 Approche collaborative basée régions pour l'analyse d'images multirésolutions	46
Partie II	
Processus collaboratif pour l'analyse descendante d'images multirésolutions	49
4 Analyse descendante d'images multirésolutions	51
4.1 Motivations	51
4.2 Proposition	53
4.3 Schéma méthodologique	54
5 Segmentation hiérarchique d'images basée sur des exemples	57
5.1 Segmentation à base d'arbres binaires de partitions	57
5.2 Le problème de l'échelle	62
5.3 Apprentissage et reproduction d'une coupe	64
5.4 Bilan scientifique	68
6 Classification de segments multirésolutions	71
6.1 Caractérisation multirésolution des segments	71
6.2 Histogramme de composition	73
6.3 Classification multirésolution	75
6.4 Bilan scientifique	78
7 PCSC, un processus d'analyse collaboratif segmentation–classification	79
7.1 Schéma méthodologique de PCSC	79
7.2 Processus collaboratif segmentation–classification à travers la résolution	80
7.3 Complexité algorithmique de PCSC	82
7.4 Bilan scientifique	84

8	Validations expérimentales	85
8.1	Protocole de validation	85
8.2	Matériel	87
8.3	Expérimentations, résultats et analyses	90
8.4	Bilan scientifique	107
Partie III		
Vers un processus itératif guidé par des connaissances		109
9	Connaissances et remise en cause des résultats	111
9.1	Limites de PCSC	111
9.2	Différents types de connaissances	114
9.3	Remise en cause des résultats	117
10	HSBD, une distance entre histogrammes basée sur les connaissances	121
10.1	Mesurer la distance entre des attributs multirésolutions	121
10.2	Distances entre histogrammes	124
10.3	Schéma méthodologique de HSBD	129
10.4	Bilan scientifique	135
11	C-PCSC, un processus de post-classification guidé par des connaissances	137
11.1	Remise en cause des résultats de classification	137
11.2	Schéma méthodologique de C-PCSC	141
11.3	Vers un enrichissement itératif des résultats	143
11.4	Bilan scientifique	145
12	Validations expérimentales	147
12.1	Protocole de validation	147
12.2	Matériel	148
12.3	Expérimentations, résultats et analyses	149
12.4	Bilan scientifique	163
Conclusion		167
	Bilan	167
	Contributions	167
	Perspectives de recherche	168
Bibliographie		170
Annexes		183
A	Publications liées aux travaux	183
B	Matériel	185
C	Classification de séries temporelles d'images satellites	189
D	Développements logiciels	193

Liste des tableaux

8.1	Typologies et niveaux d'analyse.	86
8.2	Synthèse des jeux de données utilisés pour les expériences.	90
8.3	Classes thématiques.	95
8.4	Indices de qualité.	96
8.5	Résultats d'évaluation obtenus pour les jeux de données ELSAU, ILLKIRCH, HAUTEPIERRE	98
8.6	Évaluation globale des résultats de l'expérience d'adaptation au domaine.	106
8.7	Temps de calcul et consommation mémoire.	108
8.8	Comparaison des temps de calcul et des consommations mémoire.	108
10.1	Histogrammes modélisant les compositions des trois paniers de courses C_1, C_2, C_3	123
10.2	Matrice de dissimilarité \mathcal{M}^{dis} associée aux instances du concept FRUIT.	125
10.3	Histogrammes modélisant les compositions des trois paniers de courses C_1, C_2, C_3	128
12.1	Synthèse des jeux de données utilisés pour les expériences.	149
12.2	Matrice de dissimilarité \mathcal{M}^{dis} associée aux instances du concept OBJET URBAIN.	150
12.3	Résultats d'évaluation globaux.	153
12.4	Matrice de dissimilarité \mathcal{M}^{dis} associée aux instances du concept BLOCS URBAIN.	155
12.5	Résultats d'évaluation obtenus pour les trois jeux de données.	157

Table des figures

1.1	Image thématique résultant de la classification d'une image satellite.	12
1.2	Schéma d'une approche basée pixels pour l'analyse d'images.	16
2.1	Schéma d'une approche basée régions pour l'analyse d'images.	20
2.2	Illustration du résultat d'un algorithme de segmentation.	21
2.3	Illustration du problème de segmentation.	22
2.4	Illustration des phénomènes de sur-segmentation et de sous-segmentation.	23
2.5	Processus de segmentation guidée par les connaissances.	25
2.6	Exemple de hiérarchie de six concepts urbains.	27
2.7	Illustration d'une erreur de segmentation.	29
2.8	Bloc urbain délimité en rouge sur une image THRS (60 cm)	31
2.9	Résultat d'analyse multiniveau.	32
2.10	Relations spatiales du système RCC8.	34
2.11	Exemple de segmentation hiérarchique.	35
2.12	Exemple de segmentation hiérarchique basée régions.	36
3.1	Trois cas de données multisources.	41
3.2	Illustration d'une fonction de correspondance.	42
3.3	Schéma d'une approche par fusion pour l'analyse d'images multisources.	43
3.4	Schéma d'une approche collaborative pour l'analyse d'images multisources.	45
4.1	Schéma du processus collaboratif basé régions pour l'analyse d'images multirésolutions.	55
5.1	Calcul d'une carte d'élongation.	60
5.2	Un exemple de BPT.	61
5.3	Illustration de l'extraction interactive d'une coupe à partir d'un BPT.	62
5.4	Approche de segmentation interactive, basée sur un exemple.	63
5.5	Approche de segmentation basée sur des exemples : apprentissage.	66
5.6	Approche de segmentation basée sur des exemples : reproduction automatique.	67
6.1	Illustration d'un histogramme de composition.	74
6.2	Méthodologie de classification multirésolution.	75
6.3	Étiquetage des clusters obtenus.	76
7.1	Illustration du processus PCSC appliqué au traitement d'images multirésolution.	81
8.1	Illustration de la zone urbaine de Strasbourg (France).	88
8.2	Présentation du jeu de données ELSAU (6 576 m × 2 793 m).	89
8.3	Cartes d'élongation.	91
8.4	Représentation des fonctions α et $(1 - \alpha)$	92
8.5	Cartes de vérité terrain associées aux images du jeu de données ELSAU	94
8.6	Résultats des différentes études d'impact.	97
8.7	Segmentations obtenues pour le jeu de données ELSAU (6 576 m × 2 793 m).	99
8.8	Résultats obtenus pour le jeu de données ELSAU (6 576 m × 2 793 m).	100
8.9	Résultats obtenus pour le jeu de données ILLKIRCH (6 576 m × 2 803 m).	101
8.10	Résultats obtenus pour le jeu de données HAUTEPIERRE (3 293 m × 2 246 m).	102
8.11	Illustration d'une partie des résultats de segmentation.	104
8.12	Impact du choix de l'exemple de coupe parmi $p = 20$ possibilités.	105
8.13	Illustration des résultats de l'expérience d'adaptation au domaine.	106

9.1	Extrait d'une segmentation dont les frontières sont crénelées.	112
9.2	Connaissances monorésolutions.	115
9.3	Connaissances multirésolutions liées aux images.	116
10.1	Deux types d'histogrammes modélisant la composition d'un panier de courses.	122
10.2	Dendrogramme associé à la matrice de dissimilarité.	130
10.3	Schéma méthodologique du calcul de la distance HSD.	131
10.4	Représentation graphique des fonctions \mathbf{d}^k et \mathbf{d}_{inter}^k	133
11.1	Connaissances multirésolutions liées aux images.	139
11.2	Illustration du processus de post-classification monorésolution.	140
11.3	Illustration du principe itératif ascendant de remise en cause des résultats.	144
12.1	Dendrogramme associé à la matrice de dissimilarité.	151
12.2	Dendrogrammes associés à la matrice de dissimilarité.	152
12.3	Résultats d'évaluation locaux.	154
12.4	Dendrogramme associé à la matrice de dissimilarité.	156
12.5	Résultats de post-classification pour le jeu de données ELSAU (6 576 m × 2 793 m).	158
12.6	Résultats de post-classification pour le jeu de données ILLKIRCH (6 576 m × 2 803 m).	159
12.7	Histogrammes de composition basés régions (liés au concept sémantique des objets urbains).	160
12.8	Histogrammes de composition basés régions (liés au concept sémantique des objets urbains).	161
12.9	Cartes des conflits.	162
12.10	Impact des perturbations appliquées sur les valeurs de proximité sémantique.	163
B.1	Présentation du jeu de données ILLKIRCH (6 576 m × 2 803 m).	185
B.2	Présentation du jeu de données HAUTEPIERRE (3 293 m × 2 246 m).	186
B.3	Cartes de vérité terrain associées aux images du jeu de données ILLKIRCH	187
B.4	Cartes de vérité terrain associées aux images du jeu de données HAUTEPIERRE	188
C.1	Illustration de l'approche proposée.	190
C.2	Illustration d'une partie du jeu de données considéré et d'un résultat de classification obtenu.	192
D.1	Illustration de l'interface graphique du logiciel MUSTIC.	195
D.2	Illustration de l'exportation des résultats de classification dans le logiciel © Google Earth.	195

Notations

Nous présentons ci-dessous quelques notions et notations, nécessaires à la compréhension de ce manuscrit. Le lecteur pourra s'y référer au cours de la lecture de ce mémoire.

Notations préliminaires

Ensemble et partition Soit X un ensemble fini. L'ensemble $\{Y \mid Y \subseteq X\}$ de tous les sous-ensembles de X est noté 2^X . Le cardinal de X est noté $|X|$. Si un ensemble $\{X_i\}_{i=1}^t \in 2^X$ ($t \geq 1$) de sous-ensembles de X est une partition de X , on note alors $X = \sqcup_{i=1}^t X_i$ (ou $X = X_1 \sqcup X_2 \dots \sqcup X_t$).

Liste et vecteur Un n -uplet \mathbf{v} composé de n éléments a_i ($i \in \llbracket 1, n \rrbracket$) est noté $\mathbf{v} = (a_i)_{i=1}^n = (a_1, a_2, \dots, a_n)$. Une telle structure sera généralement utilisée pour représenter des listes ou encore des vecteurs d'éléments.

Fonction Une fonction \mathcal{F} de l'ensemble X vers l'ensemble Y est notée $\mathcal{F} : X \rightarrow Y$. Pour tout $Z \subseteq X$, l'image $\{\mathcal{F}(z) \mid z \in Z\}$ de Z par \mathcal{F} est notée $\mathcal{F}(Z)$. Pour tout $T \subseteq Y$, la pré-image $\{t \mid \mathcal{F}(t) \in T\}$ de T par \mathcal{F} est notée $\mathcal{F}^{-1}(T)$.

Intervalle Un intervalle sur \mathbb{R} , borné par $a, b \in \mathbb{R}$, sera noté $[a, b]$. Un intervalle sur \mathbb{Z} , borné par $a, b \in \mathbb{Z}$, sera noté $\llbracket a, b \rrbracket$.

Domaine de définition Soit $E = \llbracket 0, d_x - 1 \rrbracket \times \llbracket 0, d_y - 1 \rrbracket \subset \mathbb{Z}^2$. L'ensemble E correspond à la discrétisation de l'espace continu (*i.e.*, une partie de \mathbb{R}^2) qui sera visualisé dans les images. Par la suite on nommera E le domaine de définition d'une image. Un élément $\mathbf{x} = (x, y) \in E$ est appelé un pixel.

Image Soit $V_b = \llbracket 0, v_b - 1 \rrbracket \subset \mathbb{Z}$. L'ensemble V_b correspond à l'échantillonnage discret des intensités observées pour une bande spectrale particulière. Une image monovaluée (ou monospectrale) est une fonction $\mathcal{I}_b : E \rightarrow V_b$ qui à chaque point \mathbf{x} de la scène, associe une valeur d'intensité spectrale v dans exactement une bande spectrale

$$\begin{aligned} \mathcal{I}_b : \quad E &\rightarrow V_b \\ \mathbf{x} = (x, y) &\mapsto \mathcal{I}_b(\mathbf{x}) = v \end{aligned} \tag{1}$$

Soit $V = \prod_{b=1}^s V_b \subset \mathbb{Z}^s$ ($s \geq 1$). L'ensemble V correspond à l'échantillonnage discret des intensités observées pour s bandes spectrales particulières. Une image multivaluée (ou multispectrale) est une fonction $\mathcal{I} : E \rightarrow V$ qui à chaque point \mathbf{x} de la scène, associe un s -uplet de valeurs d'intensités spectrales $\mathbf{v} = \prod_{b=1}^s \mathcal{I}_b(\mathbf{x})$ dans les bandes spectrales considérées

$$\begin{aligned} \mathcal{I} : \quad E &\rightarrow V \\ \mathbf{x} = (x, y) &\mapsto \mathcal{I}(\mathbf{x}) = \mathbf{v} = \prod_{b=1}^s \mathcal{I}_b(\mathbf{x}) \end{aligned} \tag{2}$$

Résolution spectrale d'une image La résolution spectrale d'une image peut être caractérisée par le nombre de bandes spectrales, leurs largeurs, et leurs positions sur le spectre.

Résolution spatiale d'une image La résolution spatiale d'une image correspond à la capacité du capteur dont elle est issue à distinguer deux objets à une certaine résolution. En pratique, cette grandeur physique peut se voir comme la surface au sol représentée par chacun des pixels (*e.g.*, 1 pixel = 10 m \times 10 m). Par la suite, on notera τ_i la résolution spatiale liée a une image \mathcal{I}^i .

Histogramme Soit x une mesure, ou un attribut, pouvant prendre w valeurs dans l'ensemble $X = \{x_1, x_2, \dots, x_w\}$. Soit A un ensemble composé de n éléments. Chaque élément de A est associé à une valeur a par la mesure x . L'ensemble d'observation résultant de cette mesure est défini comme $A_x = \{a_1, a_2, \dots, a_n\}$ où $a_i \in X$.

L'histogramme de l'ensemble A_x , relativement à la mesure x effectuée sur A , noté $H(x, A)$ est une liste de w éléments comptant le nombre des occurrences des valeurs de x parmi les a_i . Par abus de langage, on utilisera par la suite la notation $H(A)$ au lieu de $H(x, A)$. L'histogramme $H(A)$ peut être défini comme $H(A) = (H_1(A), H_2(A), \dots, H_w(A))$ où $H_i(A)$, $i \in \llbracket 1, w \rrbracket$, représente le nombre d'éléments de A_x ayant la valeur x_i . Chacun des $H_i(A)$ peut être calculé comme

$$H_i(A) = \sum_{j=1}^n c_{ij} \quad \text{avec} \quad c_{ij} = \begin{cases} 1 & \text{si } a_j = x_i \\ 0 & \text{sinon} \end{cases} \quad (3)$$

Dans la littérature, les différents $H_i(A)$ sont généralement appelés les barres (ou encore les modalités) de l'histogramme $H(A)$. Si $P_i(A)$ représente la probabilité des échantillons relativement à la i -ème barre, alors $P_i(A) = H_i(A)/n$. Ainsi, $P(A)$ peut être représenté comme un histogramme normalisé.

Les w valeurs de la mesure x sont généralement nommées *niveaux de la mesure* quand elles sont utilisées dans $H(A)$ pour indexer les distributions des valeurs des échantillons.

Histogramme d'une image L'histogramme $H(\mathcal{I}, E)$ représentant la distribution des valeurs d'une image $\mathcal{I} : E \rightarrow V$, peut être construit par le biais de la fonction $\mathcal{H}_{\mathcal{I}} : V \rightarrow \mathbb{N}$ qui associe à chaque valeur $v \in V$ le nombre $\mathcal{H}_{\mathcal{I}}(v) = |\mathcal{I}^{-1}(\{v\})|$ de pixels de \mathcal{I} possédant la valeur v .

On considère ici v comme une mesure pouvant prendre w valeurs distinctes ($w = |V|$). L'histogramme $H(\mathcal{I}, E)$ peut alors être défini comme $H(\mathcal{I}, E) = (H_1(\mathcal{I}, E), H_2(\mathcal{I}, E), \dots, H_w(\mathcal{I}, E))$ où $H_i(\mathcal{I}, E) = \mathcal{H}_{\mathcal{I}}(i)$, $i \in \llbracket 1, w \rrbracket$.

L'histogramme $H(\mathcal{I}, X)$ associé au sous-ensemble $X \subseteq E$ peut être construit d'une manière similaire par le biais de la fonction $\mathcal{H}_{\mathcal{I}, X} : V \rightarrow \mathbb{N}$ qui associe à chaque valeur $v \in V$ le nombre $\mathcal{H}_{\mathcal{I}, X}(v) = |\mathcal{I}^{-1}(\{v\}) \cap X|$ (i.e., l'histogramme restreint de E à X).

Introduction

Introduction

« À la source de toute connaissance, il y a une idée, une pensée, puis l'expérience vient confirmer l'idée. » 

— Claude Bernard (1813–1878)

Contexte des travaux

Depuis la fin des années 1980, les images optiques issues de capteurs aéroportés sont de plus en plus accessibles par une vaste communauté d'utilisateurs. En particulier, différents systèmes satellitaires sont maintenant disponibles et produisent une quantité de données importante, utilisable pour l'observation de la Terre. En raison de cet important volume de données, les méthodes analytiques manuelles couramment employées par les experts en photo-interprétation, deviennent inadaptées et/ou sur-consommatrices de temps pour un traitement efficace de ces données. Il devient donc crucial d'automatiser ces méthodes par des procédés informatiques permettant de répondre aux besoins des utilisateurs, et capables de traiter cette quantité de données hétérogènes.

Dans ce contexte a débuté, au cours des années 2000, une étroite collaboration entre les laboratoires d'Informatique (LSIIT¹) et de Géographie Environnementale (LIVE²) de l'Université de Strasbourg. De nombreux travaux ont ainsi été menés, dans le but de développer des méthodes (semi-)automatiques permettant d'extraire des connaissances utiles (*e.g.*, classes de couverture des sols, éléments urbains) à partir d'images de télédétection. La plupart de ces méthodes sont basées sur des principes empruntés aux domaines de la fouille de données et du traitement d'images. La présente thèse s'inscrit dans la continuité de ces travaux.

Problématique

Jusqu'au début des années 1990, la majorité des méthodes développées dans le cadre de l'analyse d'images de télédétection étaient fondées sur le paradigme *basé pixels*, qui consiste à considérer directement les pixels des images comme les données à traiter. Dans la plupart de ces approches, l'objectif est de catégoriser dans des classes particulières, par le biais d'algorithmes de classification, des pixels présentant des propriétés radiométriques similaires dans l'espace des données. Les classes résultantes peuvent alors fournir des informations pertinentes aux utilisateurs, relatives par exemple à l'organisation, l'occupation et la nature de la couverture des sols. Dans le cadre de l'analyse d'environnements urbains, les méthodes basées pixels étaient principalement employées pour analyser les taches urbaines (*e.g.*, agglomérations, zones agricoles) et les grandes zones thématiques (*e.g.*, quartiers urbains, parcelles agricoles) à partir d'images à basses résolutions spatiales (BRS – Basse Résolution Spatiale : > 30 m, et MRS – Moyenne Résolution Spatiale : 30 à 5 m).

Cependant, l'apparition à la fin des années 1990, des images à hautes résolutions spatiales (HRS – Haute Résolution Spatiale : 5 à 1 m, et THRS – Très Haute Résolution Spatiale : < 1 m), a entraîné une augmentation importante de la précision des images (*e.g.*, satellite IKONOS lancé en 1999). Ce phénomène a eu pour conséquence de diminuer considérablement l'efficacité de ces méthodologies, classiquement employées en routines opérationnelles. En effet, de telles images permettent une analyse directe des objets géographiques

1. Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection – UMR CNRS/UdS 7005.

2. Laboratoire Image, Ville et Environnement – ERL CNRS 7230.

simples (*e.g.*, maisons, routes). De tels objets apparaissent dans ces images comme complexes car ils sont souvent composés d'ensembles de pixels hétérogènes (*i.e.*, ayant des propriétés radiométriques différentes dans l'espace des données). En conséquence, ces objets d'intérêt ne peuvent pas être extraits directement par des méthodes basées pixels reposant principalement sur des hypothèses d'homogénéité.

Pour permettre le traitement de ces données, de nouvelles méthodes d'analyse ont été proposées. Celles-ci sont communément appelées méthodes *basées régions*³ [Blaschke, 2010]. Le principe de ces méthodes est de délimiter des parties de l'image pouvant correspondre aux objets d'intérêt de la scène avant de les analyser (*e.g.*, de les classifier). Pour ce faire, l'image est décomposée en segments par un processus de partitionnement (*i.e.*, de segmentation) qui consiste à regrouper des pixels spatialement connexes en vue de constituer des agrégats de pixels homogènes. Ces segments sont alors caractérisés par des attributs de nature variable (*e.g.*, informations radiométriques, de texture ou de forme) puis sont catégorisés dans des classes particulières par des algorithmes de classification.

Parallèlement, il est souvent nécessaire d'extraire plusieurs niveaux de connaissance et d'information à partir des images. Ainsi, dans le domaine de la télédétection, et en particulier dans celui de la cartographie urbaine, les utilisateurs emploient les images pour analyser le territoire à plusieurs niveaux d'échelle, dans le but d'extraire différents niveaux d'objets d'intérêt (*e.g.*, quartiers urbains, blocs urbains, objets urbains individuels). De plus, les objets d'intérêt relatifs à ces différents niveaux d'analyse sont souvent structurés de façon hiérarchique. Par exemple, un quartier urbain se décompose en général en un ensemble de blocs urbains de divers types, eux-mêmes pouvant se décomposer en ensembles d'objets urbains simples, formant ainsi une hiérarchie conceptuelle [Puissant, 2003].

Les travaux menés par les communautés spécialisées en analyse d'images de télédétection ont démontré que les méthodes basées régions sont bien adaptées à l'extraction d'objets d'intérêt simples car elles supportent un certain niveau d'hétérogénéité dans ces objets. En revanche, elles ne sont pas applicables directement pour extraire des objets d'intérêt complexes de niveaux sémantiques plus élevés (*e.g.*, blocs ou quartiers urbains, parcs), car ces derniers apparaissent dans l'image comme trop hétérogènes et sont le résultat d'une certaine organisation spatiale des objets simples.

Pour apporter une solution à ces problèmes, des extensions aux approches basées régions ont été récemment proposées dans la littérature. Le principe de ces méthodes est de regrouper, d'une manière ascendante, des objets d'intérêt simples extraits des images (*e.g.*, maisons, routes) pour former de nouveaux types de structures d'intérêt (*e.g.*, blocs urbains, quartiers) permettant ainsi différents niveaux d'analyse. Cependant, devant la grande complexité des données à hautes résolutions spatiales (*i.e.*, nombre élevé de pixels, niveau de détail considérable), il est souvent difficile d'employer de telles méthodes, qui nécessitent des temps de calcul importants.

Par ailleurs, il est devenu courant de disposer d'ensembles de données multisources proposant des informations différentes, et potentiellement complémentaires sur la zone à étudier. De tels ensembles peuvent être composés d'images centrées sur la même scène géographique, ayant des caractéristiques hétérogènes, et potentiellement acquises par des satellites différents. Par exemple, l'utilisateur dispose de plus en plus fréquemment d'ensembles d'images à plusieurs résolutions spatiales (*i.e.*, données multirésolutions), offrant des vues différentes de la scène étudiée (*i.e.*, à plusieurs niveaux d'échelle), ce qui peut permettre de faciliter l'extraction des objets d'intérêt. Des travaux récents [Wemmert et al., 2009] ont ainsi montré que les approches multisources, et plus particulièrement les approches multirésolutions, sont prometteuses dans le cadre de l'analyse et de l'extraction d'objets complexes à partir d'images satellites. En effet, de tels objets peuvent apparaître comme plus homogènes dans des images à plus faibles résolutions. Cependant, la limite majeure des approches multisources réside dans le fait qu'elles reposent actuellement sur un paradigme basé pixels.

Le problème principal étudié au cours de cette thèse a donc porté sur l'extension des approches multisources au paradigme basé régions, dans l'optique de permettre l'extraction de plusieurs niveaux de connaissance et d'information à partir des images.

3. Le terme « orienté objets » a également été utilisé pour décrire ces méthodes mais a été progressivement abandonné par la communauté Informatique pour éviter toute confusion avec le concept homonyme utilisé pour les langages de programmation.

Axes de recherche

Aider l'utilisateur dans le processus d'extraction d'objets ne signifie pas remplacer ce dernier. Ses connaissances relatives aux objets étudiés et aux données utilisées dans ce processus seront toujours nécessaires pour guider un outil informatique dans l'accomplissement de cette tâche. De ce fait, nous avons étudié au cours de cette thèse de nouvelles approches basées régions, les plus automatisées possibles, permettant d'aider les utilisateurs dans le processus d'extraction de différents niveaux sémantiques d'objets complexes à partir d'images de télédétection. Durant cette période, nous nous sommes particulièrement intéressés à la définition d'approches d'analyse multirésolutions descendantes, s'inspirant des mécanismes de photo-interprétation couramment employés par les utilisateurs, experts dans ce domaine. Le principe est d'extraire, dans un premier temps, les grandes zones de la scène observée à partir d'images à résolutions grossières, puis de les décomposer récursivement pour en extraire des objets d'intérêt plus simples à partir d'images à des résolutions plus fines. Le principal avantage de ces approches est de pouvoir traiter plus aisément les grandes quantités de données actuellement disponibles.

Axe 1 : extraction d'objets d'intérêt complexes à partir d'images multirésolutions

Dans un premier axe de recherche, nous nous sommes ainsi intéressés à l'élaboration d'une méthodologie basée régions, capable de traiter des données multirésolutions afin d'en extraire des hiérarchies d'objets d'intérêt. Pour ce faire, nos recherches se sont orientées vers la définition de nouveaux mécanismes de segmentation et de classification issus des domaines de la fouille de données et du traitement d'images. Les objets d'intérêt à extraire étant structurés sous la forme de hiérarchies, nos travaux se sont naturellement tournés vers l'utilisation d'approches de segmentation hiérarchique fournissant des ensembles de partitions de la scène à différents niveaux de détail. Une fois les segments extraits des images, il devient alors possible de les catégoriser par le biais d'algorithmes de classification. Nos travaux, se sont aussi intéressés à la classification non-supervisée de ces segments et plus particulièrement à la caractérisation de ces données par l'intermédiaire d'attributs multirésolutions. Finalement, dans le but de faire collaborer ces deux approches au sein d'un processus unique, nos recherches ont porté sur la définition d'un processus global d'analyse d'images multirésolutions, descendant à travers la résolution spatiale. De manière plus générale, nous nous sommes intéressés à élaborer un outil informatique reposant sur une stratégie d'analyse similaire à celle d'un utilisateur, qui consiste à interpréter la scène en considérant, en premier lieu, les grandes zones composant les territoires (structurées par des éléments linéaires) puis à affiner récursivement et itérativement le niveau d'interprétation pour en extraire des zones plus spécialisées.

Axe 2 : intégration de connaissances dans le processus d'analyse

Les travaux menés dans l'axe 1 ont abouti à la proposition d'un processus d'analyse, permettant d'extraire des hiérarchies d'objets d'intérêt complexes d'une manière descendante, à partir d'images multirésolutions. La principale originalité de ce dernier est de tirer parti de la complémentarité de ces données durant l'étape de fouille de données. De plus, il repose sur des approches non-supervisées qui ne requièrent que peu d'informations *a priori* sur le contenu des images analysées. Néanmoins, ce processus est affecté de certaines limites : des erreurs peuvent avoir été commises lors des étapes de segmentation et/ou de classification pour chacun des niveaux de résolution, et le principe descendant implique alors une propagation de ces erreurs d'une résolution à l'autre. En conséquence, les hiérarchies extraites peuvent comporter des erreurs, les rendant alors difficilement exploitables par l'utilisateur pour des tâches de cartographie. Dans un deuxième axe de recherche, nous nous sommes donc intéressés à définir un post-processus permettant de corriger les résultats de classification obtenus aux différentes résolutions. Pour ce faire, nous nous sommes appliqués, dans un premier temps, à utiliser des connaissances relatives au contenu des images, acquises pendant la phase descendante par l'analyse des données multirésolutions. Ces connaissances offrent des informations de spécialisation ou de généralisation, relatives aux différents niveaux d'analyse. Dans un deuxième temps, nous nous sommes intéressés à définir des stratégies permettant d'ajouter des connaissances du domaine pour guider cette remise en cause des résultats. De nombreuses recherches menées au cours de ces dernières années ont en effet montré que devant la complexité des données, les approches non-supervisées pouvaient s'avérer insuffisantes et qu'une solution pour y remédier consiste à y intégrer des connaissances de haut niveau.

Structure du document

Ce document est structuré en trois parties principales, plus une annexe, brièvement introduites ci-dessous :

Partie I

Approches basées régions pour l'analyse d'images : segmenter pour classifier

Dans cette partie, nous introduisons le contexte de cette thèse en présentant l'intérêt et les limites des approches basées régions classiques pour l'analyse d'images et l'extraction d'objets complexes. Nous étudions ensuite les différentes approches proposées jusqu'alors dans la littérature, concernant l'utilisation conjointe et collaborative de données multirésolutions, pour apporter des solutions aux problèmes précédemment présentés. Face aux besoins soulevés lors de cette étude, nous introduisons finalement une proposition originale d'un processus d'analyse basé régions, détaillé dans la suite de ce manuscrit.

Partie II

Processus collaboratif pour l'analyse descendante d'images multirésolutions

Au cours de cette partie, nous présentons un processus collaboratif original permettant d'analyser des images multirésolutions, pour en extraire des hiérarchies d'objets d'intérêt. Basé sur le principe d'une analyse descendante, ce processus a pour objectif d'automatiser au maximum la méthode d'analyse classiquement employée par les experts en photo-interprétation : extraire les grandes zones d'intérêt de la scène observée (*e.g.*, zones urbaines, zones agricoles) à partir des images aux résolutions les plus basses, puis les décomposer récursivement pour en extraire des zones d'intérêt plus fines (*e.g.*, objets urbains, champs) à partir des images aux résolutions les plus élevées. Ce processus implique des mécanismes de segmentation hiérarchique interactive et de classification multirésolution, eux-aussi présentés en détail dans cette partie. Les géosciences et plus particulièrement l'observation de la Terre *via* les images de télédétection étant le domaine privilégié d'application des propositions faites lors de cette thèse, nous présentons une validation expérimentale de ce processus dans le cadre de l'analyse d'environnements urbains.

Partie III

Vers un processus itératif guidé par des connaissances

Dans cette partie, nous étudions des mécanismes permettant de remettre en cause et corriger les résultats d'analyse obtenus aux différentes résolutions. Ces mécanismes sont principalement basés sur l'intégration de connaissances du domaine de haut niveau (apportées par un utilisateur) et de connaissances liées au contenu des images traitées (générées par l'intermédiaire du processus d'analyse descendant) dans un post-processus, appliqué à la suite du processus collaboratif descendant. Dans ce contexte, nous présentons une distance originale basée sur la sémantique issue des connaissances du domaine, permettant de comparer des données caractérisées par le biais d'attributs multirésolutions décrits sous la forme d'histogrammes. Nous évoquons ensuite la perspective d'un processus itératif permettant une re-segmentation et une re-classification « corrective » des résultats liés aux différents niveaux d'analyse, dans le but d'en améliorer la précision. Nous présentons une validation expérimentale de ce post-processus et de cette distance, dans le cadre de l'analyse d'environnements urbains à partir d'ensembles d'images satellites multirésolutions.

Annexe

En annexe A, nous détaillons les différentes publications scientifiques rédigées dans le cadre de la diffusion de ces recherches. En annexe B, nous illustrons les différents jeux de données utilisés dans le cadre des validations expérimentales. En annexe C, nous introduisons des travaux supplémentaires portant sur l'étude de nouvelles méthodologies pour la classification de séries temporelles d'images satellites. Finalement, en annexe D, nous présentons les principaux développements logiciels réalisés dans le cadre de cette thèse.

Première partie

Approches basées régions pour l'analyse d'images : segmenter pour classifier

★ Résumé ★

Dans cette partie, nous introduisons le contexte de cette thèse en présentant l'intérêt et les limites des approches basées régions classiques pour l'analyse d'images et l'extraction d'objets complexes. Nous étudions ensuite les différentes approches proposées jusqu'alors dans la littérature, concernant l'utilisation conjointe et collaborative de données multirésolutions, pour apporter des solutions aux problèmes précédemment présentés. Face aux besoins soulevés lors de cette étude, nous introduisons finalement une proposition originale d'un processus d'analyse basé régions, détaillé dans la suite de ce manuscrit.

La fouille de données pour l'analyse d'images optiques

Sommaire

1.1	Analyse de masses de données image	9
1.1.1	Vision par ordinateur et analyse d'images	9
1.1.2	De grandes masses de données	10
1.1.3	Vers une automatisation du processus d'analyse	10
1.2	Analyse automatique d'images	11
1.2.1	Classification et analyse d'images	11
1.2.2	Approches supervisées	12
1.2.3	Approches non-supervisées	13
1.2.4	Discussion	14
1.3	Les approches basées pixels pour l'analyse d'images	15
1.3.1	Les pixels comme données à classifier	15
1.3.2	Limites des approches basées pixels	16
1.3.3	Vers des approches basées régions	17

Nous décrivons dans ce chapitre l'intérêt de l'emploi de la fouille de données pour l'analyse d'images optiques. Pour ce faire, nous commençons par rappeler les principes de l'extraction visuelle de connaissances à partir d'images et par montrer que, devant les grandes masses de données image, ces approches s'avèrent insuffisantes (Section 1.1). Nous introduisons ensuite la fouille de données, et en particulier la classification, comme une solution possible pour l'automatisation de l'analyse d'images (Section 1.2). Finalement, nous présentons les approches de classification basées pixels ainsi que leurs limites pour l'analyse de données à (très) hautes résolutions spatiales (Section 1.3).

1.1 Analyse de masses de données image

1.1.1 Vision par ordinateur et analyse d'images

La vision nous permet de percevoir et d'interpréter le monde qui nous entoure. Depuis le début des années 1980, la vision par ordinateur (aussi nommée vision artificielle) fait l'objet d'un intérêt significatif, d'une part du fait d'une technologie à coût décroissant et de plus en plus efficace, et d'autre part du fait de besoins croissants dans de nombreux domaines applicatifs comme ceux du médical, de la sécurité ou encore de la télédétection.

La vision par ordinateur a pour but de reproduire certaines fonctionnalités de la vision humaine au travers de l'analyse d'images. Elle consiste notamment à reconnaître dans une image des éléments et/ou des objets portant une sémantique identifiable par un utilisateur. Considérons l'exemple d'une image satellite. L'ordinateur, ou la machine utilisée, doit être capable de reconnaître des objets d'intérêt, prenant dans ce type d'images la forme d'objets géographiques (*e.g.*, toits des habitations, routes, zones de végétation)¹. Ces

1. Par la suite, on nommera objet géographique, un objet physique présent dans la scène observée et représenté à travers une image. De même, on nommera objet d'intérêt, un objet géographique que l'utilisateur souhaite pouvoir étudier à partir de l'image.

objets d'intérêt peuvent être considérés comme des connaissances relatives aux contenus des images étudiées. Par exemple, reconnaître qu'une scène urbaine (que l'on observe au travers de l'image) est composée de 1 569 maisons individuelles et 2 572 m² de végétation est une connaissance liée au contenu d'une scène urbaine. Ainsi, analyser une image consiste à en extraire des objets d'intérêt généralement liés à des instances d'une connaissance particulière (*e.g.*, géographique), vis-à-vis de la scène observée.

Depuis l'apparition des premières images, la méthode classiquement employée par les experts en photo-interprétation, consiste à extraire visuellement/manuellement les informations contenues dans ces données. Pour ce faire, des clés d'interprétation ont été définies [Collet et al., 2001] pour guider et uniformiser les techniques d'analyse employées par les utilisateurs. Cependant, ce processus est particulièrement consommateur de temps et peut conduire à des erreurs humaines d'interprétation. L'automatisation de l'extraction d'informations à partir d'images semble ainsi nécessaire.

1.1.2 De grandes masses de données

Une importante croissance des capacités de stockage, liée à une diversification et une accessibilité accrue des dispositifs d'acquisition, et aux réseaux à large bande, a créé un contexte technologique qui a drastiquement changé les usages de l'information visuelle telle que l'image et la vidéo. Dans de nombreux domaines, les quantités d'images et de vidéos acquises (et manipulées) deviennent de plus en plus importantes, formant de véritables masses de données. Nous donnons ci-après quelques exemples illustrant ce phénomène.

- Dans le domaine des médias et des loisirs, la diminution du prix des appareils photos numériques induit des quantités d'images et de vidéos circulant sur les réseaux de plus en plus conséquentes. Ainsi, en 2012, plus de 3 milliards de vidéos sont regardées chaque jour sur YouTube. De plus, les utilisateurs mettent en ligne l'équivalent de 240 000 longs métrages chaque semaine ².
- Dans le domaine de la santé et de l'imagerie médicale, les moyens d'acquisition et de visualisation d'images du corps humain sont de plus en plus variés et utilisés à grande échelle. La démocratisation de l'imagerie par résonance magnétique ou par tomographie à rayons X permet aujourd'hui d'obtenir des données 3D des patients, et les chirurgiens attendent de l'analyse de telles images de pouvoir s'entraîner virtuellement sur un patient avant d'opérer réellement.
- Dans le domaine aérospatial, les données issues de capteurs satellitaires sont de plus en plus accessibles. En particulier, différents systèmes satellitaires sont maintenant disponibles pour des besoins civiles et produisent une masse de données importante utilisable pour l'observation de la Terre et de l'Univers. En 2011, la nouvelle génération de satellites PLÉIADES a été lancée et permet aujourd'hui d'obtenir des données relatives à la surface terrestre à une résolution spatiale submétrique.

La dimension des grandes masses de données s'impose aussi avec force, rendant les techniques classiques d'analyse inadéquates. Les techniques manuelles de photo-interprétation deviennent alors insuffisantes pour analyser et extraire de la connaissance à partir de ces images. La recherche dans le domaine de l'analyse d'images doit donc répondre à ces besoins par la prise en compte des difficultés engendrées par ces nouveaux volumes d'informations.

1.1.3 Vers une automatisation du processus d'analyse

Analyser une image consiste à en extraire des connaissances relatives au contenu de la scène observée. L'automatisation de cette tâche vise à réduire au maximum l'intervention humaine au sein de ce processus.

L'extraction automatique de connaissances à partir de données est définie comme un processus de découverte d'informations implicites, inconnues auparavant et potentiellement utiles à partir de données [Frawley et al., 1992]. Ce processus peut se décomposer en plusieurs étapes : préparation des données (recherche, nettoyage et codage de celles-ci), fouille des données (recherche d'un modèle, de connaissances), validation et interprétation du résultat, et enfin intégration des connaissances apprises.

La fouille de données (ou *data mining*) est un maillon important dans cette chaîne de traitements des données vers les connaissances. Ainsi de nombreuses méthodes issues de la statistique, de l'analyse de données,

2. Voir http://www.youtube.com/t/press_statistics.

de l'apprentissage automatique, et de la reconnaissance de formes ont été étudiées et mises en œuvre pour fouiller les données, les structurer et les classifier, mais aussi pour les expliquer et produire des modèles prédictifs [Cornuéjols et Miclet, 2010].

Il est important de noter que suivant les connaissances que l'on souhaite extraire d'une image, les approches de fouille de données employées pourront s'avérer très différentes. On peut discerner deux grandes familles de cas d'utilisation :

1. on peut souhaiter extraire un ou plusieurs objets d'intérêt particuliers de l'image (*i.e.*, distinction entre le *fond* de l'image et la *forme* de l'objet d'intérêt); les méthodes généralement employées sont alors des approches *ad hoc*, conçues pour ne détecter dans l'image que des objets dont les propriétés sont similaires aux propriétés (connues) de l'objet recherché ;
2. on peut souhaiter traiter la totalité des objets d'intérêt de l'image (*i.e.*, l'image n'est alors plus composée d'un *fond* et d'une *forme*, mais uniquement de *formes*). Les méthodes généralement employées sont en général plus génériques que dans le cas d'utilisation précédent.

Dans le contexte de nos travaux, l'objectif est d'extraire, à partir d'images de télédétection, la totalité des objets d'intérêt géographiques composant une scène urbaine. De ce fait, nous nous intéresserons par la suite plus particulièrement à la deuxième famille d'approches, par le biais de l'utilisation d'algorithmes de classification.

1.2 Analyse automatique d'images

1.2.1 Classification et analyse d'images

Quand la tâche d'analyse consiste à extraire des connaissances de la totalité de l'image, la classification fournit une solution efficace pour automatiser ce traitement [Gonzalez et Woods, 2002]. La classification est une opération consistant à attribuer l'étiquette (aussi appelée label) d'une catégorie particulière à chaque élément d'un ensemble de données. Ces éléments sont décrits par des caractéristiques, encore appelées *attributs*, reflétant leurs propriétés.

Soit un ensemble de données D . Chaque donnée $d \in D$ peut être décrite par un vecteur \mathbf{v} composé de p attributs appartenant à un espace quelconque \mathbb{F} . On représente ce vecteur d'attributs par $\mathbf{v} = (a_j)_{j=1}^p$. Une classification \mathcal{C} de l'ensemble de données D peut alors être définie comme

$$\begin{aligned} \mathcal{C}: \mathbb{F}_1 \times \dots \times \mathbb{F}_p &\rightarrow \mathcal{L} \\ \mathbf{v} &\mapsto l \end{aligned} \quad (1.1)$$

où l est l'étiquette de la catégorie associée à la donnée d (représentée par le vecteur d'attributs \mathbf{v}) par l'algorithme de classification, et \mathcal{L} représente l'ensemble des étiquettes considérées. Une catégorie K_l est alors définie comme

$$K_l = \{d \in D \mid \mathcal{C}(\mathbf{v}) = l\} \quad (1.2)$$

c'est-à-dire comme l'ensemble des données ayant été classifiées avec l'étiquette l . L'ensemble \mathfrak{K} des catégories issues d'une classification \mathcal{C} est défini comme $\mathfrak{K} = \{K_l\}_{l \in \mathcal{L}}$.

Dans le cadre de l'analyse d'images, les éléments à classifier sont les données (*i.e.*, les pixels) composant l'image. Ces dernières peuvent alors être décrites par des attributs construits sur les grandeurs physiques (*e.g.*, radiométriques) mesurées par des capteurs.

Un utilisateur qui tente de catégoriser visuellement les objets observés dans une image, utilise ses facultés visuelles (ainsi que des clés d'interprétation visuelles définies à cet effet) pour identifier et étiqueter des pixels qui représentent des catégories pertinentes de la scène observée. La classification peut alors permettre d'automatiser ce processus d'étiquetage de pixels, en labellisant de la même façon des données présentant des caractéristiques similaires. L'objectif est alors d'assigner une catégorie particulière ou thème (*e.g.*, eau, forêt, route) à chacun des pixels d'une image, produisant ainsi une image thématique porteuse d'informations relatives à la scène observée. Pour visualiser cette image, il est alors possible d'affecter une couleur particulière (ou aléatoire) à chaque pixel suivant le thème qui lui a été attribué, produisant ainsi une image thématique colorisée. La figure 1.1 illustre cette technique de visualisation pour l'analyse d'une image satellite.

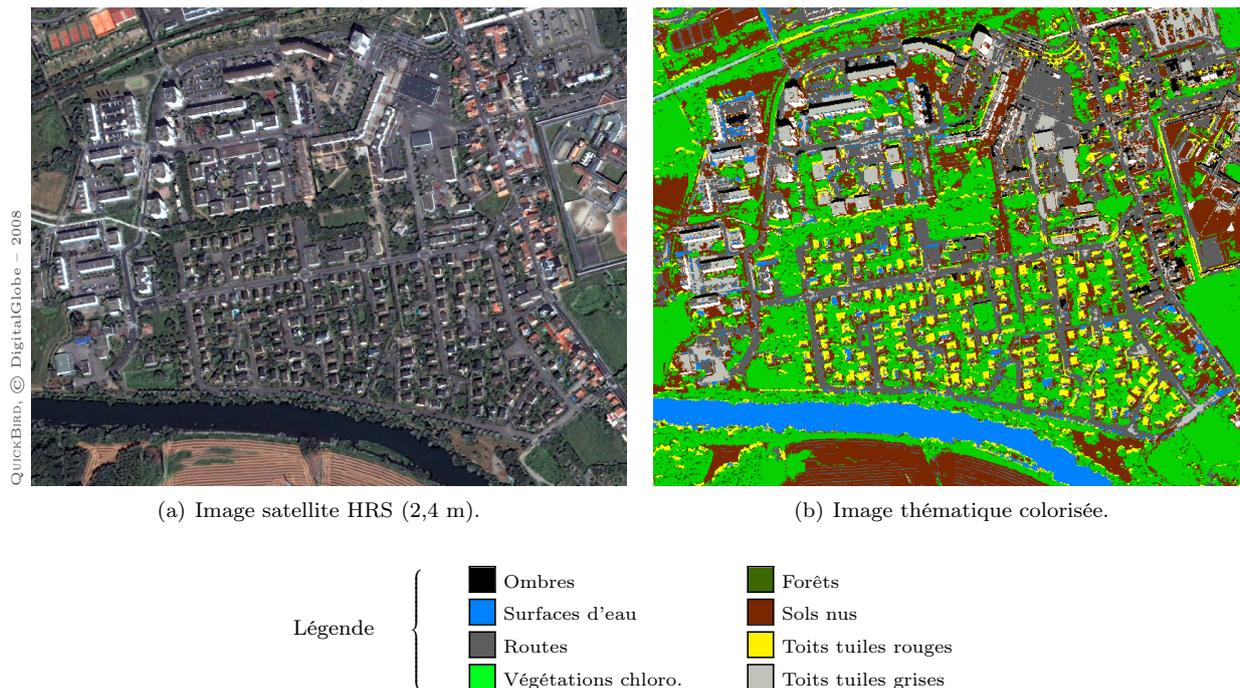


FIGURE 1.1 – Image thématique résultant de la classification d'une image satellite.

Deux types d'approches peuvent être considérés pour labelliser des éléments de manière automatique : les méthodes de classification supervisée et les méthodes de classification non-supervisée. Nous présentons par la suite les principes de ces approches et leurs avantages/inconvénients respectifs.

1.2.2 Approches supervisées

A. Principe

Les approches supervisées sont utilisées lorsque des catégories existent *a priori* dans un jeu de données, et que le problème consiste à créer un modèle permettant d'assigner des éléments à ces catégories. Dans ce cas particulier, les catégories sont appelées des classes. Chaque classe peut être identifiée par une étiquette qui porte généralement une sémantique (*e.g.*, Toit, Route, Sol nu). La classification supervisée repose sur la création d'un modèle d'apprentissage. Cette création est réalisée par le biais d'un ensemble d'exemples, c'est-à-dire un ensemble d'éléments dont la classe est connue *a priori*. Par exemple, dans le cadre de l'analyse d'images, l'utilisateur va identifier des pixels qui sont représentatifs de différents types de surfaces spécifiques, formant ainsi l'ensemble de données d'apprentissage. L'objectif, à partir de ces exemples, est de découvrir un modèle de classes pouvant être généralisé à un ensemble de données plus large, sous la forme d'un modèle prédictif. Ce processus comporte deux étapes : (1) une étape de construction (d'apprentissage) du modèle à partir des exemples (dont la classe est connue), suivie (2) d'une étape de classement des objets dont la classe est inconnue, consistant à affecter un label de classe à ces objets.

Parmi les nombreuses approches proposées dans la littérature, on peut citer les méthodes basées sur le principe des séparateurs à vaste marge (SVM) développés depuis les années 1990 [Schölkopf et Smola, 2002]. Ces méthodes de classification permettent de traiter des problèmes de discrimination non-linéaire. Pour une étude plus générale et comparative sur les différentes approches de classification supervisée proposées dans la littérature, les lecteurs pourront se référer à [Alpaydin, 2004].

B. Limites

Bien que très souvent utilisée dans la plupart des tâches de classification, la classification supervisée souffre de certaines limites :

- la première est liée au fait que les classes recherchées doivent être connues *a priori* par l'utilisateur. En conséquence, ce type d'approches ne permet de labelliser les données qu'avec des étiquettes de classes supposées *a priori* être présentes au sein de ces données. De ce fait, les approches supervisées manquent de souplesse pour classer des données ne correspondant à aucune des classes recherchées ;
- la deuxième est liée à la construction du modèle d'apprentissage. La construction de ce modèle est basée sur les connaissances de l'utilisateur, sa familiarité avec les classes pouvant être recherchées dans les données, et donc sa capacité à définir des exemples d'apprentissage. La définition de ces exemples, nécessaires à la construction du modèle d'apprentissage, nécessite généralement l'intervention d'un utilisateur, qui va manuellement affecter une classe à chaque élément. Dans certains cas, cette intervention peut s'avérer fastidieuse ;
- la troisième est liée à la validité du modèle d'apprentissage ; en effet, le modèle construit *via* ces exemples sur un jeu de données particulier n'est pas toujours réutilisable sur un jeu de données différent.

1.2.3 Approches non-supervisées

A. Principe

La classification non-supervisée (classification automatique ou encore *clustering*) procède d'une manière différente. Son objectif est de faire émerger des groupes homogènes de données (dans l'espace des caractéristiques), sans aucune information *a priori*. Les groupes créés sont appelés clusters et possèdent des étiquettes dépourvues de sémantique (*e.g.*, Cluster n°7). Le but est de diviser un ensemble de données en groupes tels que, pour une mesure de similarité donnée, les objets d'un groupe soient les plus similaires possibles (homogénéité intra-groupe) et que les objets de groupes différents soient les plus différents possibles (séparabilité inter-groupe). Cependant, un tel processus de partitionnement conduit naturellement à former un groupe différent pour chaque donnée de l'ensemble à classer. Ainsi, le problème majeur de la classification non-supervisée est de trouver un compromis entre une homogénéité intra-groupe et une séparation inter-groupe conduisant à la formation de groupes pertinents. L'objectif consiste alors à découvrir la structure sous-jacente des données pour en extraire de l'information. Dans la plupart des cas, le nombre de groupes à former est un paramètre de l'algorithme que l'utilisateur doit spécifier.

Parmi les nombreuses approches proposées dans la littérature, on peut notamment citer les méthodes basées sur une distance entre objets pour effectuer le regroupement, et en particulier les méthodes basées sur un prototype. Ces méthodes considèrent que les objets proches dans l'espace des données appartiennent au même cluster. Pour ce faire, le principe consiste à définir les clusters en fonction d'un objet représentatif. Ce dernier est appelé centroïde s'il s'agit du centre de gravité des objets appartenant au cluster (ou médoïde s'il s'agit d'un objet particulier du cluster). La plus connue de ces méthodes est l'algorithme *K*-MEANS [Macqueen, 1967; Jain, 2010]. *K*-MEANS est un algorithme itératif qui permet de diviser l'espace des données en différents groupes homogènes. Chacun de ces groupes (ou clusters) est symbolisé par la moyenne des objets dans ce regroupement. Ainsi, chaque élément appartient au cluster dont la distance au centroïde est la plus faible. L'algorithme se décompose en deux étapes successives :

1. choisir k centroïdes (et ainsi k moyennes) initiaux (*e.g.*, k moyennes au hasard) ;
2. répéter jusqu'à convergence :
 - assigner chaque objet au centroïde le plus proche (par rapport à la distance choisie pour comparer les données) ;
 - mettre à jour la moyenne associée à chaque centroïde en fonction des objets qui lui ont été associés.

Pour une étude plus générale, et comparative, sur les différentes approches de classification non-supervisée proposées dans la littérature, les lecteurs pourront se référer à [Jain et al., 1999].

B. Limites

La principale limite, inhérente à la classification non-supervisée des données, réside dans le choix du nombre de clusters. Automatiser la définition du nombre de groupes à extraire automatiquement est l'un des problèmes les plus complexes en clustering. Il est souvent nécessaire de fournir le nombre de clusters souhaités comme paramètre de l'algorithme. Le choix de ce paramètre a souvent été étudié comme un problème de sélection de modèle [Baudry, 2010]. En fonction des données considérées, on recherche le modèle de regroupement le mieux adapté pour les classifier. Dans ce cas, l'algorithme est généralement exécuté plusieurs fois, indépendamment, avec un nombre de clusters différent. Les résultats sont ensuite comparés en se basant sur un critère de sélection permettant de choisir la meilleure solution. Ce choix est souvent subjectif et fortement dépendant du critère sélectionné pour comparer les résultats.

Pour éviter ce problème de sélection de modèle, des solutions ont été proposées. Parmi celles-ci, on peut citer l'algorithme ISODATA [Jain et Dubes, 1988], basé sur l'algorithme K -MEANS, qui permet une évolution du nombre de clusters durant l'exécution. Des clusters sont fusionnés si le nombre d'objets dans l'un des clusters est inférieur à un seuil ou si deux clusters sont plus proches qu'un certain seuil. De plus, les clusters peuvent également être séparés en deux si leur dispersion (évaluée par l'écart-type) dépasse un seuil donné. Cet algorithme permet, contrairement à K -MEANS, d'avoir un nombre de clusters variable pendant l'exécution. Cependant, cette approche nécessite la définition de nombreux seuils et ne fait que transformer le problème du choix du nombre de clusters en un problème relatif à la définition de ces paramètres.

Pour résoudre ces problèmes d'une manière plus générale, différentes approches ont été proposées. La méthode SAMARAH [Wemmert, 2000; Gançarski et Wemmert, 2007; Forestier, 2010] est une approche de clustering qui consiste à faire collaborer plusieurs méthodes de classification non-supervisée pour tenter de trouver un consensus sur le clustering d'un jeu de données. Ces différentes méthodes vont partager des informations et remettre en cause leurs résultats en fonction des décisions prises par chacune. Ainsi, une discussion est engagée entre ces dernières dans le but de faire converger collectivement les différents résultats afin que ceux-ci soient suffisamment comparables pour qu'un mécanisme d'unification (basé par exemple sur un vote) puisse être mis en œuvre. L'objectif de cette collaboration est de réduire l'impact du choix d'une méthode et de ses paramètres (*e.g.*, le nombre de clusters) sur le résultat.

1.2.4 Discussion

A. Problème lié au choix de la méthode

De nombreuses méthodes de classification (supervisée ou non) sont ainsi disponibles lorsque la tâche consiste à labelliser automatiquement des éléments d'un jeu de données dans des catégories particulières. Cependant, un problème général soulevé par la classification réside dans le choix de la méthode à utiliser pour traiter un jeu de données particulier.

La plupart du temps, l'utilisateur est confronté au choix de la méthode, mais également à celui de ses paramètres. Il est aussi confronté à la question de la représentation des connaissances et de leur utilisation dans les algorithmes de classification pour améliorer le processus de fouille. En effet, quelles sont les informations les plus pertinentes pour caractériser au mieux les données à traiter ? Ces choix sont généralement liés au domaine d'application et à des connaissances *a priori* de l'utilisateur, mais aussi aux données à classer elles-mêmes.

B. Problème lié à la pondération des attributs

Un autre problème soulevé par la classification est lié au choix des pondérations affectées aux attributs caractérisant les données. Dans la perspective d'obtenir un meilleur résultat de classification, on cherche souvent à décrire ces dernières de la manière la plus détaillée possible. Les données sont alors représentées par de nombreux attributs. Ceux-ci peuvent cependant présenter les caractéristiques pénalisantes suivantes :

1. manque de pertinence : un attribut non pertinent n'apporte aucune information permettant de discriminer les classes entre elles ;
2. présence de bruit : un attribut bruité porte des informations incorrectes ;

3. corrélation entre les attributs : des attributs corrélés portent la même information ; cette information redondante aura implicitement plus de poids qu'une information portée par un unique attribut.

Ainsi, augmenter la dimension des données peut parfois nuire à la qualité de la classification.

Pour résoudre ces problèmes, différentes approches ont déjà été proposées. Dans [Gançarski et Blansché, 2008; Gançarski et al., 2008], les auteurs proposent de s'appuyer sur des algorithmes évolutionnaires afin d'optimiser la sélection/pondération de ces ensembles d'attributs. Bien que de telles approches aient fourni des résultats prometteurs, la détermination des pondérations affectées aux attributs caractérisant les données durant l'étape de classification reste un problème difficile à résoudre. Ces problèmes dépassant le cadre de ces travaux de thèse, nous en ferons abstraction dans la suite de ce manuscrit.

C. Classes thématiques et classes des caractéristiques

Lorsque l'on parle de catégories ou de classes, il convient de faire la distinction entre les classes thématiques et les classes des caractéristiques, effectivement discernables au sein des données (*i.e.*, de l'image). Les classes thématiques sont des catégories d'intérêt que l'utilisateur tente d'identifier dans les images, comme différents types de toits de bâtiments, de végétation, *etc.* Les classes des caractéristiques sont des ensembles de données (*e.g.*, des pixels) qui ont des caractéristiques (*e.g.*, valeurs radiométriques) similaires au sein d'une image. Dans la littérature, on parle souvent des classes de l'espace des caractéristiques (ou de l'espace des données).

L'objectif ultime de la classification est d'établir la correspondance entre les classes des caractéristiques et les classes thématiques. Cependant, il est souvent difficile d'établir une correspondance directe entre ces deux types d'informations. Certaines classes des caractéristiques bien définies peuvent parfois émerger sans qu'elles correspondent nécessairement à des classes d'information intéressantes pour l'analyse. De plus, une classe thématique très large peut contenir plusieurs sous-classes des caractéristiques avec des variations non négligeables. En utilisant l'exemple de la classe thématique *Toit de bâtiments*, les sous-classes des caractéristiques peuvent être liées à des variations dans la couleur, la forme ou simplement par les effets d'ombrage ou des variations dans l'illumination. L'utilisateur a alors pour rôle de déterminer l'utilité des différentes classes des caractéristiques et de valider leur correspondance à des classes thématiques utiles.

D. Positionnement

Malgré les problèmes relatifs aux différentes approches de classification (*e.g.*, définition des exemples d'apprentissage pour les approches supervisées, définition du nombre de clusters pour les approches non-supervisées), ces dernières permettent d'automatiser efficacement les processus d'extraction de connaissances à partir d'images. De plus, ces approches ont l'avantage d'avoir été largement étudiées ces trente dernières années, menant ainsi à de nombreux ouvrages quant à leurs utilisations respectives (on peut citer [Jensen et Lulla, 1987] comme exemple de travail pionnier).

Dans notre contexte applicatif (l'analyse de scènes urbaines), les données à traiter sont issues d'images à (très) hautes résolutions spatiales. Ces dernières ont la particularité de présenter un important niveau de détail, induisant un très grand nombre d'objets géographiques extractibles, qu'il est quasiment impossible de déterminer à l'avance. En conséquence, il est bien souvent difficile d'employer des approches supervisées nécessitant de connaître *a priori* les classes à extraire, et la définition de nombreux exemples d'apprentissage.

De fait, nous nous sommes uniquement intéressés, au cours de cette thèse, à l'utilisation d'approches de classification non-supervisée dans le cadre de l'analyse d'images de télédétection. Nous décrivons par la suite différentes méthodes, reposant sur des approches de classification et de fouille de données, permettant d'aider un utilisateur, expert en photo-interprétation, dans le processus d'extraction de connaissances à partir de données images.

1.3 Les approches basées pixels pour l'analyse d'images

1.3.1 Les pixels comme données à classifier

Les approches basées pixels considèrent les pixels de l'image comme les données élémentaires à traiter. Ce type d'approches a été largement étudié ces vingt dernières années dans le cadre de l'analyse d'images

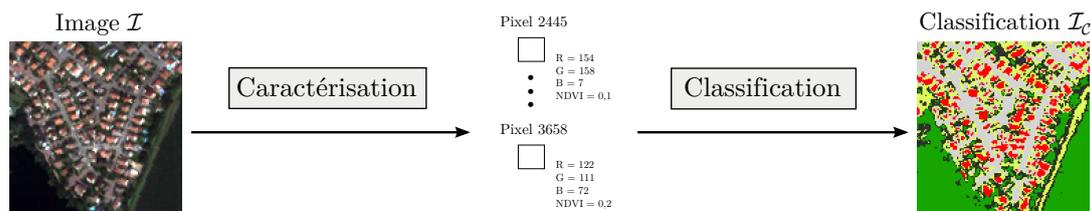


FIGURE 1.2 – Schéma d'une approche basée pixels pour l'analyse d'images.

[Jähne, 2005]. En imagerie satellitaire, de nombreuses approches (spécifiques ou non) ont été proposées et sont encore actuellement utilisées en routine opérationnelle. Pour une introduction à l'emploi de certaines de ces approches, les lecteurs pourront se référer à [Richards, 1986; Jasani et al., 2009].

La plupart des approches basées pixels classifient les pixels décrits par des attributs construits directement sur les valeurs des grandeurs physiques (*e.g.*, radiométriques, géométriques) mesurées par les capteurs et représentées en chaque point de l'image. Par exemple, un pixel $\mathbf{x} \in E$ d'une image multispectrale $\mathcal{I} : E \rightarrow V$, composée de s bandes spectrales, peut être caractérisé par un vecteur de données (*i.e.*, d'attributs) $\mathbf{v} = (\mathcal{I}_1(\mathbf{x}), \mathcal{I}_2(\mathbf{x}), \dots, \mathcal{I}_s(\mathbf{x}))$ où chaque composante correspond à la valeur du pixel dans un certain espace radiométrique. Dans cet exemple, le nombre d'attributs p caractérisant chaque pixel est égal au nombre s de bandes spectrales composant l'image. Dans le cas d'une image multispectrale acquise par le satellite SPOT-5, chaque pixel peut être caractérisé par 4 valeurs radiométriques issues des 4 bandes spectrales de l'image.

Chaque pixel peut ensuite être labellisé par une étiquette *via* l'utilisation d'une fonction de classification \mathcal{C} (supervisée ou non). Le résultat de la classification peut alors être vu comme une image composée d'une mosaïque de pixels, appartenant chacun à une catégorie particulière. Une telle image, dite classifiée, pourra ainsi être vue comme une fonction

$$\begin{aligned} \mathcal{I}_c : E &\rightarrow \mathcal{L} \\ \mathbf{x} &\mapsto l \end{aligned} \quad (1.3)$$

qui à chaque point \mathbf{x} de la scène, associe l'étiquette $l \in \mathcal{L}$ d'un cluster (*resp.*, d'une classe) spécifique.

Dans le cas où une approche de classification supervisée est utilisée, ces pixels sont intrinsèquement labellisés par l'étiquette de la classe (portant une sémantique) à laquelle ils ont été affectés par l'algorithme. À l'opposé, dans le cas où une approche de classification non-supervisée est utilisée, les clusters de pixels ainsi formés ne possèdent pas d'étiquette thématique/sémantique. L'utilisateur peut alors choisir (ou non) d'affecter à ces clusters une étiquette correspondant à une réalité thématique. Ces étiquettes seront ensuite attribuées à tous les pixels appartenant au cluster reconnu. Le schéma proposé par la figure 1.2 illustre l'utilisation d'une approche basée pixels pour la classification d'images.

1.3.2 Limites des approches basées pixels

Bien qu'encore largement utilisées en routine opérationnelle dans le cadre de l'analyse automatique d'images³, les approches basées pixels souffrent de divers inconvénients.

L'inconvénient principal de l'emploi de ces approches réside dans le fait qu'elles n'utilisent, généralement, que les informations radiométriques des pixels (ou encore des indices basés sur des ratios entre les différentes bandes spectrales [Carlson et Ripley, 1997]) pour les caractériser lors de l'étape de classification. En règle générale, les composants physiques (*i.e.*, les matériaux) des objets géographiques représentés dans les images possèdent des valeurs spectrales différentes et les données résultantes peuvent alors être différenciées dans l'espace des caractéristiques par les approches basées pixels. Cependant, quand des objets géographiques thématiquement différents présentent des valeurs radiométriques similaires (*e.g.*, une route en macadam et certains toits industriels; une ombre portée et une zone d'eau), les propriétés radiométriques de ces derniers peuvent, dans certains cas, n'être pas suffisantes pour caractériser d'une manière efficace les données. Dans ce cas précis, l'emploi des approches pixels ne permet pas de classifier correctement les images : les classes des

3. Les lecteurs pourront se référer au logiciel ENVI 5.0 proposé par EXELIS pour l'analyse automatique d'images. Pour plus d'informations, voir <http://www.exelisvis.com>.

caractéristiques peuvent notamment recouvrir plusieurs classes thématiques que l'utilisateur souhaite extraire. Par la suite, ce phénomène sera nommé le problème de la séparabilité spectrale.

Un autre inconvénient réside dans l'analyse d'images à hautes résolutions spatiales (HRS). En effet, l'augmentation de la résolution spatiale des images a conduit à un accroissement important des détails présents dans celles-ci (voir figure 1.1 (a)). Jusqu'à la fin des années 1990, l'un des principaux problèmes liés aux images satellites résidait dans le fait qu'un pixel pouvait représenter une zone composée de plusieurs objets de natures très différentes (et cela d'autant plus que la résolution spatiale était basse). On parle alors de pixel *mixte* ou de *mixel*. Par exemple, les images à moyennes résolutions spatiales (MRS) ne permettent qu'une analyse des zones thématiques (*e.g.*, zones bâties denses ou clairsemées, végétation, eau) et/ou structurelle du paysage (*e.g.*, extraction de réseaux routiers ou ferrés, de limites urbaines). Les images HRS permettent désormais une analyse des objets primitifs eux-mêmes (*e.g.*, maisons individuelles, routes) [Puissant et Weber, 2002]. Paradoxalement, l'augmentation de la précision spatiale des images entraîne également des perturbations dans l'analyse. D'une part, les objets d'intérêt sont composés, dans ces nouvelles images, de zones de pixels hétérogènes. D'autre part, l'utilisateur est confronté à des images présentant un très grand nombre de classes, dû à l'émergence de nombreux détails à ce niveau de résolution (*e.g.*, différents types de toits, ombres portées). Le nombre de classes thématiques pouvant être extraites devient donc très difficile à évaluer et les clusters/classes résultant(e)s deviennent difficilement interprétables par l'utilisateur.

Afin de résoudre ces problèmes, des extensions aux approches basées pixels ont été développées. Certaines approches proposent ainsi de caractériser les pixels par des informations calculées dans leur voisinage avant de les classifier [Kittler et Foglein, 1984; Shekhar et al., 2002]. Ces informations relatives au voisinage peuvent alors permettre de compléter les propriétés radiométriques des pixels, parfois insuffisantes pour caractériser efficacement ces derniers. La plupart des approches présentées dans la littérature proposent de caractériser les pixels à classifier par des indices de texture calculés sur une fenêtre prenant en compte leurs voisinages dans l'image. Ce type d'approches est un premier pas vers la prise en compte des informations contextuelles des pixels dans un processus de classification. Cependant, des travaux préliminaires ont montré que l'utilisation de ces informations de contexte n'est pas toujours suffisante pour l'analyse de scènes urbaines complexes à partir d'images à hautes résolutions spatiales [Weng, 2012].

Ces limites ont ainsi conduit au développement de nouvelles méthodes de traitement et d'interprétation des images.

1.3.3 Vers des approches basées régions

Afin de permettre l'extraction de connaissances à partir de grandes masses de données à hautes résolutions spatiales, de nouvelles méthodes d'analyse d'images ont été proposées. Celles-ci sont généralement appelées méthodes basées régions (ou orientées objets). Dans ces approches, l'image est divisée en un ensemble de segments par un processus de segmentation (manuel ou automatisé). Ces segments sont ensuite caractérisés par des attributs variés (*e.g.*, informations colorimétriques, texturales, géométriques ou contextuelles). Ce sont ces segments caractérisés qui sont ensuite traités par des algorithmes de classification. L'idée principale réside dans l'hypothèse que ces segments délimitent les frontières des objets géographiques de l'image ayant une réalité sémantique (*e.g.*, portions de routes, toits des bâtiments).

Ce nouveau paradigme basé régions s'inspire des principes de la vision humaine. En effet, les spécialistes de la *Gestalt* (la psychologie de la forme) tout comme les psycho-physiciens reconnaissent qu'un processus de segmentation intervient lors des toutes premières étapes du processus de perception visuelle [Li, 2000]. De plus, ils ont prouvé que ces premières étapes sont totalement indépendantes d'un quelconque apprentissage ou d'une connaissance *a priori* sur le monde observé [Roelfsema et al., 1998].

Les approches basées régions [Blaschke, 2010] représentent ainsi une solution potentielle pour résoudre certains problèmes liés à l'utilisation d'approches basées pixels pour l'analyse de données à hautes résolutions spatiales. En effet, reconstruire les objets d'intérêt recherchés avant de les classifier peut permettre :

1. d'obtenir une meilleure caractérisation des données. Combiner plusieurs informations contenues dans les images, comme des informations géométriques et radiométriques, peut notamment accroître la séparabilité des données représentant des objets géographiques ayant des réponses spectrales similaires ;

2. de traiter des objets d'intérêt apparaissant comme hétérogènes dans les images. La classification directe des segments de l'image peut permettre de réduire considérablement le nombre de classes des caractéristiques à traiter.

Les approches basées régions pour l'analyse d'images

Sommaire

2.1 Principe des approches basées régions	19
2.1.1 Une collaboration segmentation–classification	19
2.1.2 Création des segments de l'image	20
2.1.3 Méthodologies de segmentation	22
2.1.4 Ajout de connaissances dans le processus de segmentation	25
2.2 L'apport de la segmentation d'images pour la classification	26
2.2.1 Caractérisation des segments <i>via</i> des attributs pertinents	26
2.2.2 Classifier des éléments de plus haut niveau sémantique	27
2.2.3 Utilisation de connaissances de haut niveau	27
2.3 Limites des approches basées régions	28
2.3.1 Un aspect paramétrique important	29
2.3.2 Des objets d'intérêt de plus en plus complexes	30
2.3.3 Vers des approches multiniveaux	32

Ce chapitre présente l'intérêt et les limites des approches basées régions pour l'analyse d'images. Pour ce faire, nous commençons par introduire les principes sur lesquels reposent ces approches (Section 2.1). Nous décrivons ensuite les avantages relatifs à l'utilisation d'une étape de segmentation avant l'étape de classification en analyse d'images (Section 2.2). Finalement, nous présentons les limites des approches basées régions pour l'extraction d'objets d'intérêt complexes à partir d'images à très hautes résolutions spatiales et nous introduisons les approches multiniveaux comme une solution potentielle à ces problèmes (Section 2.3).

2.1 Principe des approches basées régions

2.1.1 Une collaboration segmentation–classification

Les approches basées régions consistent à faire suivre une première étape de segmentation de l'image, par une seconde étape de classification (Figure 2.1). L'idée principale réside dans l'hypothèse que les segments résultant de l'étape de segmentation délimitent les frontières des objets géographiques de l'image et que ces segments peuvent permettre de mieux caractériser les données.

L'étape de segmentation consiste à regrouper des pixels en vue de constituer des ensembles connexes de pixels homogènes, appelés segments, formant une partition de l'image. Plus formellement, une segmentation d'une image $\mathcal{I} : E \rightarrow V$ est une partition $\mathfrak{S} = \{S_i\}_{i=1}^r$ ($r \geq 1$) de l'ensemble E . Un résultat de segmentation peut alors être vu comme une image composée d'une mosaïque de pixels, appartenant chacun à un segment particulier. Une telle image dite segmentée (ou de segments) pourra être vue comme une fonction

$$\begin{aligned} \mathcal{I}_{\mathfrak{S}} : E &\rightarrow \llbracket 1, r \rrbracket \\ \mathbf{x} &\mapsto i \end{aligned} \tag{2.1}$$

qui à chaque point \mathbf{x} de la scène, associe l'indice i d'un segment S_i particulier. Par équivalence, on a $S_i = \mathcal{I}_{\mathfrak{S}}^{-1}(\{i\})$ pour tout $i \in \llbracket 1, r \rrbracket$.

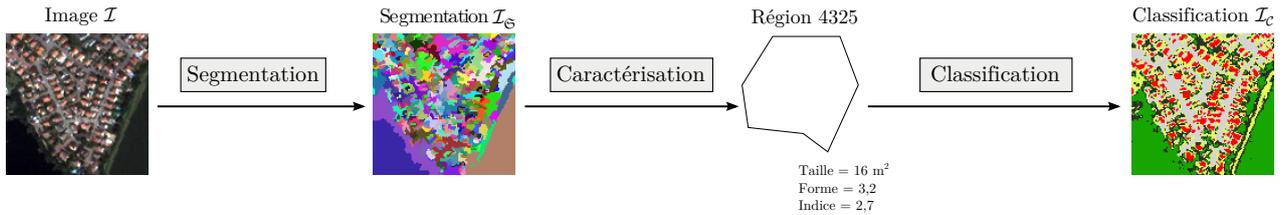


FIGURE 2.1 – Schéma d'une approche basée régions pour l'analyse d'images.

L'image segmentée résultante peut alors être visualisée en affectant une couleur aléatoire à chaque groupe de pixels formant un segment particulier, produisant ainsi une image de segments colorisée.

Chaque segment $S \in \mathfrak{S}$ peut ensuite être caractérisé par un vecteur d'attributs $\mathbf{v} = (a_j)_{j=1}^p$ où $a_j \in \mathbb{F}_j$, modélisant ses p propriétés intrinsèques. Les valeurs de ces attributs sont généralement dérivées des données brutes de l'image (*e.g.*, informations colorimétriques, taille, forme). Un segment $S \in \mathfrak{S}$, caractérisé par un vecteur d'attributs \mathbf{v} , est alors appelé une région de l'image \mathcal{I} .

Les régions sont ensuite labellisées avec différentes étiquettes *via* l'utilisation d'une fonction de classification \mathcal{C} qui, à chacun des segments $S_i \in \mathfrak{S}$ (caractérisé par le vecteur \mathbf{v}_i), associe le label $l \in \mathcal{L}$ (*resp.*, l'étiquette) d'un cluster (*resp.*, d'une classe) spécifique. L'image classifiée \mathcal{I}_C peut être obtenue *via* la fonction $\mathcal{C} \circ \mathcal{I}_{\mathfrak{S}}$. Le résultat de classification peut ici aussi être vu comme une image composée d'une mosaïque de pixels, appartenant chacun à une catégorie particulière.

Lors de l'étape de classification, il est alors possible d'utiliser une approche supervisée ou non-supervisée [Herold et al., 2003]. La définition présentée ci-dessus est suffisamment générique pour supporter ces deux types d'approches. Dans le cas de l'utilisation d'une approche supervisée, les exemples d'apprentissage à définir ne sont alors plus des pixels mais des segments caractérisés (*i.e.*, des régions) issus de l'étape de segmentation suivie de l'étape de caractérisation.

Les approches basées régions permettent ainsi d'aborder le problème de l'analyse d'images différemment des approches basées pixels. En effet, l'objectif de ce nouveau paradigme est de faire le lien entre des régions, délimitées et caractérisées au sein d'une image, et des objets géographiques présents dans la scène.

2.1.2 Création des segments de l'image

Pour former des ensembles homogènes de pixels connexes au sein d'une image, de nombreuses approches reposent sur l'utilisation d'algorithmes de segmentation. Nous présentons par la suite une définition de la segmentation ainsi que certaines notions nécessaires à la compréhension de son fonctionnement.

A. Segmentation d'images

La segmentation est une opération de traitement d'images qui a pour but de rassembler des pixels voisins entre eux afin de trouver les segments homogènes et les contours délimitant les objets représentés dans ces images [Gonzalez et Woods, 2002]. Les pixels connexes sont ainsi regroupés en segments et constituent une partition de l'image. Ces groupes de pixels sont supposés être pertinents, c'est-à-dire que les segments doivent correspondre aux parties significatives des objets d'intérêt recherchés. La figure 2.2 présente un exemple de résultat de segmentation. Pour visualiser ce résultat, deux approches sont possibles : la figure 2.2 (b) représente l'image des segments colorisée tandis que la figure 2.2 (c) représente l'image des frontières des segments représentés sur l'image considérée.

Une définition formelle de la segmentation complète d'une image a été proposée par Pavlidis [Pavlidis, 1977]. Cette définition est retranscrite ci-après.

Définition 1. Soit E le domaine de l'image et \mathcal{I} la fonction qui associe à chaque pixel \mathbf{x} une valeur $\mathcal{I}(\mathbf{x})$. Si nous définissons un prédicat P sur l'ensemble des parties de E , la segmentation de E est définie comme une partition \mathfrak{S} de E en r sous-ensembles S_1, \dots, S_r telle que :

1. $E = \bigsqcup_{i=1}^r S_i$;

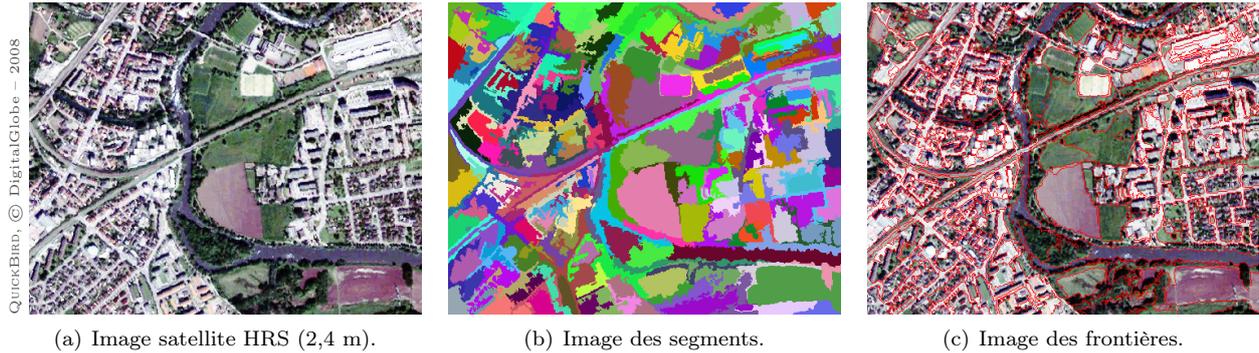


FIGURE 2.2 – Illustration du résultat d'un algorithme de segmentation.

2. $\forall i \in \llbracket 1, k \rrbracket, S_i$ est connexe ;
3. $\forall i \in \llbracket 1, k \rrbracket, P(S_i) = \text{vrai}$;
4. $\forall i, j \in \llbracket 1, k \rrbracket \times \llbracket 1, k \rrbracket, (S_i \text{ est adjacent à } S_j \text{ et } i \neq j) \Rightarrow (P(S_i \cup S_j) = \text{faux})$.

Le prédicat P est utilisé pour tester l'homogénéité des ensembles S_i . D'après la définition de Pavlidis, la segmentation d'une image correspond à sa décomposition en un ensemble de segments homogènes, le critère d'homogénéité P restant néanmoins à déterminer. Les conditions de la définition formelle de Pavlidis ont été exprimées, à la même période, par Zucker [Zucker, 1976] comme suit :

- (1.) la première condition implique que tout pixel de l'image doit appartenir à un et un seul segment. Cela implique notamment que l'algorithme de segmentation ne doit pas se terminer avant d'avoir exploré l'ensemble des pixels de l'image ;
- (2.) la seconde condition impose que tout segment doit être connexe (ce qui n'est pas le cas dans la définition d'une partition d'un ensemble). La connexité des segments est généralement induite par le voisinage défini sur l'image ;
- (3.) la troisième condition impose que chaque segment doit être homogène ;
- (4.) enfin, la quatrième condition est une condition de maximalité relativement à cette propriété d'homogénéité.

Néanmoins, ces quatre conditions nécessaires à l'obtention d'une segmentation complète d'une image, soulèvent certaines imprécisions rendant le processus de segmentation complexe. Les problèmes liés à ces imprécisions sont décrits ci-après.

B. Un problème mal posé

Il est important de remarquer que le nombre de segments formant la partition de l'image n'est pas considéré dans la définition précédente. Il peut, en particulier, exister plusieurs segmentations valides pour un prédicat P donné. Si les conditions (1.) et (2.) sont vérifiées par toute segmentation, les conditions (3.) et (4.) dépendent directement du choix de P . En conséquence, il peut exister une multitude de segmentations possibles pour chaque image¹. Considérons l'exemple présenté par la figure 2.3. La segmentation représentée par la figure 2.3 (b) sépare le champ non cultivé en six régions, chacune de ces régions représentant une partie du champ délimitée par des routes. À l'inverse, la segmentation représentée par la figure 2.3 (c) (obtenue par le même algorithme de segmentation, paramétré différemment) donne une région pour le champ dans sa totalité. Ces deux segmentations sont adaptées pour extraire les surfaces agricoles de l'image et les régions obtenues ont un sens potentiel pour l'utilisateur. On ne peut pas juger objectivement qu'une des segmentations est meilleure que l'autre.

1. Le nombre de segmentations différentes que l'on peut obtenir pour un ensemble donné à n éléments est borné par le n -ième nombre de Bell qui correspond au nombre de partitions différentes d'un ensemble à n éléments. Les nombres de Bell peuvent se calculer de proche en proche par la relation de récurrence suivante : $B_{n+1} = \sum_{k=0}^n \binom{n}{k} B_k$.

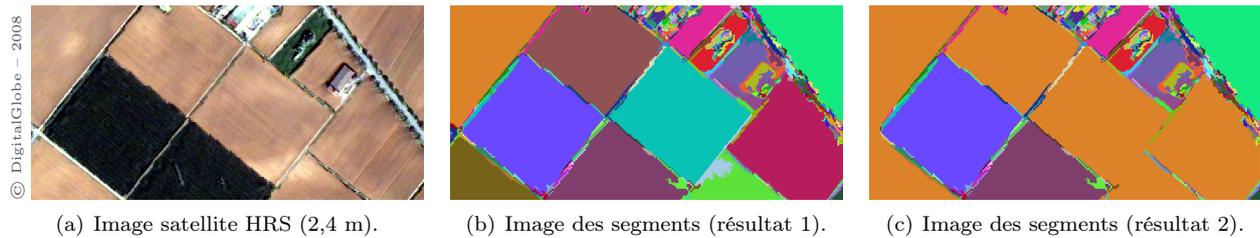


FIGURE 2.3 – Illustration du problème de segmentation. Les deux résultats de segmentation ont été obtenus avec le même algorithme de segmentation, paramétré différemment. (b) Le champ non cultivé est divisé en six segments. (c) Le champ non cultivé est formé d'un seul segment.

Le concept mathématique de problème bien posé provient d'une définition de Hadamard [Hadamard, 1902] qui postulait qu'un problème est bien posé s'il possède une solution et si cette dernière est unique. En conséquence, la segmentation est un problème complexe que l'on peut qualifier de mal posé.

Comme le précise Jonathan Weber dans sa thèse [Weber, 2011], le problème n'est donc pas lié au processus de segmentation mais à son application pour l'obtention d'une segmentation particulière, répondant aux besoins liés à un contexte applicatif spécifique. L'adaptation de la segmentation en fonction des besoins contextuels est un problème important qui sera abordé dans ce manuscrit.

C. Sur-segmentation et sous-segmentation

Le nombre de segments est un critère important pour apprécier la qualité d'une segmentation. Un nombre trop élevé ou trop faible de segments peut notamment rendre une segmentation inutilisable. Ces deux cas sont nommés respectivement sur-segmentation et sous-segmentation. L'exemple représenté par la figure 2.4 illustre ces phénomènes.

Une sur-segmentation est une segmentation comprenant un nombre trop important de segments. Les différents objets d'intérêt contenus dans l'image sont alors divisés en sous-objets ne portant pas forcément de sémantique. Dans ce cas, les segments ne sont pas significatifs et ne comportent, en particulier, pas assez de pixels pour pouvoir être utilisés en l'état (Figure 2.4 (b)). Afin d'obtenir une segmentation exploitable, on réduit généralement la sur-segmentation en fusionnant des segments adjacents similaires.

À l'inverse, une sous-segmentation est une segmentation comprenant un nombre trop faible de segments (Figure 2.4 (c)). Ce problème est plus difficile à résoudre que la sur-segmentation. Une solution consiste à re-segmenter le contenu des segments obtenus. Cependant, dans certains cas, comme par exemple quand un segment intersecte (*i.e.*, chevauche) plusieurs segments de la sous-segmentation, la solution nécessitera de décomposer ces segments pour provoquer une sur-segmentation. La résolution de cette sur-segmentation avec des fusions appropriées permettra alors d'obtenir la segmentation voulue par l'utilisateur.

Il semble ainsi préférable d'éviter ces deux phénomènes. Malheureusement, il n'existe pas à l'heure actuelle de méthode générique répondant à ces problèmes. Cependant, la sur-segmentation d'une image est généralement préférée car elle peut être employée comme pré-traitement en vue d'une segmentation. Un tel cas de figure sera abordé dans ce manuscrit.

2.1.3 Méthodologies de segmentation

Il existe un très grand nombre de méthodes de segmentation et leur développement a donné lieu à d'importantes contributions dans les domaines de l'analyse et du traitement d'images. On peut notamment regrouper ces approches en trois groupes principaux [Pal et Pal, 1993] :

- les approches par seuillage ;
- les approches basées frontières ;
- les approches par croissance.

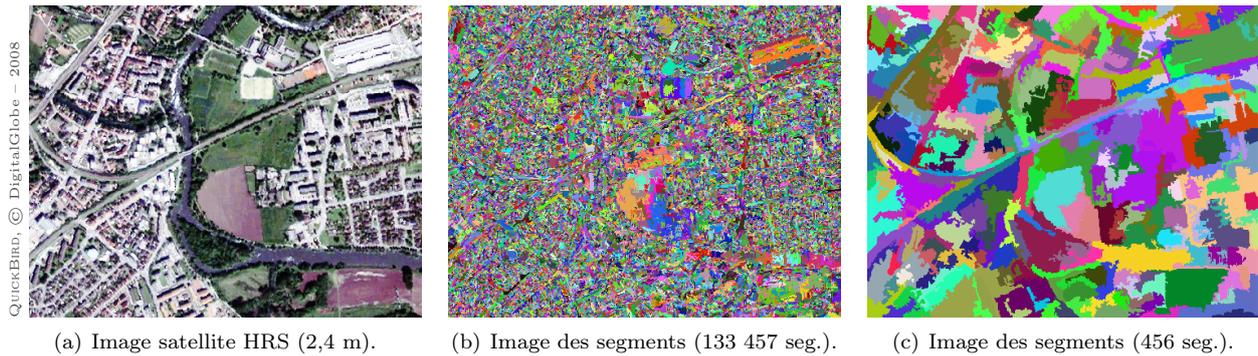


FIGURE 2.4 – Illustration des phénomènes de sur-segmentation et de sous-segmentation. (b) Résultat de sur-segmentation. (c) Résultat de sous-segmentation.

Nous présentons ci-après un bref aperçu des techniques de segmentation existantes, utilisables dans le cadre de l’analyse d’images satellitaires. La qualité de ces techniques de segmentation est un point crucial pour obtenir par la suite de bons résultats de classification.

A. Les approches par seuillage

Ces approches de segmentation définissent des règles permettant la séparation entre différentes zones de l’image selon un ou plusieurs seuils appliqués directement sur les valeurs radiométriques des données (*i.e.*, pour chaque bande spectrale disponible).

Le principe des approches par seuillage automatique repose sur l’analyse de la distribution en fréquences des niveaux radiométriques associés à l’image [Weszka et al., 1974]. L’objectif est de déceler plusieurs « populations » dans l’histogramme, et de proposer les valeurs permettant de les séparer au mieux. Pour ce faire, différentes méthodes de traitement des données peuvent être utilisées pour trouver les valeurs de ces seuils (*e.g.*, maximisation d’entropie, minimisation de la variance intra-classes). Cependant, quand l’histogramme d’une image ne présente pas de populations distinctes, la valeur de seuil obtenue automatiquement n’est généralement pas pertinente.

Dans le cadre de la segmentation d’images satellitaires, ces approches ont été largement étudiées [Zamperoni, 1992; Pal et al., 2000]. Cependant ces dernières héritent de deux inconvénients majeurs. Premièrement, elles ne considèrent pas les pixels dans leur voisinage et ne sont donc pas adaptées à extraire des objets d’intérêt caractérisés par de nombreuses propriétés d’homogénéité spatiale. Deuxièmement, ces approches ne sont pas adaptées à traiter des données à hautes résolutions spatiales, présentant un nombre très important d’objets d’intérêt différents (*i.e.*, conduisant à de nombreux niveaux radiométriques distincts et ainsi à de nombreuses populations dans l’histogramme).

B. Les approches basées frontières

Ces approches de segmentation consistent à exploiter les transitions détectables entre les segments connexes de l’image. Parmi celles-ci, l’une des plus utilisées est la ligne de partage des eaux (LPE) [Vincent et Soille, 1991] issue de la morphologie mathématique.

Le principe consiste à partager l’image en différentes zones homogènes appelées « bassins versants ». Pour ce faire on peut considérer une image à niveaux de gris comme un relief topographique (un niveau de gris correspond alors à la « hauteur » du point considéré). La « montée des eaux » consiste à immerger ce relief topographique. Un nouveau bassin est créé chaque fois que le niveau atteint un pixel formant un minimum local. Lorsque deux bassins se rencontrent, on empêche leur fusion en créant une digue, nommée ligne de partage des eaux.

Cette définition reste valable dans le cas d’images couleurs en appliquant par exemple une LPE sur chacune des bandes indépendamment puis en fusionnant les résultats [Tarabalka et al., 2010]. Les bassins versants ainsi

obtenus correspondent aux segments de la partition. Notons que l'on n'applique généralement pas la LPE sur l'image originale mais sur son gradient (obtenu par le biais de la dérivée de la fonction image).

Le principal problème de la LPE est la sur-segmentation. Ce phénomène est lié au fait que les images induisent généralement un grand nombre de minima locaux (dus par exemple au bruit du système d'acquisition, engendrant des variations locales non-significatives des niveaux de gris), ce qui génère un nombre important de bassins versants dans la LPE.

Il existe plusieurs méthodes qui permettent de pallier cette sur-segmentation. La première consiste à filtrer l'image originale afin de tenter de supprimer les minima non-significatifs. En effet, les images satellitaires sont généralement fortement texturées : elles en comportent ainsi un très grand nombre. Le problème de cette stratégie est la perte d'information occasionnée par le filtrage. La seconde consiste à choisir la position et/ou le nombre de minima locaux *a priori* (*i.e.*, approche par marqueurs (*swamping*)). On doit alors connaître à l'avance le nombre de zones que l'on souhaite mettre en évidence grâce à la LPE. Le problème majeur de cette technique réside dans le fait que la recherche de marqueurs spécifiques nécessite une connaissance *a priori* sur la position dans l'image des objets d'intérêt que l'on souhaite segmenter. Ceci est difficilement applicable dans notre cas d'étude car nous ne disposons d'aucune information *a priori* sur les caractéristiques et la position spatiale des objets à extraire. Une autre approche est proposée dans [Chen et al., 2005]. Dans leurs travaux, les auteurs utilisent une méthode par « seuillage de gradient » en combinaison de la LPE classique.

C. Les approches par croissance de segments

Ces approches de segmentation manipulent directement les contenus des segments potentiels de l'image. La plupart de ces approches partent généralement d'un ensemble de segments puis les fusionnent itérativement en suivant l'évolution d'un critère d'agrégation donné.

Les méthodes à base de croissance de segments figurent parmi les plus utilisées dans le cas d'images satellitaires [Meinel et Neubert, 2004; Carleer et al., 2005]. Le principe de ces méthodes consiste à réunir, d'une façon itérative, un ensemble de points connectés en segments de plus en plus larges, en fonction de différents critères d'homogénéité. Deux approches sont alors envisageables : la croissance à base de « graines » et la croissance par fusion de segments.

Les approches à base de graines procèdent en deux étapes. Dans une première étape, des zones réduites de l'image sont choisies pour jouer le rôle de graines. Dans une deuxième étape, ces zones sont étendues en fonction d'un critère d'homogénéité à leurs pixels voisins [Mueller et al., 2004; Couprie et al., 2011]. Cette première technique est efficace dans le cas d'images peu texturées, où le contraste entre le fond et la forme est élevé. De plus, le choix du nombre de graines à initialiser dépend du nombre d'objets d'intérêt que l'on souhaite extraire, et les positions initiales de celles-ci ne sont pas évidentes à déterminer. Cette approche n'est donc pas bien adaptée dans le cas d'images de télédétection, où les acquisitions sont souvent fortement texturées et où le nombre d'objets d'intérêt à extraire est en général élevé et inconnu à l'avance.

Le principe des approches par fusion de segments consiste à partitionner l'image en un ensemble de segments, puis à fusionner itérativement les segments voisins présentant des propriétés similaires, voire à scinder ceux qui ne sont pas suffisamment homogènes (méthode du « split and merge ») [Cross et al., 1988]. À chaque étape, la paire de segments adjacents qui minimise une fonction d'hétérogénéité est fusionnée. La majorité de ces approches utilise une structure de graphe d'adjacence de segments. Les sommets du graphe correspondent aux segments élémentaires et les arêtes aux relations de voisinage entre ces segments. À partir de ce graphe, on peut fusionner des segments élémentaires en supprimant certaines arêtes et en fusionnant les nœuds correspondants. Pour ce faire, les auteurs de [Benz et al., 2004] utilisent comme fonction d'hétérogénéité une composition entre les hétérogénéités radiométrique et géométrique des segments.

Certaines techniques hybrides [Chen et al., 2005] consistent à utiliser d'abord la LPE pour partitionner l'image, puis à fusionner les segments voisins ayant des critères communs. L'inconvénient de cette méthode réside dans le fait que si la LPE commet des erreurs, celles-ci risquent d'être propagées par le processus de fusion. Pour pallier ce problème, les auteurs de [Crespo et al., 1997] proposent une approche par fusion de zones plates. Les zones plates peuvent être vues comme une extension du concept de composantes connexes aux images non-binaires. Une zone plate représente un ensemble de pixels connexes, dont les valeurs radiométriques sont identiques. L'intérêt principal des zones plates est la précision de leurs frontières. En effet, en

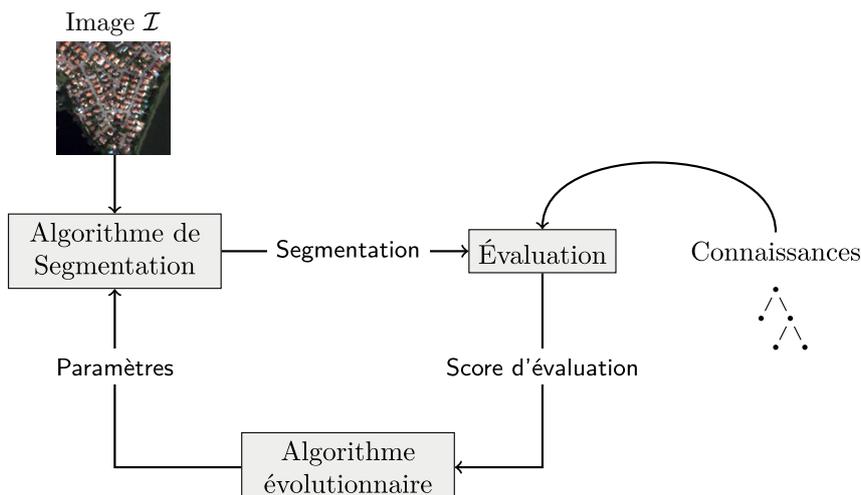


FIGURE 2.5 – Processus de segmentation guidée par les connaissances [Forestier, 2010].

partant de l’hypothèse que les frontières des objets d’intérêt, représentés par une image, sont situées entre des pixels adjacents de valeurs différentes (sauf si ces objets se touchent), on en déduit que les frontières de ces objets sont incluses dans celles des zones plates de l’image. En conséquence, les objets d’intérêt sont souvent décomposables en ensembles de zones plates. Une fusion judicieuse des zones plates permet, de ce fait, de segmenter les objets d’une image. Cependant, il est rare de trouver de vastes zones composées de pixels ayant la même valeur. Les zones plates sont ainsi souvent composées de quelques pixels seulement (en particulier dans les images multibandes). Ce phénomène induit une sur-segmentation très importante, ce qui rend les zones plates peu adaptées à la segmentation et uniquement utilisables comme partition initiale d’un processus de fusion de segments.

Afin de réduire cette sur-segmentation tout en conservant les propriétés intéressantes des zones plates, un critère de construction moins restrictif (proposé initialement dans [Nagao et al., 1979] dans le cadre d’un processus de fusion de segments) peut être employé. Ce critère conduit à la production de zones plus grandes, que l’on nomme zones quasi-plates (ZQP) [Meyer et Petros, 1999], ayant l’avantage de présenter un coût calculatoire peu élevé. Cependant les frontières des objets d’intérêt peuvent ne plus être incluses dans les frontières de ces ZQP. La construction des ZQP a été largement étudiée dans la littérature et de nombreuses extensions en ont été proposées [Soille, 2008]. Ces méthodes de segmentation ont notamment été appliquées avec succès à la segmentation d’images satellitaires [Soille, 2010]. Cependant, la taille des ZQP et la précision de leurs frontières (relativement à celles des objets d’intérêt) sont dépendantes de différents paramètres qu’il est souvent difficile d’estimer *a priori*.

Ainsi, le principal problème des approches par croissance réside dans le fait qu’elles nécessitent généralement un ou plusieurs seuils (*e.g.*, le « paramètre d’échelle » dans [Benz et al., 2004]) fournis par l’utilisateur, dans le but de stopper le processus de fusion (ou de construction) des segments. Ces paramètres sont souvent difficiles à choisir et peuvent dépendre du type et de la résolution spatiale de l’image à segmenter.

2.1.4 Ajout de connaissances dans le processus de segmentation

L’un des points communs aux approches présentées précédemment réside dans le fait qu’elles n’utilisent *a priori* que très peu de connaissances expertes sur les données. Partant du constat que la qualité de la segmentation a un fort impact sur l’identification des régions, certaines stratégies proposent d’utiliser des connaissances disponibles pour tenter de guider la segmentation. En effet, de même que pour les approches de classification supervisée (ou guidée), il est possible d’utiliser des connaissances pour identifier ou former les segments délimitant au mieux les objets d’intérêt de la scène observée.

La majorité des algorithmes de segmentation fonctionne avec des paramètres ou seuils qui conditionnent les résultats obtenus (*e.g.*, géométrie des segments obtenus, cardinal de la partition). Pour influencer sur la

construction des segments au cours de l'étape de segmentation, de nombreuses approches opèrent sur le choix de ces paramètres [Bhanu et al., 1995; Pignalberi et al., 2003]. L'idée est de les faire varier afin de trouver le jeu de paramètres qui permet d'obtenir une partition dont les segments sont identifiables grâce aux connaissances disponibles. Pour ce faire, les auteurs de [Derivaux et al., 2010] utilisent un algorithme génétique dont l'objectif est d'optimiser la qualité de la segmentation, qui est évaluée *via* des exemples d'objets segmentés manuellement. Cette approche, illustrée par la figure 2.5, est détaillée d'une manière plus approfondie dans la thèse de Germain Forestier [Forestier, 2010].

2.2 L'apport de la segmentation d'images pour la classification

Les approches basées régions consistent à appliquer une première étape de segmentation de l'image afin de partitionner celle-ci en segments pouvant être ensuite labellisés par une seconde étape de classification. Nous présentons dans cette section comment l'étape de segmentation peut permettre d'améliorer l'étape de classification.

2.2.1 Caractérisation des segments *via* des attributs pertinents

Dans les approches de classification basées régions, les méthodes ne classifient plus les pixels eux-mêmes mais les régions de l'image, obtenues par un processus de segmentation suivie d'une caractérisation des segments résultants.

Pour former ces régions, il est possible de caractériser les segments en fonction de différents types d'attributs divisés en trois catégories principales : les attributs radiométriques (*e.g.*, moyenne, écart-type, indice de texture), les attributs de formes ou géométriques (*e.g.*, surface, périmètre, élongation) ainsi que les attributs contextuels (*e.g.*, le nombre et les types de régions qui entourent la région à classifier).

Les attributs radiométriques sont dérivés des données brutes de l'image. Pour chaque segment, on calcule par exemple la moyenne et l'écart-type des valeurs radiométriques associées aux pixels composant le segment. Cette nouvelle donnée permet de fournir une information pertinente sur la composition spectrale de la région et de synthétiser les informations radiométriques attachées aux pixels. Des travaux menés au sein de l'équipe [Forestier et al., 2008] ont montré qu'il pouvait également être pertinent de caractériser un segment en fonction de son hétérogénéité radiométrique. Pour modéliser cette information, une solution consiste à stocker, pour chaque segment, la répartition des labels des clusters issus d'une classification basée pixels sous forme d'un histogramme. De cette manière, on résume les informations radiométriques tout en conservant l'information d'hétérogénéité liée aux objets géographiques représentés.

Les attributs géométriques sont eux-aussi dérivés directement de l'image. Pour chaque segment, on calcule par exemple la forme de la boîte englobante et/ou la superficie de ce dernier. Ces attributs géométriques fournissent une information pertinente sur la représentation spatiale de la donnée à classifier. Une description détaillée des différents types d'attributs géométriques pouvant être employés pour caractériser ces segments (ainsi qu'une étude relative à leur contexte d'utilisation) est proposée dans [Carleer et Wolff, 2006] ainsi que dans la thèse d'Anne Puissant [Puissant, 2003].

Les attributs contextuels peuvent être calculés après une première phase de reconnaissance des régions de l'image [Bruzzone et Carlin, 2006; Aksoy et al., 2011]. Pour chaque région, on peut alors déterminer, par exemple, le type des régions adjacentes à cette dernière (dans un voisinage proche ou plus éloigné), ce qui permet d'obtenir une information de haut niveau, relative au contexte spatial du segment considéré.

Ces nouveaux attributs permettent ainsi de compléter les simples informations radiométriques, caractérisant les pixels de l'image, généralement employées dans les approches basées pixels. Dans le cas où des objets d'intérêt de types différents ont des valeurs radiométriques similaires, ces caractéristiques (calculées au niveau des régions) peuvent permettre de séparer ces objets dans un nouvel espace de données. Considérons l'exemple d'un segment correspondant à une surface d'eau et d'un segment correspondant à l'ombre portée d'un bâtiment. Les régions liées à ces segments sont généralement caractérisées par des propriétés radiométriques similaires ne permettant pas leur différenciation. Cependant, ces régions sont différenciables par le biais de leurs propriétés géométriques ou contextuelles qui apparaissent souvent comme différentes (*e.g.*, l'ombre portée d'un bâtiment est souvent située à côté d'un bâtiment, tandis qu'une surface d'eau est souvent située

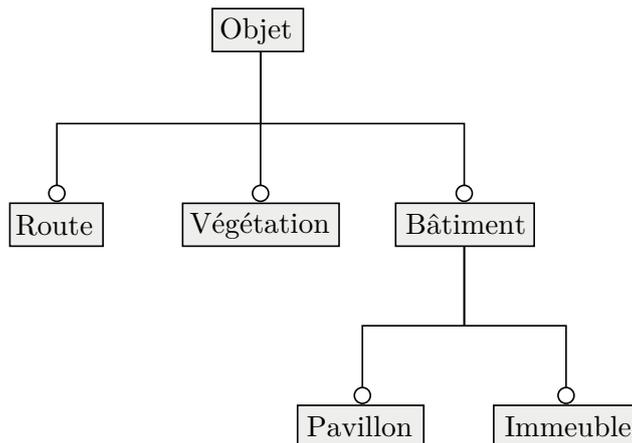


FIGURE 2.6 – Exemple de hiérarchie de six concepts urbains [Forestier, 2010].

à côté d'une zone de végétation). Une telle séparation pourra notamment permettre d'améliorer les résultats produits par les méthodes de classification.

2.2.2 Classifier des éléments de plus haut niveau sémantique

Les approches basées régions permettent d'aborder le problème de l'interprétation d'images à hautes résolutions spatiales d'une manière différente des approches basées pixels. En effet, ce nouveau cadre de raisonnement permet de faire directement le parallèle entre les objets d'intérêt présents dans l'image et les régions de cette dernière pouvant correspondre à ces objets, ce qui n'était pas possible avec les pixels. Dans les approches basées pixels, les éléments individuels à classer (*i.e.*, les pixels) ne portaient pas de sémantique « réelle » car ils ne représentaient que peu d'information individuellement (*e.g.*, une information radiométrique). La sémantique était portée par les groupes de pixels (*i.e.*, les clusters ou les classes).

Dans le paradigme régions, il devient plus aisé d'établir une sémantique sur les éléments individuels. On peut par exemple chercher à identifier les objets géographiques du type *Pavillon* qui est un concept que l'on peut directement relier à sa représentation physique dans le monde réel : un pavillon est généralement représenté dans une image à haute résolution spatiale sous la forme d'une région de forme rectangulaire, d'aire inférieure à 90 m², et ayant une réponse radiométrique proche du rouge (toit en tuiles). On nommera alors concept, une représentation générale et abstraite de la réalité.

2.2.3 Utilisation de connaissances de haut niveau

Dans un but d'automatisation du processus d'interprétation des images, les approches régions vont permettre d'utiliser plus aisément des connaissances (de haut niveau) de l'utilisateur, sans pour autant passer par la définition d'exemples d'apprentissage (*e.g.*, classification supervisée). Parmi ces approches, une stratégie consiste à modéliser les connaissances de haut niveau sous la forme d'ontologies.

Une ontologie [Gruber, 1995] est un réseau sémantique qui regroupe un ensemble de concepts visant à décrire un domaine. Ces concepts sont liés les uns aux autres par des relations taxonomiques (hiérarchisation des concepts) d'une part, et sémantiques d'autre part. Les ontologies fournissent un cadre formel pour la représentation de connaissances. Un exemple de hiérarchie de concepts est présenté par la figure 2.6. En haut se situe le concept le plus général *Objet*. Dans cet exemple, trois concepts le spécialisent : *Bâtiment*, *Route* et *Végétation*. Le concept *Bâtiment* peut lui-même se spécialiser en *Immeuble* et *Pavillon*.

Pour pouvoir utiliser les connaissances de haut niveau sémantique ainsi formalisées, il est indispensable de les compléter afin de leur faire correspondre une réalité physique dans l'image. Plusieurs types de caractéristiques peuvent être utilisés : intervalles sur des attributs quantitatifs, méta-connaissances sur les attributs discriminants du concept, attributs qualitatifs, *etc.* Par exemple, si l'on reprend l'exemple du concept *Pavillon*, les connaissances affectées à ce concept seront (non-exhaustivement) :

Concept *Pavillon*²

- Forme = rectangulaire ;
- Aire < 90 m² ;
- Radiométrie [R, G, B] = [255, 10, 15].

Une fois ces informations obtenues, les concepts présents dans la hiérarchie peuvent être utilisés pour procéder à l'identification et la classification des objets d'intérêt présents dans une image. Dans les approches basées régions, le but est alors d'affecter un concept à chacune des régions issues de la segmentation. Pour chaque région (ou cluster de régions), on recherche le concept le plus similaire à celui-ci dans l'ontologie (si celui-ci existe). Ce processus, qui consiste à chercher le ou les concepts correspondant à une région (ou à un cluster), est appelé appariement ou encore identification.

Les auteurs de [Athanasiadis et al., 2007] présentent une approche pour la segmentation et l'identification d'objets d'intérêt utilisant une ontologie dans le domaine du multimédia. Une autre stratégie, introduite dans [Ogier et al., 2000], consiste à évaluer les objets mal identifiés par l'étape d'appariement. Dans le domaine de la télédétection, ce genre d'approches a aussi été utilisé avec succès. Dans [Forestier et al., 2012], les auteurs proposent un nouveau score d'appariement basé sur la similarité entre les valeurs des attributs pour une région donnée et l'espace de définition de cet attribut pour le concept donné. Ce score est composé d'une similarité locale (*i.e.*, correspondance d'une région avec un concept donné) et d'une similarité globale (*i.e.*, correspondance d'une région avec une hiérarchie de concepts). Des recherches sont aussi actuellement en cours pour permettre d'affecter aux concepts de ces ontologies, des connaissances (sous la forme de propriétés) relatives aux relations spatiales (*e.g.*, le concept *Pavillon* peut être caractérisé par la propriété : souvent adjacent à une parcelle de végétation) et/ou de composition (*e.g.*, le concept *Pavillon* peut entrer dans la composition du concept *Bloc urbain d'habitations individuelles*) pouvant relier ces concepts dans le but de mieux les caractériser.

Trois problèmes principaux sont cependant inhérents aux approches utilisant des connaissances de haut niveau :

1. ces approches nécessitent que l'utilisateur ait une forte connaissance *a priori* sur les différents objets d'intérêt identifiables au sein des images étudiées, ce qui n'est pas forcément le cas dans les environnements urbains complexes ;
2. il est souvent très difficile pour l'utilisateur de définir et de formaliser des connaissances relatives à des objets urbains complexes : à titre d'exemple, alors qu'il est plus ou moins aisé pour l'utilisateur de formaliser des connaissances quant au concept *Pavillon*, la formalisation de connaissances pour le concept *Bloc urbain industriel* n'est pas une tâche évidente. Deux blocs urbains industriels peuvent apparaître comme très différent dans une même image ;
3. l'instanciation de l'ontologie du domaine n'est bien souvent valable que pour un seul type d'images, dû aux différences de résolutions spatiales et radiométriques des données.

En conséquence, nous préférons dans ces recherches nous limiter à employer des approches basées régions, plus automatiques, proposant à l'utilisateur des clusters d'intérêt non labellisés qu'il pourra ensuite étiqueter ou non, suivant ses connaissances du domaine. Cependant, un objectif à plus long terme sera de combiner les travaux proposés au cours de cette thèse aux solutions futures, proposées par l'équipe, relatives à l'utilisation de connaissances de haut niveau.

2.3 Limites des approches basées régions

Dans la section précédente nous avons introduit les approches basées régions comme une solution adaptée aux problèmes liés à l'analyse d'images HRS. Nous montrons par la suite que ces approches souffrent néanmoins de limites paramétriques et qu'elles peuvent s'avérer insuffisantes pour l'extraction d'objets d'intérêt complexes, en particulier à partir d'images à très hautes résolutions spatiales (THRS). Après avoir décrit les limites de ces approches, nous évoquons leur adaptation à l'analyse d'images multisources.

2. Ce concept est extrait du dictionnaire des données décrit dans [Forestier et al., 2012].

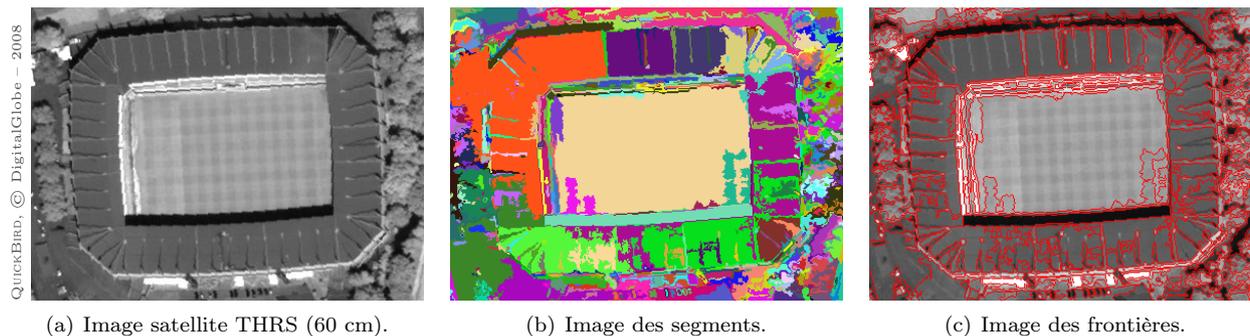


FIGURE 2.7 – Illustration d’une erreur de segmentation.

2.3.1 Un aspect paramétrique important

L’utilisation d’approches basées régions pour l’analyse d’images HRS pose certains problèmes. La construction des segments de l’image va en effet induire une étape paramétrique supplémentaire pour l’utilisateur. De plus, les algorithmes de segmentation utilisés peuvent commettre des erreurs dans la création de ces segments, qui pourront être répercutées sur les résultats de classification. Finalement, la caractérisation des segments produits n’est pas une étape triviale. Nous détaillons chacun de ces problèmes ci-après.

A. Choix de l’algorithme et des paramètres de segmentation

Pour utiliser un algorithme de classification, l’utilisateur est confronté à la configuration de ses paramètres (*e.g.*, nombre de clusters, nombre d’itérations), influant alors de manière plus au moins importante sur les résultats obtenus. De tels paramètres sont nécessaires afin que l’algorithme soit suffisamment générique pour être applicable sur des jeux de données variés. La plupart du temps, le choix des valeurs de ces paramètres est effectué en fonction des connaissances de l’utilisateur relatives aux données (*e.g.*, nombre de classes thématiques recherchées, résolutions des images) et aux algorithmes utilisés (*e.g.*, l’algorithme de classification K -MEANS converge, en moyenne, au bout de 10 itérations).

Pour utiliser une approche basée régions, l’utilisateur est de plus confronté au choix de l’algorithme de segmentation, et de ses paramètres respectifs. Par ailleurs, la paramétrisation d’un algorithme de segmentation est souvent moins intuitive que celle d’un algorithme de classification. Considérons par exemple, un algorithme par fusion de segments couramment utilisé dans le cadre de l’analyse d’images de télédétection [Benz et al., 2004]. Cet algorithme nécessite un seuil (« paramètre d’échelle »), fourni par l’utilisateur, destiné à stopper le processus de fusion (ou de construction) des régions. Ce seuil est difficile à choisir et dépend du type et de la résolution spatiale de l’image à segmenter. Généralement, les utilisateurs procèdent par un mécanisme d’essais-erreurs jusqu’à obtenir le partitionnement souhaité. Ce dernier est souvent lié à la classe thématique recherchée.

B. Influence des erreurs de segmentation sur la classification

Il est important de noter que la construction des segments n’est pas une étape triviale. En effet, il n’est généralement pas possible d’obtenir une segmentation parfaite des objets d’intérêt de la scène : des pixels correspondant à certains objets géographiques peuvent être inclus dans le segment associé à un autre objet ; réciproquement, des pixels correspondant à l’objet considéré peuvent avoir été omis. Enfin, certains segments peuvent ne correspondre à aucun objet géographique de la scène observée.

Ces différentes erreurs altèrent fortement le processus de classification. Sur l’exemple de la figure 2.7, un stade de sport est représenté. Le processus de segmentation utilisé a agrégé des pixels de végétation au segment correspondant au toit du stade et réciproquement (*i.e.*, segment orange), produisant ainsi deux segments erronés. Cette erreur entraînera de façon évidente un défaut dans le processus de classification qui ne pourra alors traiter correctement aucune des deux régions résultantes.

C. Caractérisation des segments

L'association, à chaque segment, d'une description en termes de caractéristiques radiométriques, géométriques ou encore contextuelles, pose le problème de la pertinence et de la redondance de ces descripteurs.

L'étude réalisée dans la thèse d'Anne Puissant [Puissant, 2003] illustre ce problème. Dans ces travaux, l'auteur a effectué une analyse de corrélation pour évaluer la redondance d'information entre différents attributs géométriques (*e.g.*, surface, périmètre, élongation). L'objectif était alors d'estimer la pertinence de ces attributs pour la caractérisation de certains types de segments (correspondant à différents types d'objets d'intérêt), à partir de données existantes sur des éléments urbains. Ces travaux ont montré qu'il est difficile de généraliser des connaissances quant à la pertinence ou non de ces différents attributs. En effet, on peut citer les conclusions de l'auteur : « *En résumé, le choix d'un critère géométrique pertinent dépend d'abord du type d'objets recherché ; si un critère de surface peut suffire pour identifier des objets de types surface en eau et zone de végétation, un indice de forme est plus adapté pour les bâtiments ou la végétation intra-urbaine aux formes spécifiques.* »

Le choix des attributs à retenir lors de la caractérisation des segments reste encore un problème ouvert. De plus, il est évident que la qualité de ces descripteurs dépend directement de celle de la segmentation. Dans notre cas, l'utilisateur possède *a priori* peu de connaissances sur les objets d'intérêt représentés dans l'image. Il lui est donc difficile de faire une sélection fine des indices à utiliser.

2.3.2 Des objets d'intérêt de plus en plus complexes

Face à la grande complexité des objets d'intérêt pouvant être analysés dans les environnements urbains à partir d'images à hautes résolutions spatiales, les approches basées régions présentent certaines limites.

A. Des images à très hautes résolutions spatiales

Ces vingt dernières années ont été marquées, dans de nombreux domaines de l'imagerie, par un saut technologique induit par l'arrivée des images à très hautes résolutions spatiales (THRS) conduisant à une augmentation importante des détails présents dans ces images. Les données à traiter sont devenues, en conséquence, de plus en plus complexes.

En imagerie satellitaire, l'arrivée d'une nouvelle génération de capteurs (*e.g.*, satellites QUICKBIRD, IKONOS, et maintenant PLÉIADES) de résolution submétrique a conduit à la production de telles images à très hautes résolutions spatiales, améliorant ainsi l'aptitude à différencier des objets géographiques en fonction de leurs matériaux et de leurs géométries [Donoghue, 2000]. Pour illustrer ce phénomène, la figure 2.8 présente un extrait d'une image THRS issue du satellite QUICKBIRD (© DigitalGlobe Inc.).

Le traitement des masses de données engendrées par ces nouvelles images représente un véritable défi technologique pour le domaine de l'analyse d'images. Par exemple, une image satellite THRS peut être composée de plus de $20\,000 \times 20\,000$ pixels. Une telle quantité de données rend difficilement applicables, dans un contexte opérationnel, les méthodologies basées régions classiquement employées, car ces dernières peuvent nécessiter une étape de segmentation parfois coûteuse en ressources de calcul.

De plus, la réalisation de nouveaux projets liés aux mosaïquages d'images contribue à amplifier ce phénomène dans les prochaines années. En effet, dans le domaine de l'observation de la Terre, de nombreux projets [Soille, 2006] ont pour objectif de regrouper des grands ensembles d'images THRS dans le but d'obtenir des mosaïques d'images. Ces mosaïques peuvent ainsi permettre d'obtenir une couverture des sols à une échelle globale (*e.g.*, tout un continent). Les quantités de données résultant de tels ensembles d'images s'avèrent alors difficilement exploitables par des approches basées régions classiques. Bien que certaines approches aient été proposées pour analyser des images dites *GigaPixels* [Gueguen et al., 2011], le traitement de telles quantités d'informations demeure encore un problème difficile à résoudre.

B. Différents niveaux d'analyse

En télédétection, et en particulier dans le domaine de la cartographie urbaine, les utilisateurs, experts en photo-interprétation, utilisent les images pour analyser le territoire à plusieurs niveaux d'échelle dans le but d'extraire différents types d'objets d'intérêt (*e.g.*, quartiers, blocs urbains, objets urbains individuels). Ces



FIGURE 2.8 – Bloc urbain délimité en rouge sur une image THRS (60 cm).

différents niveaux d'échelle conduisent à différents niveaux d'analyse d'une scène permettant une meilleure compréhension de cette dernière. La figure 2.9 illustre un résultat d'analyse à deux niveaux : le premier correspond à une analyse de l'image au niveau des quartiers urbains, et le second à une analyse de la scène au niveau des blocs urbains.

De plus, les objets d'intérêt relatifs à ces différents niveaux d'analyse sont souvent organisés de façon hiérarchique. Par exemple, un quartier urbain se décompose en général en un ensemble de blocs urbains de divers types pouvant eux-mêmes se décomposer en ensembles d'objets urbains simples, formant alors une hiérarchie multiniveau. Plus un niveau d'analyse est élevé au sein de cette hiérarchie, plus le niveau sémantique des objets d'intérêt y étant associés est complexe. Les approches basées régions classiques étant de nature mononiveau, elles ne permettent pas d'extraire simultanément ces différents niveaux d'objets d'intérêt.

C. Des hiérarchies d'objets d'intérêt hétérogènes

La forte croissance de la résolution spatiale des données a conduit à l'augmentation du niveau de détail des objets d'intérêt observables dans les images. Ces objets d'intérêt deviennent ainsi de plus en plus complexes et hétérogènes, entraînant alors des perturbations dans l'analyse.

D'une part, les objets simples (*e.g.*, maisons, arbres, routes) qui apparaissaient comme globalement homogènes dans les images HRS apparaissent maintenant comme hétérogènes dans des images THRS, où ils sont analysables *via* leurs matériaux (*e.g.*, toits oranges ou gris, différents types d'asphaltes, différents types de végétation). En particulier, ces objets d'intérêt simples couvrent souvent, dans ces nouvelles images, une zone de l'image composée de pixels très hétérogènes (tant radiométriquement que spatialement).

D'autre part, les objets de plus hauts niveaux sémantiques (*e.g.*, blocs urbains, quartiers urbains) apparaissent maintenant dans ces images comme des objets composites. Ils peuvent alors être qualifiés d'objets complexes car ils sont composés d'ensembles d'objets simples organisés d'une façon structurée. La figure 2.8 illustre ce phénomène. Un bloc urbain est délimité en rouge sur une image THRS dont la résolution spatiale est de 60 cm. On peut observer que cette structure urbaine est composée d'un ensemble de toits d'habitations individuelles ainsi que de portions de végétation et de routes.

Devant cette importante hétérogénéité, la précision des résultats produits par les approches basées régions classiques devient insuffisante pour une utilisation de ces résultats dans le cadre de routines opérationnelles. De telles approches, fondées sur des critères et des hypothèses d'homogénéité radiométrique et spatiale, sont adaptées à l'extraction d'objets simples car elles supportent un certain niveau d'hétérogénéité dans ces objets d'intérêt. En revanche, elles deviennent difficilement applicables pour l'extraction d'objets plus complexes à partir d'images THRS, dans lesquelles ces derniers apparaissent comme très hétérogènes. De plus, ces

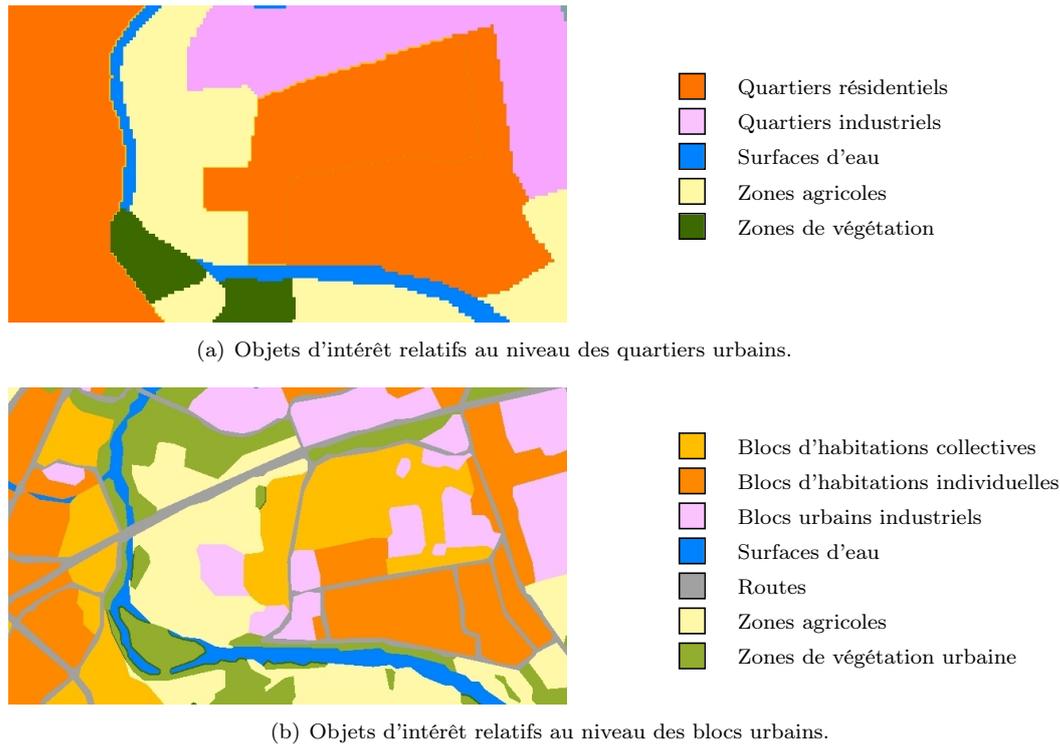


FIGURE 2.9 – Résultat d'analyse multiniveau de la scène urbaine représentée par la figure 2.8.

approches ne sont pas appropriées pour l'extraction simultanée de plusieurs niveaux d'objets d'intérêt.

En conséquence, pour traiter l'intégralité des objets d'intérêt pouvant apparaître dans des images HRS et THRS, il devient nécessaire d'employer des approches permettant d'extraire ces différents niveaux d'objets complexes tout en considérant leurs structures hiérarchiques.

2.3.3 Vers des approches multiniveaux

Face à ces différents problèmes, de nouvelles méthodologies ont été proposées afin d'extraire des hiérarchies d'objets complexes à partir d'images à (très) hautes résolutions spatiales. Pour obtenir un résultat multiniveau en utilisant une approche d'analyse basée régions, deux stratégies sont envisageables :

1. utiliser un algorithme de segmentation mononiveau classique, puis utiliser une approche de classification hiérarchique permettant d'obtenir différents niveaux de catégorisation ;
2. utiliser un algorithme de segmentation fournissant un résultat de segmentation multiniveau, puis classer les régions relatives à ces différents niveaux *via* un algorithme de classification.

Nous décrivons par la suite les avantages et inconvénients respectifs de ces deux stratégies.

A. Approches par classification hiérarchique

Dans le premier type d'approches, l'étape initiale consiste à utiliser un algorithme de segmentation mononiveau pour former une partition fine de l'image (*i.e.*, une sur-segmentation), permettant ainsi une extraction des objets d'intérêt les plus basiques (*i.e.*, correspondant à des sous-parties des objets d'intérêt recherchés). Ensuite, un algorithme de classification permettant la création de différents niveaux de labellisation est employé dans le but d'obtenir des résultats de classification d'un plus haut niveau sémantique. Ainsi, dans le cadre de la classification non-supervisée, une première approche est basée sur l'utilisation d'un algorithme de classification hiérarchique. Ce type d'algorithme a la particularité de fournir un résultat dans lequel un cluster peut être décomposé en plusieurs sous-clusters, l'ensemble des clusters formant alors une hiérarchie (sous la forme d'un arbre ou d'un dendrogramme).

Parmi les approches de classification hiérarchique, la plus couramment utilisée repose sur l'algorithme AHC (Ascendant Hierarchical Clustering) [Ward, 1963]. Cet algorithme procède par des fusions successives de clusters dans le but de créer des groupes de clusters plus importants. Les clusters initiaux peuvent être dans ce cas précis, les régions initiales issues d'une segmentation suivie d'une caractérisation (chaque région est alors considérée comme un cluster initial), ou encore des clusters de régions issus d'une pré-étape de classification. À chaque étape, les deux clusters les plus similaires, en termes de caractéristiques choisies, sont fusionnés. Le processus s'arrête lorsque les deux clusters restants fusionnent dans l'unique cluster contenant toutes les données.

Ce type d'approches permet de classifier les données à différents niveaux de granularité. À partir de l'arbre ou du dendrogramme obtenu, on peut alors extraire différents niveaux de classification, autorisant ainsi différents niveaux d'analyse. Cependant, les clusters formés dans les plus hauts niveaux de la hiérarchie ne permettent généralement pas de regrouper, dans une même catégorie, l'ensemble des régions basiques formant les objets composites (*e.g.*, former un cluster regroupant toutes les régions basiques d'un bloc urbain). Les clusters formés dans les plus hauts niveaux de la hiérarchie correspondent bien souvent à des clusters moins spécifiques et plus généraux que les clusters initiaux (*e.g.*, un cluster regroupant les régions correspondant à des zones de végétation urbaines et non-urbaines). En effet, les regroupements effectués sont généralement réalisés en considérant uniquement les caractéristiques radiométriques et géométriques des régions qui ne sont généralement pas suffisantes pour inférer de tels objets complexes.

B. Approches par segmentation multiniveau

Une solution pour extraire les objets complexes des images consiste à utiliser des approches prenant en compte le voisinage des régions basiques. L'objectif est alors de reconstruire les segments/régions correspondant aux objets complexes avant de les classifier. Pour ce faire, il est possible d'utiliser un algorithme de segmentation fournissant un résultat de segmentation multiniveau, puis de classifier ces différents niveaux *via* un algorithme de classification.

Ce type d'approches a été particulièrement étudié ces dernières années, donnant lieu à l'élaboration de méthodes hybrides par groupement ou encore de méthodes de segmentation hiérarchique.

Approches par groupement basées sur les relations spatiales Lorsque l'on analyse visuellement une image satellite (T)HRS, on peut remarquer que différents objets d'intérêt simples (*e.g.*, maisons, routes) peuvent être regroupés pour former de nouveaux types d'objets d'intérêt (*e.g.*, blocs urbains, quartiers) permettant ainsi différents niveaux d'analyse. Ce mécanisme de groupement intervient naturellement lors des premières étapes du processus de perception visuelle, et se base sur les différentes relations spatiales entre les objets considérés [Marr, 1982].

Pour tenter d'automatiser ce processus de groupement, de nombreuses approches ont été proposées. Le principe est d'essayer de construire directement les segments/régions correspondant aux objets complexes de l'image étudiée avant de les classifier. À partir d'une partition initiale de l'image (assez fine pour que les segments correspondent à des objets d'intérêt élémentaires), un processus spécifique est appliqué afin de fusionner des segments voisins dans le but de construire des segments pouvant représenter des objets d'intérêt d'un plus haut niveau sémantique.

Un principe commun à de nombreuses approches par groupement, consiste à maximiser les propriétés d'homogénéité radiométrique des segments correspondant aux objets complexes recherchés (*e.g.*, fusionner des segments basiques jusqu'à ce que ces fusions n'aient plus de cohérence radiométrique). Pour ce faire, des approches permettant de regrouper plusieurs segments basiques en utilisant, par exemple, des structures de graphes, ont été proposées [Guigues et al., 2003]. Dans ces travaux, des sous-graphes remarquables appelés *cocons* sont construits et analysés suivant leurs propriétés radiométriques. Ces sous-graphes particuliers peuvent alors représenter les objets d'intérêt complexes recherchés. D'autres approches recherchent, au sein des graphes relatifs aux images, des relations spatiales fréquentes entre les segments [Barnsley et Barr, 1997; Dogrusoz et Aksoy, 2007; Zamalieva et al., 2009; Guo et al., 2009; Aksoy et Cinbis, 2010]. Elles font alors l'hypothèse que ces relations fréquentes caractérisent des segments correspondant à des objets d'intérêt particuliers. Certaines approches exploitent également les propriétés texturales des segments. Par exemple, les

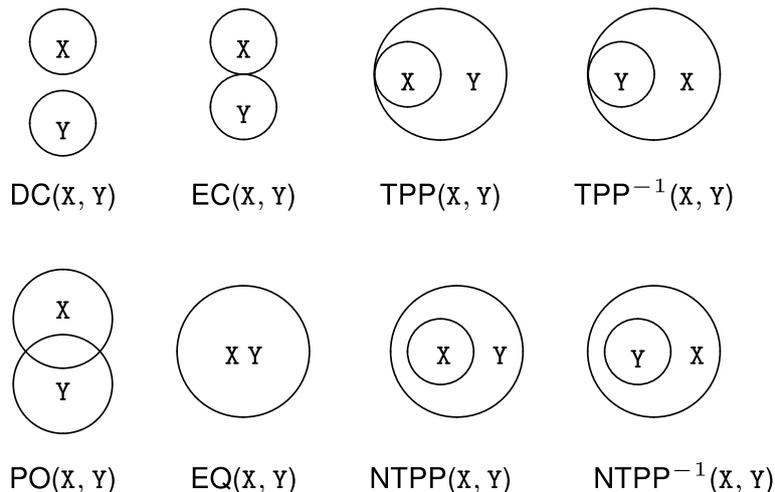


FIGURE 2.10 – Relations spatiales du système RCC8. Ces dernières sont employées pour caractériser les positions relatives des régions de l'image dans l'espace de définition considéré.

auteurs de [Gaetano et al., 2009] proposent une méthode par groupement basée sur le contexte spatial des segments basiques. Les textures sont alors représentées par des chaînes de Markov modélisant les relations « de passage » dans l'image, entre les différentes catégories de segments (*e.g.*, eau, végétation, toits d'habitations). Les catégories correspondant aux objets complexes recherchés sont finalement induites par ces différentes textures. Cependant, ces approches sont coûteuses en temps de calcul, et leur capacité à extraire des objets d'intérêt complexes est directement liée à la qualité des partitions initiales des images.

Une autre stratégie pour automatiser ce processus de regroupement consiste à utiliser des connaissances du domaine pour former des segments/régions correspondant aux objets complexes recherchés au sein des images. Généralement ces connaissances sont basées sur des règles de fusion spécifiques définissant des priorités relatives aux segments basiques à regrouper. Les auteurs de [Inglada et Michel, 2009] proposent une approche par graphe où l'image est décrite par un ensemble de segments primitifs reliés entre eux par des relations spatiales caractérisées *via* le système RCC8 (Region Connection Calculus 8). La figure 2.10 illustre ce système de raisonnement spatial composé de huit relations possibles entre segments. Une fois ces relations établies, il devient possible de définir des exemples de fusion conduisant à la formation de régions correspondant à des objets complexes, puis de rechercher dans le graphe l'ensemble des sous-graphes similaires à ces exemples.

D'autres approches [Belarte et Wemmert, 2012], emploient des connaissances de haut niveau relatives au regroupement des segments basiques. Ces connaissances sont basées sur des règles de fusion issues d'ontologies du domaine. Par exemple le concept *Pavillon* peut être décrit dans une ontologie par des relations de voisinage « fréquentes » : un pavillon est situé en général à côté d'une parcelle de végétation et/ou d'une route. De telles règles peuvent ensuite être utilisées pour inférer des segments correspondant à des objets composites à partir de segments primitifs/basiques.

Cependant, dû à la grande complexité des données, ces connaissances et règles de fusion sont difficiles à obtenir. L'utilisateur est ainsi confronté à des images présentant un très grand nombre de classes induites par les multiples détails présents à ce niveau de résolution (*e.g.*, arbres, différents types de toits). Le nombre important de concepts liés à ces classes rend alors particulièrement difficile l'identification de ces règles. De plus, ces dernières sont très spécifiques et peuvent varier entre les différents types de scènes analysées (*e.g.*, des règles valides pour un environnement urbain en Alsace ne le seront plus forcément aux États-Unis).

En conclusion, les approches par groupement sont pertinentes pour l'extraction d'un premier niveau d'objets complexes (*e.g.*, bâtiments complexes) mais semblent difficilement utilisables pour extraire des structures de plus hauts niveaux sémantiques (*e.g.*, quartiers, blocs urbains). De plus, le caractère mononiveau de ces approches n'autorise pas l'extraction directe de hiérarchies d'objets complexes, permettant une analyse simultanée de l'ensemble des objets d'intérêt pouvant être considérés dans une scène urbaine. Ainsi, dans le but d'extraire directement des hiérarchies d'objets d'intérêt, des approches hiérarchiques ont été proposées.

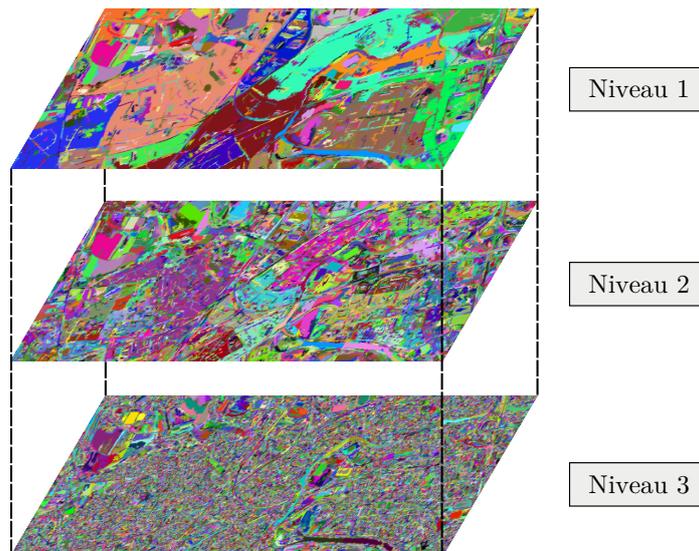


FIGURE 2.11 – Exemple de segmentation hiérarchique réalisée sur l'image représentée par la figure 2.8. Dans la hiérarchie de partitions présentée, les trois partitions obtenues autorisent trois niveaux d'analyse distincts.

Segmentation hiérarchique Les méthodes de segmentation hiérarchique fournissent une série de partitions à partir d'une image comportant des niveaux de détail distincts (croissants ou décroissants), permettant d'extraire simultanément des segments de différentes tailles (*i.e.*, à différentes échelles) pouvant correspondre à différentes catégories d'objets d'intérêt. Pour illustrer ce type de méthodes, la figure 2.11 présente le résultat d'un algorithme de segmentation hiérarchique appliqué sur une image HSR. Trois niveaux de la hiérarchie de partitions, conduisant à trois échelles différentes, y sont visualisés. Ces dernières décennies, ce type de méthodes a été largement étudié. On peut notamment citer [Pietikainen et Rosenfeld, 1981] comme un exemple de travail pionnier.

En télédétection, et plus particulièrement dans le domaine de l'analyse d'images THRS, de nombreuses approches ont été proposées. Les auteurs de [Pesaresi et Benediktsson, 2001] proposent d'utiliser des compositions d'opérations d'ouvertures et de fermetures morphologiques basées sur des éléments structurants de tailles croissantes afin de générer des profils morphologiques pour chaque pixel, permettant ainsi leur caractérisation. On trouve de nombreuses approches basées sur ce principe dans la littérature [Benediktsson et al., 2003, 2005]. Dans [Gueguen et al., 2010], une approche hybride est présentée. Les pixels de l'image sont dans un premier temps caractérisés par des profils morphologiques générés par des éléments structurants de tailles croissantes, formant ainsi un nouvel espace des données. Il devient alors possible de partitionner ce nouvel espace en utilisant une méthode de segmentation. Ces méthodes hiérarchiques basées pixels ont ainsi montré des résultats encourageants. Le problème de ce type d'approches réside dans le fait qu'elles ne fournissent pas de correspondance entre les différents niveaux des hiérarchies obtenues. Un segment extrait d'un niveau particulier ne se décomposera généralement pas en sous-segments dans les niveaux inférieurs.

Une hiérarchie de segmentations basée régions est un ensemble de segmentations à différents niveaux de détail, dans laquelle une segmentation à un certain niveau peut être produite par la fusion de segments appartenant à des niveaux plus fins. Un exemple de ce type de hiérarchie est illustré en figure 2.12. Ces hiérarchies peuvent être construites de deux manières opposées.

Dans les approches descendantes (*top-down*), les segmentations les plus grossières sont d'abord produites, puis transformées en segmentations plus fines par le biais de processus spécifiques de décomposition. Dans [Shi et Malik, 2000], la segmentation est traitée comme un problème de partitionnement de graphe. Le processus est initialisé à partir d'une segmentation grossière, puis affine progressivement les segments. Cependant, ces approches font souvent l'hypothèse que les images traitées ne contiennent que peu d'objets d'intérêt et ne sont donc pas adaptées à traiter des images THRS dans lesquelles le nombre d'objets extractibles est très important. Les auteurs de [Goffe et al., 2011] proposent d'utiliser un schéma descendant de construction de

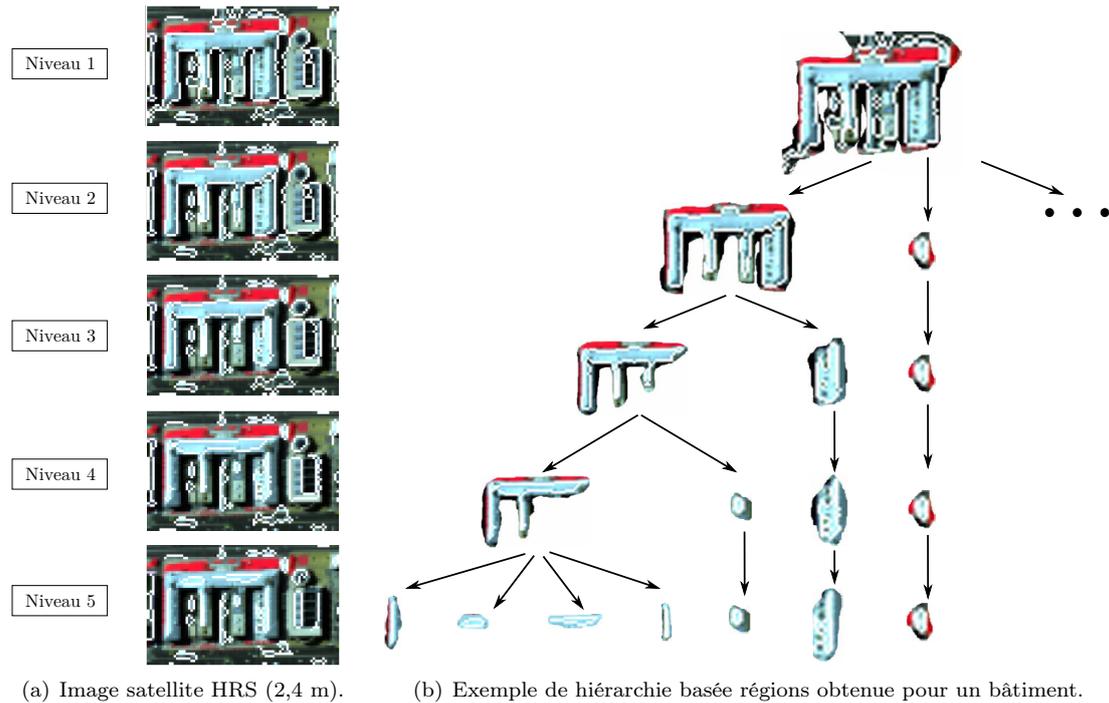


FIGURE 2.12 – Exemple de segmentation hiérarchique basée régions [Akçay et Aksoy, 2008].

pyramides irrégulières. Partant d'une carte topologique initiale, les segments sont successivement raffinés par des opérations de scission. Cependant, trouver une fonction de séparation pertinente (et robuste) demeure une question ouverte.

Les approches les plus couramment utilisées sont celles reposant sur un principe ascendant (*bottom-up*). Dans ces approches, les segmentations les plus fines sont d'abord produites. Les couples de segments adjacents les plus similaires (relativement à des propriétés particulières) sont ensuite fusionnés pour former des segments plus grossiers pouvant correspondre à des objets d'intérêt plus complexes. Les méthodes par « fusion de segments » ou « split and merge » (évoquées précédemment, voir Section 2.1.3) peuvent ainsi être étendues pour conduire à des segmentations hiérarchiques. En sauvegardant l'historique des fusions relatives à la segmentation d'une image, on peut, par le biais de ces méthodes, obtenir des hiérarchies de partitions. En analyse d'images de télédétection, différents algorithmes sont basés sur ce principe [Baatz et Schape, 2000; Scarpa et al., 2007]. Les auteurs de [Tilton, 2003] proposent notamment un algorithme de segmentation hiérarchique combinant une approche de classification non-supervisée (réalisée sur les valeurs radiométriques des pixels de l'image) à un algorithme itératif par croissance de segments.

Les auteurs de [Akçay et al., 2010] proposent une solution hybride. Dans une première étape, l'image est sur-segmentée par une méthode classique de segmentation. Ensuite, les transitions entre les segments adjacents sont caractérisées et classifiées *via* une méthode de classification non-supervisée. Finalement, par le biais d'un processus itératif, les régions adjacentes appartenant aux clusters les mieux représentés (dans l'espace des données) sont fusionnées pour obtenir les différents niveaux de la hiérarchie.

Dans [Beaulieu et Goldberg, 1989], la segmentation est réalisée par le biais d'un algorithme de fusion de segments. L'originalité de ces travaux est de déterminer l'ordre des fusions par le biais d'un processus d'optimisation. Une solution alternative est proposée dans [Gaetano et al., 2009]. Dans cette approche, le but est de détecter des objets d'intérêt complexes en utilisant un modèle de chaînes de Markov multiples. L'image est considérée comme une collection de textures complexes, émergeant à différents niveaux d'observation. Le processus de fusion exploite les propriétés texturales de l'image, couplées à des propriétés spatiales et radiométriques, dans le but de reconnaître des clusters de régions complexes. Cependant la plupart de ces approches cherchent à extraire des segments correspondant à des objets d'intérêt texturés, ce qui n'est pas nécessairement pertinent lorsque les objets recherchés sont formés d'autres éléments fortement hétérogènes.

En morphologie mathématique, les opérateurs connexes [Serra et Salembier, 1993; Salembier et Wilkinson, 2009] peuvent être utilisés pour former des segmentations hiérarchiques en utilisant, par exemple, des structures à base d'arbres. Des notions comme les arbres de type « component-tree » [Salembier et al., 1998] ou « level-lines tree » [Monasse et Guichard, 2000] peuvent permettre de réaliser des segmentations hiérarchiques à base de fusion de zones plates. Les feuilles des arbres correspondent alors aux segments basiques de l'image tandis que les nœuds correspondent aux segments issus des fusions de ces régions élémentaires. Toutefois, ces structures reposent principalement sur les valeurs extrêmes des intensités des images. Ainsi, les résultats obtenus ne sont pas forcément pertinents dans le cas d'images satellites.

Par opposition, l'arbre binaire de partitions (BPT – Binary Partition Tree) [Salembier et Garrido, 2000] reflète une mesure de similarité entre des segments voisins, et modélise une hiérarchie entre ces segments *via* une structure d'arbre. Le BPT représente un ensemble de segments à différentes échelles et ses nœuds fournissent une estimation (en fonction d'une mesure choisie) des objets d'intérêt de la scène [Vilaplana et al., 2008]. Ce type de structures a récemment été étudié dans le cadre de la segmentation d'images satellites hyperspectrales [Valero et al., 2010b] et les résultats obtenus se sont montrés prometteurs.

Une autre approche, étendant les ZQP et basée sur le paradigme de connectivité contrainte, a récemment été introduite [Soille, 2008] et appliquée pour traiter des images (T)HRS dans [Soille, 2010]. Dans cette approche, la relation de connectivité permet de générer une partition du domaine de définition de l'image. Les hiérarchies de partitions sont alors produites en faisant varier la valeur de seuil associée à chaque contrainte de connectivité. Une amélioration de ces approches traitant le problème de l'effet de « chaînage » a récemment été introduite dans [Soille, 2011].

Finalement, une dernière approche issue de la morphologie mathématique consiste à appliquer un filtre connexe pour filtrer une image puis à extraire une ligne de partage des eaux (LPE) de l'image résultante. En faisant croître le paramètre du filtre connexe, on peut alors obtenir une hiérarchie de segmentations. De nombreux travaux décrits dans la littérature [Cousty et al., 2008, 2009; Najman, 2011] reposent sur ce principe hiérarchique.

Limites des approches hiérarchiques Les approches de segmentation hiérarchique, et en particulier celles basées sur les opérateurs connexes, fournissent des résultats encourageants pour le traitement d'images (T)HRS. Parmi ces approches, on distingue celles basées sur les BPT qui, après le choix d'une mesure de similarité entre nœuds, ne nécessitent pas la définition de paramètres de segmentation complexes. Les niveaux les plus bas des hiérarchies liées à ces arbres peuvent généralement être utilisés pour l'extraction de segments correspondant aux objets d'intérêt les plus simples (*e.g.*, maisons, arbres, routes) tandis que les niveaux intermédiaires peuvent être utilisés pour l'extraction de segments correspondant à des objets complexes faiblement hétérogènes (*e.g.*, bâtiments complexes). En revanche, il semble difficile d'exploiter les plus hauts niveaux de ces hiérarchies pour extraire directement des segments représentant des objets d'intérêt fortement complexes (*e.g.*, quartiers, blocs urbains). En effet, dans les images à (très) hautes résolutions spatiales, les objets complexes recherchés apparaissent comme trop hétérogènes pour être extraits par ces méthodes qui sont pour la plupart basées sur des critères d'homogénéité radiométrique ou texturale.

Pour apporter une solution à ces problèmes, des travaux récents ont montré qu'il est possible d'adapter ces méthodes au traitement de données multisources/multirésolutions [Goffe et al., 2011]. La complémentarité des informations apportées par de telles données peut permettre de pallier le problème d'hétérogénéité radiométrique relatif à l'utilisation d'approches hiérarchiques sur des images (T)HRS. Le prochain chapitre de ce manuscrit s'intéresse à l'apport et à l'intérêt des images multisources pour les approches hiérarchiques basées régions.

Un autre problème relatif à la segmentation hiérarchique réside dans le choix de la partition à extraire de cette hiérarchie. Par exemple, partant d'une hiérarchie de segmentations, comment choisir une coupe permettant, au mieux, d'extraire les frontières des objets que l'on souhaite segmenter ? Ce problème de choix de partition n'est actuellement pas résolu. Certaines approches [Bruzzone et Carlin, 2006] contournent cette difficulté en utilisant des attributs calculés sur les segments à travers toute la hiérarchie pour caractériser les pixels issus de ces segments. Cependant, ces approches ne classifient alors plus des régions mais des pixels caractérisés *via* des hiérarchies de segments. Elles n'exploitent donc pas toutes les propriétés basées « régions » intrinsèques à ces hiérarchies. Nous reviendrons sur ce problème particulier dans la suite de ce manuscrit.

L'intérêt des images multisources

Sommaire

3.1 Images multisources	39
3.1.1 Différents types de données multisources	39
3.1.2 Des données complémentaires	40
3.1.3 Les images multisources sont des données complexes	40
3.1.4 Des représentations différentes des objets d'intérêt	40
3.2 Analyser des images multisources	43
3.2.1 Différentes sources à analyser	43
3.2.2 Analyse par fusion des données et/ou des résultats	43
3.2.3 Analyse collaborative des différentes sources	44
3.3 Approche collaborative basée régions pour l'analyse d'images multirésolutions . .	46
3.3.1 Analyse multirésolution et approches basées régions	46
3.3.2 Remettre en cause les résultats <i>via</i> une collaboration	46
3.3.3 Vers un processus collaboratif basé régions	47

Ce chapitre présente l'intérêt des images multisources dans le contexte de l'analyse d'images. Dans un premier temps nous montrons que de nombreuses données multisources sont actuellement disponibles, et que ces données sont généralement complémentaires (Section 3.1). Nous discutons ensuite de l'intérêt de cette propriété particulière pour pallier les problèmes des approches d'analyse multiniveau basées régions, en s'abstrayant de l'hétérogénéité des données (T)HRS (Section 3.2). Nous introduisons finalement l'approche d'analyse collaborative basée régions mise en œuvre dans le cadre de cette thèse (Section 3.3).

3.1 Images multisources

3.1.1 Différents types de données multisources

La dernière décennie a été marquée par une très forte augmentation du nombre de capteurs d'images disponibles (*e.g.*, terrestres, aériens, spatiaux), permettant de diversifier les sources de collecte des données, mais également de s'adapter aux besoins des utilisateurs. Ainsi, dans le domaine de la télédétection, la diversité des missions spatiales ou aériennes, liées à des besoins variés tels que la récolte d'informations sur la nature et l'étendue des sols, sur leur composition, ou encore sur l'étalement urbain, permet de recueillir une quantité importante de données relatives à la surface terrestre. Selon le contexte et la mission, les données images issues de ces acquisitions peuvent être dotées de caractéristiques différentes. Par exemple, les images peuvent avoir des résolutions spatiales kilométriques ou infra-métrique, et des résolutions spectrales très différentes : visible, infrarouge, hyperspectraux, RADAR et, plus récemment, laser aéroporté (LiDAR) fournissant des informations sur l'élévation des objets observés. De plus, ces images peuvent avoir été acquises à des dates différentes. Pour qualifier d'une manière générale ces masses de données hétérogènes, on emploiera par la suite le terme d'images ou de données multisources. Dans le cadre de ces travaux, on se focalisera sur le cas où les données sont exclusivement issues de capteurs optiques.

L'accès à des données image de plus en plus nombreuses, issues de différents capteurs aux résolutions spatiales, spectrales et temporelles très diverses, impose de repenser les modes de traitement et d'extraction de connaissances à partir des images télédéteectées. Ainsi, la multiplicité des données tend à favoriser l'utilisation conjointe de données multisources dans les méthodes de segmentation et de classification.

3.1.2 Des données complémentaires

Les ensembles de données multisources peuvent être constitués d'images possédant des caractéristiques très différentes. Dans le cadre de l'imagerie satellitaire, ces caractéristiques sont directement liées aux propriétés des capteurs. Parmi celles-ci on peut distinguer les résolutions spectrale et spatiale d'un capteur, ainsi que son temps de re-visite d'une scène particulière. Des images acquises sur la même scène par ces différents capteurs peuvent alors offrir des propriétés complémentaires :

- disposer d'images à des résolutions spectrales différentes permet d'obtenir une plus large description radiométrique des objets géographiques représentés. En effet, des études [Miller et Small, 2003] ont montré que la résolution spectrale des capteurs est une propriété essentielle, particulièrement pour discriminer différents types d'occupation du sol dans des environnements complexes tels que les milieux urbains ;
- disposer d'images à des résolutions spatiales différentes permet d'enrichir et de faciliter l'analyse d'une scène. Par exemple, pour cartographier le territoire au niveau des grandes zones urbaines, les experts utilisent généralement des images MRS car leurs résolutions spatiales sont adaptées à ce niveau d'analyse. Les objets d'intérêts (*e.g.*, quartiers, blocs urbains) pouvant apparaître comme complexes et hétérogènes au sein d'images (T)HRS y sont notamment plus homogènes et donc plus faciles à délimiter. En revanche, pour cartographier le territoire au niveau des objets urbains simples, les images (T)HRS sont utilisées, car ces objets ne sont pas distinguables à partir d'images MRS. Au demeurant, cette information multirésolution peut permettre d'améliorer un résultat d'analyse. Par exemple, pour interpréter une image à une résolution donnée, il peut être utile de prendre en compte des informations obtenues à des résolutions plus élevées (régions plus spécialisées) ou plus basses (zones thématiques) : cette interprétation doit pouvoir être utilisée pour confirmer ou infirmer des interprétations faites à ces autres résolutions ;
- disposer de séries temporelles d'images permet d'améliorer la capacité à détecter des objets géographiques et à analyser leur comportement au cours du temps. En effet, deux objets au sol peuvent avoir une signature spectrale similaire à un instant donné mais un comportement temporel différent. Les séries d'images multitemporelles peuvent ainsi être utilisées pour discriminer ces différents objets et remettre en cause le résultat d'analyse à une date particulière. Ces caractéristiques temporelles offrent ainsi une importante propriété de complémentarité.

3.1.3 Les images multisources sont des données complexes

Les images multisources offrent ainsi des représentations différentes mais complémentaires des objets d'intérêt de la scène. Cependant, cette propriété particulière induit une forte complexité sur ces données. Dans les domaines liés à la fouille de données, une donnée peut être qualifiée de complexe [Zighed et al., 2009] quand :

1. elle est représentée par différents types d'informations et/ou de caractéristiques ;
2. elle est multistructurée (*e.g.*, spatialement, temporellement, radiométriquement) ;
3. elle provient de sources hétérogènes ;
4. le phénomène qu'elle représente est décrit selon plusieurs points de vue ;
5. ses caractéristiques peuvent évoluer dans le temps.

Ainsi, les ensembles d'images multisources engendrent une importante masse de données complexes. Ce phénomène rend de plus en plus difficile l'application de méthodes classiques d'analyse d'images pour extraire des informations pertinentes à partir de ces données.

3.1.4 Des représentations différentes des objets d'intérêt

A. Des vues différentes

Dans le domaine de la télédétection, de nombreuses méthodes d'analyse de données multisources considèrent qu'il existe un objet géographique « réel » (ici la scène à analyser) qui peut être observé suivant

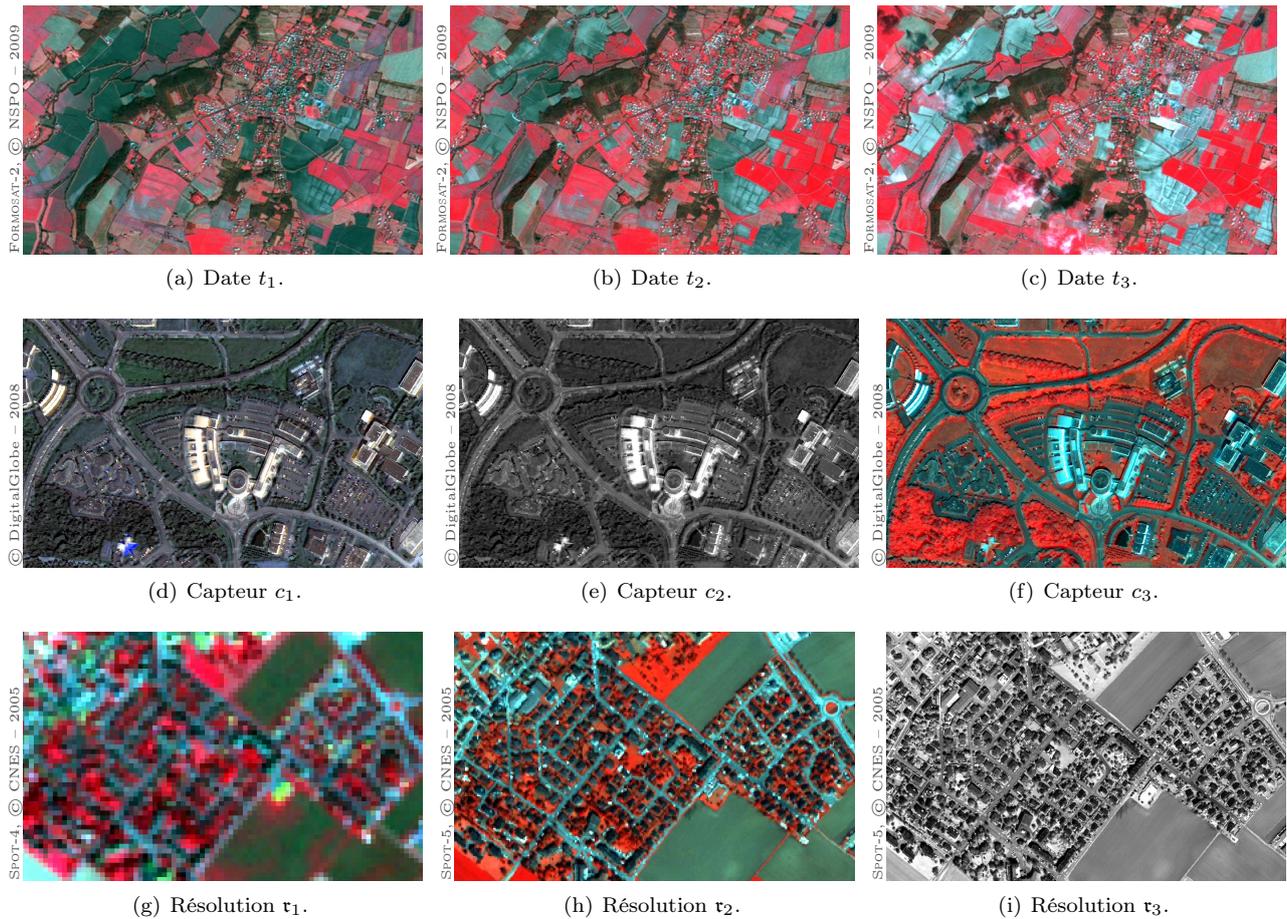


FIGURE 3.1 – Trois cas de données multisources. (a–c) Même résolution / capteur ; dates différentes. (d–f) Même résolution ; capteurs différents. (g–i) Différentes résolutions ; capteurs différents.

différents points de vue. Le but est alors de trouver une description de cet objet géographique prenant en compte ces différentes représentations. Chaque vue \mathcal{V}^i est liée à une image \mathcal{I}^i dont le domaine de définition est E^i . Plusieurs situations sont alors envisagées (voir la figure 3.1) :

1. toutes les images contiennent les mêmes éléments, et ceux-ci sont décrits par les mêmes attributs radiométriques avec des valeurs différentes dans les différentes images (*e.g.*, deux images de la même zone, du même satellite mais à des dates différentes) ;
2. toutes les images contiennent les mêmes éléments, et ceux-ci sont décrits par des attributs radiométriques différents (*e.g.*, deux images de la même zone, à la même résolution mais provenant de capteurs différents) ;
3. les images ne contiennent pas les mêmes éléments, et ceux-ci sont décrits par des attributs radiométriques différents (*e.g.*, deux images de la même zone, à des résolutions différentes provenant de capteurs différents).

Les images multisources peuvent ainsi être composées d'ensembles hétérogènes d'images, chacune pouvant fournir une représentation différente (mais souvent complémentaire) des objets géographiques représentés. Suivant ces différentes configurations, les méthodes d'analyse utilisées pour extraire des connaissances de ces ensembles de données pourront être sensiblement différentes.

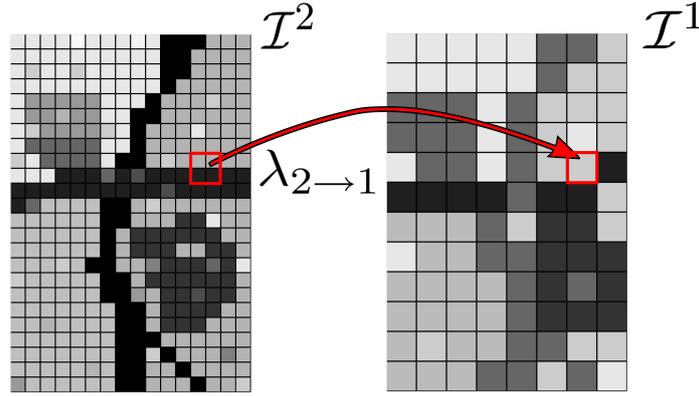


FIGURE 3.2 – Illustration de la fonction de correspondance $\lambda_{2 \rightarrow 1}$ entre deux images \mathcal{I}^2 et \mathcal{I}^1 . Dans cet exemple, $\alpha = 2$, et un pixel de \mathcal{I}^1 correspond à 4 pixels de \mathcal{I}^2 .

B. Mise en correspondance des données

Pour pouvoir analyser plusieurs vues caractérisant simultanément les mêmes objets géographiques, de nombreuses méthodes d'analyse proposées dans la littérature nécessitent de mettre en relation les données décrites par ces vues.

Grâce au géoréférencement, il est possible de mettre en correspondance toutes les acquisitions disponibles (à des résolutions différentes ou non) d'une même scène. Celui-ci permet, pour chaque pixel d'une image, de connaître sa position réelle sur le globe terrestre (en fonction d'un géoréférentiel particulier). On peut donc obtenir une nouvelle description des données en considérant chacune des images disponibles.

Soit deux images $\mathcal{I}^1 : E^1 \rightarrow V^1$ et $\mathcal{I}^2 : E^2 \rightarrow V^2$ (avec $E^* = \llbracket 0, d_x^* - 1 \rrbracket \times \llbracket 0, d_y^* - 1 \rrbracket$) représentant la même scène. Lorsque ces images ont la même résolution spatiale $E^1 = E^2$, la correspondance est directe. Dans le cas où les résolutions spatiales de ces images sont différentes, la correspondance n'est pas immédiate. Considérons le cas où \mathcal{I}^1 et \mathcal{I}^2 ont deux résolutions spatiales différentes $r^1 < r^2$.

On pose $\alpha = d_x^2/d_x^1 = d_y^2/d_y^1 \in \mathbb{N}^*$. Nous nous situons ici dans le cas particulier où $d_x^2 = \alpha \cdot d_x^1$ et $d_y^2 = \alpha \cdot d_y^1$. Le coefficient α caractérise la « différence » de résolution entre les deux images. Ainsi, un point $\mathbf{x} \in E^1$ correspond « physiquement » à un ensemble composé de $\alpha \times \alpha$ points dans E^2 . Le point $\mathbf{x} \in E^1$ et les $\alpha \times \alpha$ points de E^2 couvrent la même zone géographique.

Dans le but de modéliser la correspondance entre les points des images \mathcal{I}^1 et \mathcal{I}^2 , on définit deux fonctions de correspondance

$$\lambda_{2 \rightarrow 1} : E^2 \rightarrow E^1 \quad \text{et} \quad \lambda_{1 \rightarrow 2} : E^1 \rightarrow 2^{E^2} \\ \mathbf{x} \mapsto (x/\alpha, y/\alpha) \quad \text{et} \quad \mathbf{x} \mapsto \alpha \times (x, y) + \llbracket 0, \alpha - 1 \rrbracket^2 \quad (3.1)$$

La figure 3.2 illustre la fonction de correspondance $\lambda_{2 \rightarrow 1}$.

C. Limites

Il est à noter que les fonctions de mise en correspondance, présentées dans le paragraphe précédent, reflètent une vision simplifiée du problème de mise en correspondance des objets géographiques entre deux acquisitions différentes. En effet, cette propriété n'est bien souvent pas garantie entre les différents objets terrestres représentées par un ensemble d'images multisource.

Pour illustrer ce phénomène, considérons deux images THRS représentant la même scène, acquises par le même capteur à deux dates différentes. Entre ces deux acquisitions, le capteur a pu être décalé ou les paramètres d'acquisition ont pu être différents (*e.g.*, dépointement du satellite) produisant alors des représentations différentes des objets géographiques. Dans le cas extrême, un bâtiment peut être représenté par son toit et l'une de ses façades dans une acquisition, et uniquement par son toit dans l'autre acquisition. De même, certains objets géographiques peuvent être masqués au sein d'une image (*e.g.*, occlusions nuageuses) mais peuvent être visibles au sein d'une autre image acquise à une date différente.

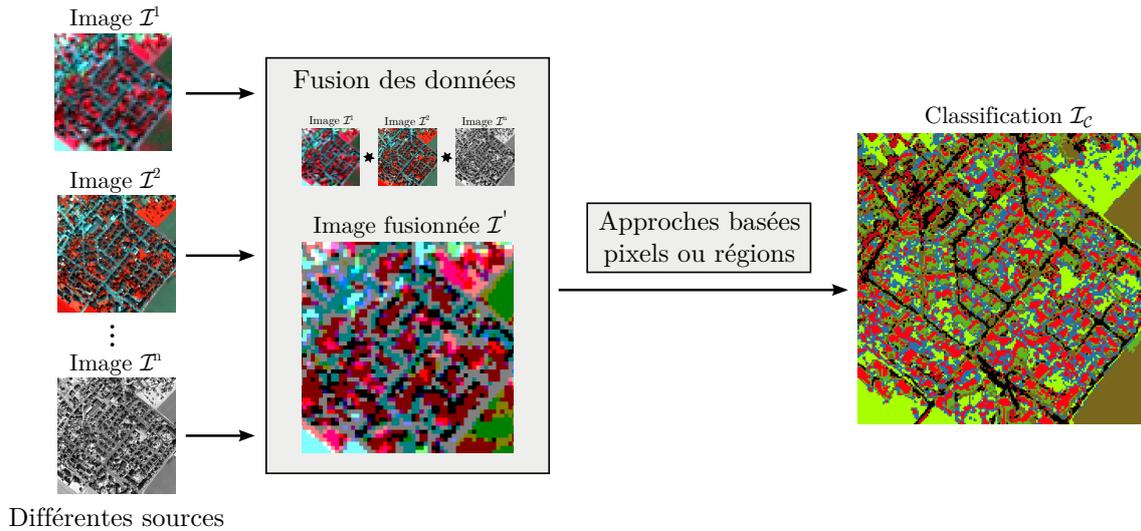


FIGURE 3.3 – Schéma d'une approche par fusion pour l'analyse d'images multisources.

La gestion de ces problèmes ne rentrant pas dans le cadre de ces travaux, on considérera par la suite que les pré-traitements effectués sur les données (*i.e.*, orthorectification) permettent une mise en correspondance pertinente entre les objets d'intérêt représentés entre les différentes images.

3.2 Analyser des images multisources

3.2.1 Différentes sources à analyser

Les données multisources peuvent ainsi être considérées comme des ensembles de données complémentaires et permettent d'obtenir des représentations différentes des objets géographiques représentés. Pour pouvoir analyser ces données à différents niveaux d'échelle (*i.e.*, analyse multiniveau), il est nécessaire d'en extraire différents niveaux d'objets d'intérêt. Cependant, les approches d'analyse énoncées précédemment (*i.e.*, approches basées pixels ou régions) ne peuvent pas être employées directement car ces dernières sont destinées au traitement de données monosources.

Face à ce problème, des approches d'analyse spécialement dédiées au traitement de données multisources ont été proposées. L'analyse de données multisources consiste à prendre en compte différentes sources d'information dans un processus global de fouille de données. On peut distinguer deux familles d'approches : celles procédant par fusion des différentes données disponibles, et celles analysant indépendamment (et/ou conjointement) les différentes sources. Nous présentons par la suite ces approches.

3.2.2 Analyse par fusion des données et/ou des résultats

Une première stratégie pour analyser des images multisources consiste à fusionner les données afin de construire une vue unique sur laquelle est alors appliquée une approche d'analyse basée pixels ou régions afin d'extraire les classes recherchées (Figure 3.3).

Dans [Chibani, 2005] une méthode a été proposée pour la combinaison d'images multispectrales, panchromatiques et radars. Cette méthode utilise conjointement la transformation Intensité-Teinte-Saturation et la décomposition en ondelette. Dans [Chang et al., 2007] une méthode basée sur la génération artificielle de bandes spectrales est proposée. Ces bandes sont dans un premier temps générées automatiquement à partir de plusieurs sources pour caractériser les pixels. Dans un second temps ces pixels sont classifiés par le biais d'une approche spécifique.

Dans le cas particulier d'images multirésolutions, un procédé spécifique nommé *pansharpening* a été proposé. En effet, certains satellites produisent simultanément lors d'une prise de vue, plusieurs images de la

même zone : une image panchromatique et une image multispectrale. L'image panchromatique présente une forte résolution spatiale mais une faible résolution spectrale, alors que l'image multispectrale a une forte résolution spectrale mais une faible résolution spatiale. Une solution pour utiliser ces deux sources d'information consiste à fusionner l'image panchromatique et l'image multispectrale. De nombreuses méthodes ont été étudiées au cours des dernières années pour fusionner ces deux types d'images et ainsi produire une image de bonne qualité spatiale et spectrale [Dou et al., 2007].

Une autre approche, dite fusion des caractéristiques, consiste à créer de nouveaux objets (*e.g.*, segments, clusters) à partir des différentes données, puis à les fusionner. Par exemple, les auteurs de [Dubuisson et Jain, 1995] proposent de fusionner des segmentations issues du traitement d'images multisources. Dans [Germain et al., 2004], les auteurs présentent une méthode basée sur la théorie des croyances proposée par Dempster-Shafer [Shafer, 1976].

Cette classification à sources multiples peut aussi se faire au niveau des décisions : les résultats de la classification de chacune des images sont fusionnés en un résultat unifiant. Cette fusion peut alors être réalisée conformément à la décision prise par chacune des méthodes de classification. Dans [Benediktsson et Kanellopoulos, 1999], les auteurs proposent une méthode de fusion de décisions fondée sur la combinaison de schémas statistiques avec les réseaux neuronaux pour la classification des sources multiples.

Néanmoins, les approches par fusion souffrent des problèmes suivants : (1) la fusion de données engendre une perte d'information ; (2) il n'existe pas de méthode de fusion pour tous les types d'images ; (3) le nombre de bandes ou d'attributs peut augmenter considérablement, au risque de se heurter à la « malédiction de la dimensionnalité » [Hughes, 1968] : la majorité des méthodes de segmentation et de classification basées sur des calculs de distance ne sont plus efficaces devant la surabondance de bandes ou d'attributs car les distances entre les données ne présentent alors plus de différences significatives.

Dans le cas particulier d'images dégradées, certaines techniques d'analyse utilisent des résolutions différentes d'une même image, générées artificiellement à partir de la décomposition par ondelettes d'une image initiale [Mallat, 1989, 1996; Sheikholeslami et al., 2000]. Cette opération transforme l'espace des données en un espace où les clusters sont plus distinctifs et donc plus facilement séparables. En particulier, dans [Scheunders et Sijbers, 2002; Kim et Kim, 2003], des algorithmes basés sur la LPE ont été proposés pour traiter des images multirésolutions créées *via* une transformation par ondelettes. D'autres méthodes utilisent des moyens de détérioration comme dans [Rezaee et al., 2000] où les auteurs emploient une technique de décomposition pyramidale (*i.e.*, chaque étage de la pyramide correspond à une résolution artificielle). Le problème inhérent à ces méthodes réside dans le fait que la détérioration d'une image à une résolution donnée vers une résolution inférieure n'apporte pas les mêmes informations qu'une image acquise directement à cette résolution. En effet, la détérioration d'une image peut introduire de nouveaux contours et de nouvelles valeurs spectrales n'existant pas « physiquement » dans la scène observée et altérant ainsi le processus d'analyse.

En conséquence, la fusion de données multisources, en particulier d'images multirésolutions, peut potentiellement améliorer les résultats fournis par des méthodes d'analyse basées pixels ou régions. En particulier les propriétés d'homogénéité induites par l'utilisation d'images à des résolutions plus grossières que les images (T)HRS, peuvent permettre d'améliorer les résultats fournis par des approches multiniveaux. Cependant, la fusion des données engendre, dans la plupart des cas, une perte d'information. Les premières expériences que nous avons menées ont montré que l'utilisation d'approches d'analyse d'images (basées pixels ou régions) sur des données multirésolutions fusionnées ne permettaient pas d'extraire efficacement les différents niveaux d'informations recherchés [Wemmert et al., 2009].

3.2.3 Analyse collaborative des différentes sources

Comme nous l'avons vu précédemment, les approches par fusion ne considèrent souvent qu'une source de données à la fois (pouvant être obtenue *via* une fusion de différentes sources). Il n'existe donc pas, la plupart du temps, d'échange d'information entre les processus liés aux traitements des différentes sources. Pour pouvoir utiliser toutes les informations disponibles et tirer ainsi parti de la complémentarité de ces sources, des approches récentes se sont intéressées à l'utilisation conjointe de plusieurs sources de données dans un processus unique d'analyse (Figure 3.4).

De nombreux travaux ont déjà été réalisés sur la collaboration de méthodes de classification non-supervisée

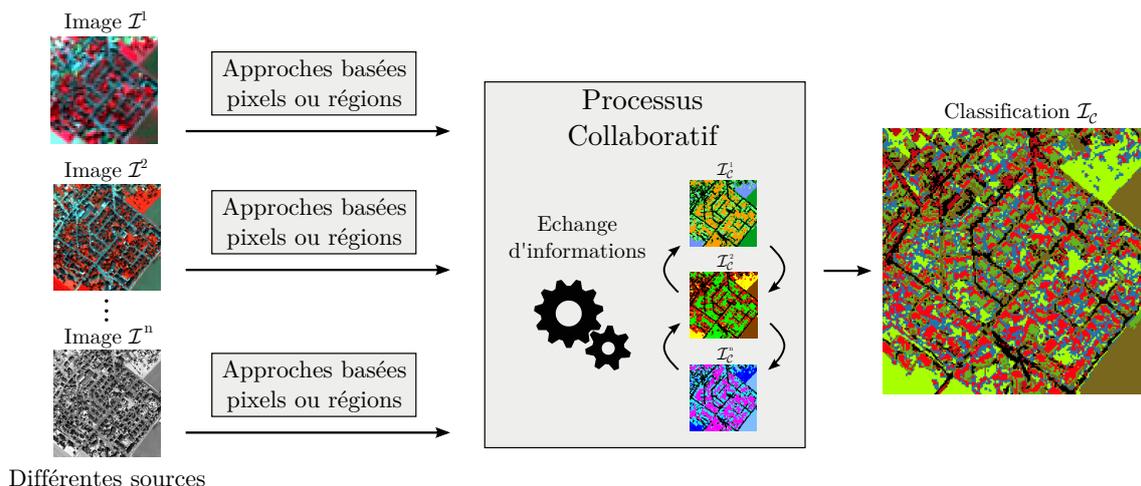


FIGURE 3.4 – Schéma d'une approche collaborative pour l'analyse d'images multisources.

pour le cas multisource [Bruzzone et al., 2002]. Parmi les différentes approches proposées, on trouve la méthode SAMARAH [Wemmert, 2000; Forestier, 2010]. Cette méthode est basée sur une approche de clustering qui consiste à faire collaborer plusieurs méthodes de classification non-supervisée pour tenter de trouver un consensus sur le clustering d'un jeu de données. Ces différentes méthodes vont partager des informations et remettre en cause leurs résultats en fonction des décisions prises par chacune. Ainsi, une discussion est entreprise entre ces méthodes dans le but de faire converger collectivement les différents résultats afin que ceux-ci soient suffisamment comparables pour qu'un mécanisme d'unification (basé sur un vote) puisse être réalisé. L'objectif de cette collaboration est de réduire l'impact du choix d'une méthode et de ses paramètres sur le résultat. Cette approche a été initialement proposée dans le cadre de l'analyse de données monosources.

Pour pouvoir traiter des données multisources, la méthode SAMARAH a été, dans un premier temps, étendue afin de permettre la classification d'images à des résolutions spatiales et/ou spectrales proches (*i.e.*, représentant les mêmes catégories d'objets urbains) en essayant de trouver un consensus entre les différents résultats de classification de chacune des sources mais avec un même nombre de clusters [Forestier et al., 2008]. L'amélioration obtenue grâce à l'utilisation de différentes sources s'effectue autant sur les résultats finaux des différentes méthodes que sur le résultat unifiant. Néanmoins, en fonction de l'hétérogénéité de la résolution spatiale des images considérées, le résultat unifiant peut ne pas être pertinent. En particulier, ce problème apparaît lorsque les différentes images traitées présentent des résolutions spatiales très différentes. En effet, la méthode SAMARAH, qui cherche un consensus entre les différents résultats de classification de chacune des sources (avec un même nombre de clusters) n'est pas adaptée pour obtenir un résultat permettant une analyse de ces différents niveaux. Afin de résoudre ce problème, des travaux ont été réalisés dans le but d'étendre cette approche aux données multirésolutions.

Le principal verrou rencontré lors de cette extension a résidé dans le fait que les différents résultats peuvent ne pas comporter le même nombre de clusters. Il a donc été nécessaire de redéfinir les critères de similarité entre résultats ainsi que le mécanisme d'unification. En particulier, la contrainte de similarité inter-résultats, qui impose que les clusters soient identiques, a été atténuée afin que les clusters d'un résultat puissent n'être que des spécialisations (en termes de partitions de l'espace des données) d'un cluster à une résolution plus basse. En effet, comme nous l'avons vu précédemment, les clusters résultant des images considérées sont totalement différents d'une résolution à une autre. Il a donc été nécessaire d'étudier la mise en correspondance des clusters entre les résultats à différentes résolutions. Cette étude a débouché sur la définition d'un nouveau critère de comparaison des résultats basé sur leur cohérence.

Les résultats multiniveaux fournis par cette approche collaborative se sont montrés prometteurs. Cependant l'approche est basée sur un paradigme basé pixels et ne permet donc pas de tirer parti des avantages offerts par les paradigmes régions, et en particulier les approches hiérarchiques.

3.3 Approche collaborative basée régions pour l'analyse d'images multirésolutions

3.3.1 Analyse multirésolution et approches basées régions

Les approches collaboratives, et en particulier la méthode SAMARAH, se sont montrées prometteuses pour obtenir un résultat d'analyse multiniveau à partir de données multirésolutions. Cependant, ces approches reposent directement sur un paradigme basé pixels. Elles n'exploitent donc pas les avantages offerts par l'utilisation des paradigmes basés régions, et en particulier ceux offerts par les approches hiérarchiques, permettant le traitement de segments pouvant correspondre à des objets d'intérêt complexes.

L'extension des approches collaboratives à des paradigmes basés régions pose cependant de nombreux problèmes. En effet, une telle extension nécessite de pouvoir mettre en correspondance les différents objets à classifier à travers les différentes résolutions. Alors que cette mise en correspondance est simple dans les approches basées pixels (*i.e.*, géoréférencement), elle reste un problème ouvert dans le cadre de la manipulation des segments des images produits par l'étape de segmentation.

Néanmoins, dans une vision de prospective, une approche de classification collaborative basée régions a été proposée. Son principe repose sur la cohérence des clusters de régions issus des différentes images [Wemmert et al., 2009]. Cette cohérence, qualifiée de cohérence de zones, consiste pour un cluster particulier d'une source donnée, à utiliser l'hétérogénéité des clusters associés aux pixels connexes dans l'autre source, pour enrichir la description/caractérisation des données. Considérons un couple d'images $(\mathcal{I}^i, \mathcal{I}^j)$. Chaque segment de l'image \mathcal{I}^i peut être caractérisé par un histogramme modélisant la répartition des labels de clusters des pixels de l'image \mathcal{I}^j qui lui correspondent directement par géoréférencement (et réciproquement). Ces nouvelles données sont alors classifiées *via* SAMARAH ou une approche non-collaborative de clustering. Bien que cette approche soit basée sur un paradigme de collaboration des régions entre les différentes sources, la mise en correspondance de ces dernières est effectuée au niveau des pixels de l'image.

Ces premiers résultats sur l'analyse de données multirésolutions, par le biais d'approches collaboratives, ont permis de montrer que la difficulté majeure, pour pouvoir employer un paradigme régions, résidait dans le fait de pouvoir mettre en relation les régions formées au sein des différentes sources. En conséquence, la phase collaborative ne doit ainsi pas être réalisée uniquement au cours de l'étape de classification mais aussi pendant l'étape de segmentation. Une collaboration au niveau des régions des différentes images peut ainsi permettre d'améliorer les résultats d'analyse dans chacune des résolutions. Cependant, la collaboration de plusieurs algorithmes de segmentation (travaillant sur des sources de données distinctes) est un procédé complexe qui, à notre connaissance, n'a pas été jusqu'à présent étudié dans la littérature.

3.3.2 Remettre en cause les résultats *via* une collaboration

Un autre avantage induit par les approches collaboratives est la possibilité de remettre en cause les résultats d'analyse obtenus sur les différentes sources. Comme il a été vu précédemment, les approches impliquant une collaboration entre plusieurs processus de fouille de données cherchent généralement à obtenir un consensus entre les résultats partagés. Cette recherche de consensus entre résultats est souvent réalisée par la remise en cause des résultats réciproques fournis par les processus impliqués dans cette collaboration. En conséquence, le résultat d'analyse unifiant obtenu s'en trouve amélioré.

Dans le cadre de la collaboration entre plusieurs processus de classification non-supervisée, la méthode SAMARAH cherche à trouver un consensus sur le clustering d'un jeu de données. Les différents processus impliqués vont partager des informations et remettre en cause leurs résultats de clustering en fonction des décisions prises par chacun. Dans le cadre de l'utilisation d'approches basées régions pour l'analyse d'images, une collaboration entre un processus de segmentation et un processus de classification est réalisée. Cette collaboration peut aussi permettre la recherche d'un consensus entre ces deux processus et la remise en cause des résultats de segmentation et de classification. Dans [Derivaux et al., 2010], les auteurs proposent un schéma itératif impliquant un algorithme de segmentation et un algorithme de classification. Dans ces travaux, les résultats obtenus par le biais de l'algorithme de segmentation sont remis en cause par ceux obtenus par l'algorithme de classification. Si ce dernier n'arrive pas à classifier correctement les régions formées par l'étape

de segmentation suivie de l'étape de caractérisation, alors les paramètres de l'algorithme de segmentation sont automatiquement modifiés dans le but d'obtenir une segmentation dont les régions seront plus fines (ou plus larges), permettant ainsi une nouvelle classification de ces dernières. La remise en cause est effectuée ici au niveau global (*i.e.*, remise en cause de toute la segmentation). Ainsi, ces travaux ont montré que l'étape de classification pouvait remettre en cause les résultats de segmentation.

L'utilisation d'approches basées régions pour l'analyse d'images multirésolutions peut également permettre une remise en cause des résultats par le biais d'un consensus multirésolution. En effet l'étude de la classification d'une image à une résolution donnée peut remettre en cause la classification d'une autre image à une résolution différente. Dans les approches basées régions, le résultat de classification de chacune des images est directement dépendant de deux processus : (1) la segmentation de l'image et (2) la classification des régions de l'image segmentée. En conséquence, la remise en cause des résultats de classification d'une image, à une résolution donnée, peut donc être réalisée soit au niveau de la segmentation de cette image, soit au niveau de la classification des régions résultantes. Ceci est valable d'une manière globale (remise en cause de toute la segmentation ou la classification) ou locale (remise en cause d'une ou plusieurs régions uniquement).

3.3.3 Vers un processus collaboratif basé régions

L'utilisation de données multirésolutions peut ainsi faciliter l'extraction de structures d'intérêt complexes pouvant apparaître comme hétérogènes à partir d'images (T)HRS. Parallèlement, les approches les plus efficaces pour traiter de telles données sont les approches collaboratives qui permettent, de plus, d'obtenir des résultats conduisant à des analyses multiniveaux. Cependant, ces approches reposent directement sur une stratégie basée pixels. Bien que certaines approches aient été proposées pour les adapter au paradigme basé régions, ces dernières n'exploitent pas totalement les avantages offerts par l'utilisation d'un tel paradigme. Il semble ainsi difficile d'étendre les approches d'analyse hiérarchique basées régions au traitement de données multirésolutions. La principale limite réside dans le fait de parvenir à mettre en relation les régions issues des différentes images.

Nous présentons, dans la partie suivante (Partie II), un processus collaboratif permettant de combiner les avantages liés aux approches hiérarchiques basées régions et les avantages liés au traitement d'images multirésolutions. Ce processus repose sur une collaboration entre des algorithmes de segmentation et de classification, appliqués aux différentes résolutions afin d'obtenir des résultats d'analyse multiniveaux. Pour s'abstraire des problèmes liés à la mise en correspondance des régions, nous proposons de construire directement la hiérarchie de régions d'une manière descendante à travers la résolution des images. Le principe de ce processus consiste à extraire des segments/régions grossiers de l'image à la plus faible résolution, puis à les affiner progressivement en utilisant des images à des résolutions plus fines. L'objectif est ainsi de faciliter l'extraction de hiérarchies d'objets complexes et hétérogènes à partir d'images à (T)HRS.

Parallèlement, nous avons montré que les approches collaboratives offraient la possibilité de remettre en cause les résultats d'analyse obtenus sur les différentes résolutions. Nous verrons ensuite (Partie III) que le processus collaboratif descendant (présenté en Partie II) est affecté de problèmes spécifiques, principalement liés à sa nature descendante et à son caractère non-supervisé. Nous montrerons alors qu'il peut être complété par un post-processus ascendant permettant de remettre en cause les résultats d'analyse multiniveaux. Notre objectif est d'étudier la faisabilité d'un processus de remise en cause de la segmentation et de la classification, basé sur l'analyse conjointe des différentes résolutions disponibles : cette analyse collaborative correspond à un processus d'enrichissement mutuel des phases de segmentation et de classification.

Deuxième partie

Processus collaboratif pour l'analyse descendante d'images multirésolutions

★ Résumé ★

Au cours de cette partie, nous présentons un processus collaboratif original permettant d'analyser des images multirésolutions, pour en extraire des hiérarchies d'objets d'intérêt. Basé sur le principe d'une analyse descendante, ce processus a pour objectif d'automatiser au maximum la méthode d'analyse classiquement employée par les experts en photo-interprétation : extraire les grandes zones d'intérêt de la scène observée (*e.g.*, zones urbaines, zones agricoles) à partir des images aux résolutions les plus basses, puis les décomposer récursivement pour en extraire des zones d'intérêt plus fines (*e.g.*, objets urbains, champs) à partir des images aux résolutions les plus élevées. Ce processus implique des mécanismes de segmentation hiérarchique interactive et de classification multirésolution, eux-aussi présentés en détail dans cette partie. Les géosciences et plus particulièrement l'observation de la Terre *via* les images de télédétection étant le domaine privilégié d'application des propositions faites lors de cette thèse, nous présentons une validation expérimentale de ce processus dans le cadre de l'analyse d'environnements urbains.

Analyse descendante d'images multirésolutions

Sommaire

4.1 Motivations	51
4.1.1 Faciliter l'extraction d'objets d'intérêt complexes à partir des images	52
4.1.2 Permettre une analyse multiniveau des territoires observés	52
4.1.3 Mise en correspondance des objets d'intérêt issus des différents niveaux	52
4.2 Proposition	53
4.2.1 Un processus collaboratif basé régions pour le traitement d'images multirésolutions . . .	53
4.2.2 Mise en correspondance des segments à travers les différentes images	53
4.2.3 Une stratégie collaborative descendante à travers la résolution	54
4.2.4 Avantages	54
4.3 Schéma méthodologique	54
4.3.1 Principe	54
4.3.2 Présentation des briques méthodologiques	55

Ce chapitre présente le schéma méthodologique du processus collaboratif basé régions proposé au cours de cette thèse. Pour ce faire, nous commençons par énoncer les motivations qui ont conduit au développement de ce processus (Section 4.1). Nous décrivons ensuite notre proposition méthodologique afin d'extraire différents niveaux d'information à partir d'images multirésolutions, permettant d'apporter une solution aux différents problèmes levés dans ces motivations (Section 4.2). Finalement, nous présentons plus en détail le processus global collaboratif proposé, nommé PCSC, impliquant une approche de segmentation hiérarchique basée sur des exemples et une approche de classification multirésolution, développées elles-aussi au cours de ces travaux (Section 4.3). Le principe de ce processus repose sur une collaboration entre ces deux approches et sa principale originalité est de s'appuyer sur les propriétés offertes par les images multirésolutions, pour faciliter l'analyse des images (T)HRS.

4.1 Motivations

Dans la partie précédente, nous avons pu mettre en évidence trois objectifs principaux qui ont motivés ces travaux :

1. **faciliter l'extraction d'objets d'intérêt complexes à partir d'images satellites optiques ;**
2. **permettre une analyse multiniveau des territoires observés ;**
3. **pouvoir mettre en correspondance les objets d'intérêt issus des différents niveaux.**

Nous détaillons dans la suite de cette section, ces motivations, aussi bien liées à des besoins thématiques qu'à des besoins méthodologiques, et ayant conduit par la suite au développement du processus proposé dans le cadre du premier axe de recherche de ces travaux.

4.1.1 Faciliter l'extraction d'objets d'intérêt complexes à partir des images

Dans le domaine de la télédétection, et en particulier dans celui de la cartographie urbaine, de nombreux objets d'intérêt représentés par le biais d'images (T)HRS sont composés d'ensembles d'objets simples comme des toits de bâtiments, des parcelles de végétation, ou encore des tronçons de routes organisés de façon structurée. Ces objets sont alors qualifiés d'objets géographiques complexes car ils peuvent apparaître comme hétérogènes au sein de telles images.

Les approches basées régions offrent une solution potentielle pour l'extraction automatique de telles structures d'intérêt, mais elles sont limitées par l'hétérogénéité radiométrique importante de ces objets d'intérêt. Bien que des extensions aient été proposées à ces approches pour tenter de reconstruire ces objets d'intérêt complexes par des regroupements d'objets simples (de manière ascendante), ces dernières demeurent difficiles à employer en raison de la forte complexité des données et de leur coût algorithmique élevé (Section 2.3.3).

Face à ces limites, il semble alors pertinent de proposer une méthodologie permettant de pallier ces problèmes, et d'extraire des objets d'intérêt géographiques complexes à partir d'images (T)HRS.

4.1.2 Permettre une analyse multiniveau des territoires observés

Par ailleurs, en analyse d'images, il est souvent nécessaire d'extraire plusieurs niveaux de connaissance et d'information à partir des images. Ces différents niveaux sont le plus souvent inhérents à la structure spatiale des objets d'intérêt observés. Dans le domaine de la télédétection, et en particulier dans celui de la cartographie urbaine, les utilisateurs emploient les images pour extraire différents niveaux d'objets d'intérêt dans le but d'analyser le territoire à plusieurs niveaux d'échelle (*e.g.*, de la tache urbaine aux bâtiments, en passant par les tissus urbains).

Pour permettre une analyse multiniveau de la scène étudiée, deux familles d'approches peuvent être considérées : (1) les approches d'analyse hiérarchique basées régions, appliquées au traitement d'images (T)HRS, et (2) les approches d'analyse collaboratives appliquées au traitement d'images multirésolutions. Nous avons vu précédemment que les approches d'analyse hiérarchique basées régions sont limitées par l'hétérogénéité des images à (très) hautes résolutions spatiales (en particulier pour l'extraction d'objets d'intérêt complexes) et qu'une solution pour les améliorer consiste à les étendre au traitement de données multirésolutions (Section 2.3.3). En effet, de tels objets d'intérêt apparaissent comme plus homogènes à travers des images de plus faibles résolutions, et donc plus faciles à extraire. Parallèlement, nous avons vu que les approches les plus adaptées pour traiter des données multirésolutions sont les approches collaboratives, qui fonctionnent directement sur des paradigmes pixels et analysent les images sans les fusionner.

En conséquence, il semble pertinent de proposer une méthodologie permettant d'obtenir une analyse multiniveau des territoires observés, profitant des avantages liés aux approches hiérarchiques basées régions et des avantages liés au traitement d'images multirésolutions. Cependant, l'extension des approches collaboratives au paradigme régions implique de pouvoir mettre en correspondance les segments extraits des différentes images.

4.1.3 Mise en correspondance des objets d'intérêt issus des différents niveaux

Dans le contexte applicatif de l'analyse d'environnements urbains, les objets d'intérêt relatifs aux différents niveaux d'analyse sont organisés de façon hiérarchique. Par exemple, un quartier urbain se décompose en général en un ensemble de blocs urbains de divers types, eux-mêmes pouvant se décomposer en ensembles d'objets urbains simples, formant ainsi une hiérarchie multiniveau. L'étude de ces relations hiérarchiques est d'un grand intérêt pour les experts, qui peuvent utiliser ces relations pour mieux comprendre la composition des environnements urbains à différentes échelles ainsi que les relations existant entre ces dernières.

En conséquence, il semble judicieux de proposer une méthodologie d'analyse multiniveau permettant d'obtenir une correspondance hiérarchique entre les segments issus des différents niveaux. Cette motivation applicative rejoint ici la motivation méthodologique présentée dans le paragraphe précédent, concernant l'utilisation d'approches collaboratives pour le traitement de données multirésolutions. Ainsi, garantir cette mise en correspondance peut permettre (1) d'offrir à l'utilisateur des informations de décomposition multiniveaux permettant une meilleure compréhension de l'organisation spatiale des structures d'intérêt recherchées, et (2) de faire collaborer et remettre en cause les résultats d'analyse obtenus aux différentes résolutions.

4.2 Proposition

Face à ces différentes motivations, nous présentons ici le cœur de notre travail. Nous introduisons par la suite un processus collaboratif produisant des résultats permettant une analyse multiniveau de la scène géographique observée. La principale originalité de ce processus est de s'appuyer sur le traitement d'images multirésolutions, afin d'améliorer les résultats de classification produits.

4.2.1 Un processus collaboratif basé régions pour le traitement d'images multirésolutions

Pour extraire des hiérarchies d'objets complexes, permettant d'obtenir une analyse à plusieurs niveaux de la scène observée, nous proposons d'étendre les approches collaboratives au paradigme basé régions pour l'analyse d'images multirésolutions. Pour ce faire, notre solution consiste à combiner une approche hiérarchique basée régions à une approche de classification non-supervisée, conduisant à une méthodologie hybride. Le principe de cette dernière est de faire collaborer ces deux approches à travers la résolution pour traiter des images à différentes résolutions spatiales.

Notre hypothèse principale repose sur le fait que le traitement de ces ensembles d'images permettra d'apporter une solution aux problèmes soulevés par l'utilisation d'approches hiérarchiques basées régions pour l'extraction de hiérarchies d'objets d'intérêt complexes à partir d'images (T)HRS. De plus, l'utilisation d'approches collaboratives permettra, par la suite, une remise en cause des résultats par la recherche de consensus multirésolutions.

Pour garantir une correspondance entre les segments des différentes résolutions (et ainsi, entre les différents niveaux d'objets d'intérêt considérés), nous proposons une stratégie permettant de construire les segments extraits, directement à travers la résolution spatiale.

4.2.2 Mise en correspondance des segments à travers les différentes images

A. Construire les segments à travers la résolution

Partant du constat que les différentes catégories d'objets d'intérêt recherchés s'organisent en une hiérarchie dont les niveaux sont en correspondance directe par des relations d'inclusion (*e.g.*, un blocs urbain est inclus dans un quartier urbain), on peut faire l'hypothèse que les segments issus des images (correspondant à ces objets d'intérêt) sont organisés de manière similaire à travers la résolution.

Pour obtenir une correspondance directe entre les segments des différentes résolutions, nous proposons par analogie aux approches de segmentation hiérarchique basées régions, de construire conjointement les différents niveaux de segments à travers les images. Pour ce faire, une solution consiste à partitionner une image donnée en un ensemble de segments pouvant correspondre à des objets d'intérêt de la scène puis à projeter cette segmentation dans une image à une résolution inférieure (ou supérieure) afin de guider le processus de segmentation de cette image. Notre idée consiste ainsi à s'appuyer sur les différentes résolutions disponibles pour construire les différents niveaux de segments liés aux images. Par ailleurs, cette stratégie peut permettre de faciliter la tâche de segmentation : l'homogénéité radiométrique des segments extraits des images MRS peut par exemple faciliter l'extraction des segments issus des images (T)HRS (et réciproquement).

De fait, par construction, les segments issus d'une image à une résolution particulière sont en correspondance directe avec les segments issus d'une image à une autre résolution. Une telle stratégie permet de s'abstraire des problèmes relatifs à l'utilisation d'approches de segmentation collaboratives entre les résolutions.

B. Approches ascendantes et descendantes à travers la résolution

Pour construire directement ces hiérarchies à travers la résolution, deux stratégies sont envisageables : les approches ascendantes et les approches descendantes.

Les auteurs de [Gaetano et al., 2009] ont ainsi proposé une méthode de segmentation hiérarchique multirésolution permettant d'extraire, à partir de ces images, des segments représentant des objets complexes. Basé sur une approche ascendante à travers la résolution, l'algorithme proposé commence par segmenter les

données à la résolution la plus fine, produisant une sur-segmentation dans le but de préserver les détails géométriques les plus élémentaires. Cette sur-segmentation initiale produit ainsi un nombre important de segments élémentaires qui sont ensuite fusionnés hiérarchiquement, en utilisant des informations radiométriques et géométriques des images aux différentes résolutions. Cette approche a fourni des résultats prometteurs et permet de mettre en correspondance les segments des images entre les résolutions considérées. Cependant, elle n'exploite pas toutes les informations portées par les images aux résolutions les plus basses.

En adoptant une stratégie opposée (*i.e.*, analyser l'image à la résolution la plus grossière puis affiner progressivement cette résolution), il est possible d'extraire et de classifier des segments correspondant à des objets complexes, tout en s'abstrayant du bruit sémantique induit par les détails [Goffe et al., 2011]. Cette stratégie, qui est aussi comparable au principe de la vision humaine [Mallat, 1996], a déjà été considérée dans [Sun et al., 2003] pour la création de cartes thématiques à partir d'images HRS et MRS.

4.2.3 Une stratégie collaborative descendante à travers la résolution

Pour mettre en œuvre cette stratégie, nous proposons de faire collaborer une approche de segmentation à une approche de classification non-supervisée afin de construire directement des hiérarchies de segments d'une manière descendante à travers la résolution. Le principe général consiste à extraire des segments grossiers de l'image à la plus faible résolution, puis à les décomposer récursivement et itérativement en segments plus fins en utilisant des images à des résolutions plus élevées.

L'utilisation d'une approche multirésolution descendante permet de diminuer les problèmes liés à l'analyse d'images (T)HRS. Elle permet à l'utilisateur de restreindre la quantité de données traitées (*e.g.*, par un « zoom » progressif) et ainsi de diminuer les problèmes liés à la complexité mémoire des données analysées.

Cette proposition permet de construire et de classifier des segments, à travers la résolution, pouvant être mis en correspondance d'une image à une autre. Une telle propriété rend possible, par la suite, la mise en place d'une collaboration entre les régions issues des différentes images. En conséquence, le processus proposé peut s'inscrire dans le cadre de la méthodologie globale, introduite dans la partie précédente (Section 3.3.2), dont l'objectif est la remise en cause de la segmentation et de la classification par l'analyse conjointe des différentes résolutions.

4.2.4 Avantages

Notre proposition combine ainsi plusieurs caractéristiques que nous récapitulons ci-après :

1. **une approche hiérarchique basée régions** : cette approche permet d'extraire des hiérarchies de segments permettant d'obtenir des résultats pouvant conduire à une analyse multiniveau ;
2. **le traitement d'ensembles d'images multirésolutions** : cette caractéristique permet de construire les segments en utilisant l'ensemble des données disponibles, et de maximiser le potentiel offert par leur complémentarité pour faciliter l'extraction des objets d'intérêt complexes ;
3. **une approche descendante à travers la résolution** : cette stratégie permet, par construction, de garantir une mise en correspondance directe des segments à travers la résolution et de réduire les problèmes mémoire liés à la complexité des données analysées ;
4. **une stratégie collaborative entre les différentes images** : cette stratégie permet d'éviter la fusion des images et de remettre en cause conjointement les résultats obtenus aux différentes résolutions ;
5. **l'utilisation d'approches non-supervisées** : par ce biais, ce processus collaboratif descendant ne nécessite que peu de connaissances *a priori* relatives au contenu des images traitées.

4.3 Schéma méthodologique

4.3.1 Principe

Notre proposition a conduit au développement d'un processus collaboratif hiérarchique basé régions, descendant à travers la résolution. Ce processus traite n images d'une même scène à différentes résolutions

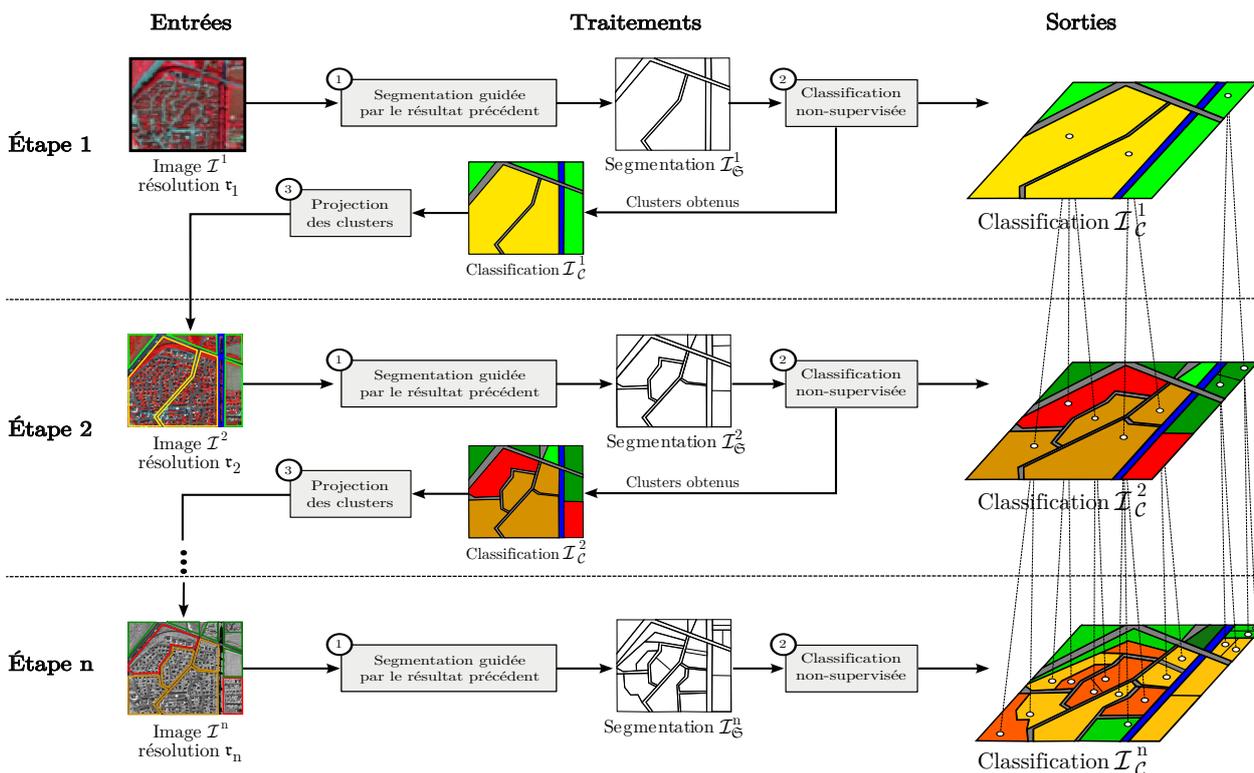


FIGURE 4.1 – Schéma du processus collaboratif basé régions pour l'analyse d'images multirésolutions.

spatiales, en partant de la plus faible pour aller jusqu'à la plus élevée, donnant lieu à n résultats permettant n niveaux différents d'interprétation. Il fonctionne en n étapes successives, chaque étape analysant une seule résolution, et produit n résultats de segmentation et de classification. Le principe consiste à extraire des segments grossiers de l'image à la plus faible résolution, puis à les décomposer récursivement en segments plus fins, en utilisant des images à des résolutions plus élevées.

Chacune des n étapes est composée de trois sous-étapes, reposant sur :

1. une approche de segmentation guidée par le résultat d'analyse à la résolution précédente ;
2. une approche de classification des régions ;
3. une projection des clusters obtenus dans la résolution suivante.

Pour traiter des ensembles d'images multirésolutions, ces trois sous-étapes sont appliquées d'une manière successive à travers la résolution : à chaque étape, le résultat de classification obtenu à la résolution précédente est projeté dans la résolution courante et est traité comme entrée de la méthode pour guider le processus de segmentation. Le principe est alors de segmenter ces clusters de régions dans la résolution courante afin de les décomposer en ensembles de segments plus fins. Cette stratégie permet de guider le processus de segmentation à la résolution courante par le résultat de classification issu de la résolution précédente, garantissant ainsi une correspondance directe entre les segments obtenus aux différentes résolutions. La figure 4.1 illustre ce principe de collaboration descendante.

Ce processus collaboratif de segmentation–classification à travers la résolution sera désigné par la suite par l'acronyme PCSC (Processus Collaboratif de Segmentation–Classification). La prochaine section introduit les trois sous-étapes proposées dans ces travaux.

4.3.2 Présentation des briques méthodologiques

Le processus PCSC est un processus générique : il peut supporter l'utilisation de différentes approches de segmentation ou de classification, que nous qualifierons par la suite de « briques de base ». Dans le cadre de ces

travaux, nous avons instancié ces dernières par trois approches adaptées à nos besoins applicatifs, fortement liés à l'analyse multiniveau d'environnements urbains.

A. Une approche de segmentation hiérarchique basée sur des exemples

Pour instancier l'approche de segmentation guidée par le résultat d'analyse obtenu à la résolution précédente, nous avons défini une approche de segmentation hiérarchique basée régions reposant sur la notion d'arbres binaires de partitions (BPT), issue de la morphologie mathématique [Salembier et Garrido, 2000]. Les feuilles des arbres correspondent aux zones élémentaires de l'image, tandis que les nœuds correspondent aux segments issus des fusions de ces segments élémentaires. Un BPT modélise ainsi une hiérarchie entre ces segments *via* une structure d'arbre. Une telle structure permet de représenter un ensemble de segments à différentes échelles et ses nœuds fournissent une bonne estimation des objets d'intérêt de la scène.

Une fois le BPT construit, il est nécessaire de définir une coupe au sein de cette hiérarchie pour obtenir une partition de l'image. Parallèlement, dans le cadre de l'analyse d'environnements urbains à partir d'images (T)HRS, différents objets d'intérêt peuvent être extraits à différentes échelles, rendant ainsi complexe la définition d'une telle coupe. Nous nous sommes ainsi intéressés, dans ces travaux, à définir un mécanisme permettant d'apporter une solution à ce problème. Pour ce faire, la méthodologie proposée implique une approche de segmentation interactive basée sur les BPT, définie interactivement par l'utilisateur sur une partie de l'image (sous la forme d'exemples de coupes), et automatiquement reproduite sur le reste des données.

B. Une approche de classification multirésolution

Une fois les segments extraits à partir d'une image à une résolution particulière par l'étape de segmentation définie précédemment, il est nécessaire de les catégoriser. Pour instancier l'approche de classification nécessaire à la catégorisation des segments, nous proposons une approche de classification non-supervisée multirésolution.

Le principe de cette approche est de caractériser les segments à classifier en prenant en compte leurs décompositions à travers la résolution. Pour ce faire, ces segments peuvent être caractérisés par des attributs multirésolutions, modélisés sous la forme d'histogrammes de composition. Une telle caractérisation offre une meilleure description des régions à classifier et permet ainsi d'améliorer les résultats de classification produits par le processus d'analyse collaboratif proposé.

C. Projection des clusters obtenus à travers la résolution

Pour traiter des ensembles d'images multirésolutions et permettre une correspondance directe entre les segments issus des différentes résolutions, nous proposons de faire collaborer à travers la résolution les approches de segmentation et de classification décrites ci-avant, d'une manière descendante.

À chaque étape $t \in n$, le résultat de l'étape précédente $t - 1$ (un ensemble de régions regroupées en différents clusters) est projeté dans la résolution courante et traité comme entrée de la méthode. Le principe est alors de décomposer (*i.e.*, segmenter puis classifier) ces différents clusters dans la résolution courante afin de former des segments plus fins. Pour ce faire, à chaque étape t , l'approche de segmentation est appliquée afin de partitionner l'ensemble des différents clusters fournis par l'étape $t - 1$. Une fois que toutes les régions composant ces clusters ont été décomposées, il est possible de créer une partition globale de l'image courante en regroupant tous les segments résultant de cette étape. Ces derniers sont ensuite caractérisés et classifiés par l'approche de classification multirésolution en différents ensembles homogènes de régions partageant des caractéristiques similaires. Ce résultat est alors projeté dans l'image suivante (de résolution plus fine) pour que le processus y soit ré-appliqué dans le but de décomposer ces régions pour obtenir un niveau d'analyse encore plus fin.

Les deux approches principales composant ce processus ((1) l'approche de segmentation basée sur des exemples et (2) l'approche de classification multirésolution) sont détaillées dans les chapitres 5 et 6. Finalement, les détails méthodologiques du processus PCSC sont énoncés dans le chapitre 7.

Segmentation hiérarchique d'images basée sur des exemples

Sommaire

5.1	Segmentation à base d'arbres binaires de partitions	57
5.1.1	Arbre binaire de partitions	57
5.1.2	Construction d'arbres binaires de partitions	58
5.1.3	Obtenir une segmentation à partir d'un arbre binaire de partitions	60
5.2	Le problème de l'échelle	62
5.2.1	Des objets d'intérêt à plusieurs échelles	62
5.2.2	Diviser l'étape de segmentation	63
5.3	Apprentissage et reproduction d'une coupe	64
5.3.1	Approche naïve	65
5.3.2	Approche par apprentissage	65
5.4	Bilan scientifique	68
5.4.1	Contributions	68
5.4.2	Valorisation scientifique	69

Ce chapitre présente l'approche de segmentation hiérarchique d'images proposée au cours de cette thèse. Cette approche de segmentation est basée sur la notion d'arbres binaires de partitions (BPT) et implique un mécanisme reposant sur des exemples de segmentations. Pour ce faire, nous commençons par détailler les principes de construction d'un BPT (Section 5.1). Nous montrons ensuite que la segmentation d'images (T)HRS, par le biais de BPT, peut engendrer des problèmes d'échelles (Section 5.2). Finalement, nous présentons une solution à ce problème, impliquant un mécanisme interactif de définition d'exemples de coupes à travers des BPT définis pour des zones spécifiques de l'image, et un mécanisme automatique de reproduction de ces derniers à travers les BPT associés au reste de l'image (Section 5.3). Dans une dernière section (Section 5.4), nous énonçons le bilan scientifique de ces travaux.

5.1 Segmentation à base d'arbres binaires de partitions

5.1.1 Arbre binaire de partitions

Soit $\mathcal{I} : E \rightarrow V$ une image. Un arbre binaire de partitions (BPT) [Salembier et Garrido, 2000] d'une image \mathcal{I} est une structure de données fournissant une hiérarchie de régions de E relativement à la fonction définie par \mathcal{I} . Ainsi, un BPT de \mathcal{I} est un couple (\mathcal{N}, φ) tel que $\mathcal{N} \subseteq 2^E$ est un ensemble de sous-ensembles de E vérifiant $E \in \mathcal{N}$, et $\varphi : \mathcal{N} \setminus \{E\} \rightarrow \mathcal{N}$ est une fonction vérifiant la propriété suivante :

Propriété Pour tout $N \in \mathcal{N} \setminus \{E\}$, on a $\varphi^{-1}(\{N\}) = \{N_1, N_2\}$ tel que $N_1 \neq N_2 \in \mathcal{N}$ et $N = N_1 \sqcup N_2$.

Les éléments de \mathcal{N} sont nommés les nœuds du BPT. La fonction φ modélise la relation de parenté entre les nœuds : plus précisément, si $N = \varphi(N')$, alors N (respectivement N') est le père (respectivement le fils) de N' (respectivement N). Le nœud E est la racine du BPT. Les nœuds de $\mathcal{N} \setminus \varphi(\mathcal{N} \setminus \{E\})$ (*i.e.*, ceux n'ayant pas d'enfant) sont nommés les feuilles du BPT.

D'une manière plus pratique, φ permet de diviser récursivement E en plusieurs partitions, obtenues successivement en divisant exactement un élément de la partition courante en deux sous-ensembles. En particulier, l'ensemble $\mathcal{N} \setminus \varphi(\mathcal{N} \setminus \{E\})$ (respectivement $\{E\}$) constitue la partition la plus fine (respectivement la plus grossière) de E relativement à la fonction φ .

Chaque sous-ensemble $C \subseteq \mathcal{N}$ de nœuds tel que C soit une partition de E est appelé une coupe de l'arbre. D'une manière plus pratique, les nœuds de C induisent un sous-arbre du BPT initial, de racine E et dont les feuilles forment C (cet arbre étant aussi un BPT).

5.1.2 Construction d'arbres binaires de partitions

Le BPT d'une image $\mathcal{I} : E \rightarrow V$ est construit d'une manière ascendante (*i.e.*, de ses feuilles à sa racine). D'un point de vue algorithmique, partant d'une segmentation initiale de E (généralement composée de tous les singletons $\{\mathbf{x}\}$, pour tout $\mathbf{x} \in E$, ou par les zones plates de \mathcal{I}), les nœuds de \mathcal{N} (et ainsi φ) sont définis successivement par la fusion de couples de nœuds de \mathcal{N} pour lesquels φ n'a pas encore été définie. Dans le contexte de la segmentation d'images, de tels couples de nœuds sont généralement choisis comme adjacents spatialement, conduisant ainsi à des nœuds connexes dans \mathcal{N} .

De nombreux BPT peuvent être obtenus à partir d'une unique partition initiale de E . Dans le but de décider quel arbre parmi ceux-ci est le plus intéressant, il est nécessaire de définir un « ordre de fusion » (pour décider de la priorité de fusion parmi les couples de nœuds adjacents). La construction d'un BPT conduit ainsi à deux notions principales : le *modèle de région* (qui spécifie comment les segments/nœuds sont caractérisés/représentés), et le *critère de fusion* (qui définit la similarité entre les nœuds voisins ainsi que leur ordre de fusion).

A. Ajout d'un biais applicatif

Les modèles et critères de base, utilisés dans la plupart des approches de segmentation par fusion de segments, sont généralement fondés sur l'homogénéité radiométrique des nœuds [Garrido et al., 1998]. Cependant, les choix du modèle de région et du critère de fusion peuvent être influencés par les connaissances *a priori* de l'utilisateur relatives aux données traitées et aux objets d'intérêt recherchés. Ainsi, il peut être pertinent d'induire un biais applicatif dans la définition de ces éléments, afin d'obtenir des résultats de segmentation reflétant mieux les attentes de l'utilisateur. Ces considérations peuvent permettre à ce dernier d'influencer la construction d'un BPT particulier en fonction de ses connaissances sur le contenu de l'image traitée (*e.g.*, pour segmenter des surfaces de végétation, il n'est généralement pas nécessaire de considérer les propriétés géométriques des régions).

Lors du traitement d'images à (très) hautes résolutions spatiales, centrées sur des scènes urbaines, les détails géométriques doivent être pris en considération. Les zones urbaines sont en effet souvent caractérisées par une forte organisation géométrique des objets d'intérêt autour d'éléments linéaires (*e.g.*, routes, chemins de fer, rivières). En conséquence, nous proposons pour caractériser les nœuds à fusionner, de considérer leurs propriétés radiométriques et leurs propriétés géométriques (aires et élongation) dans le but de fusionner, en priorité, des segments correspondant à des objets d'intérêt qui ne structurent pas la scène. Cette stratégie permet de conserver, jusqu'aux niveaux les plus hauts de la hiérarchie du BPT, une organisation géométrique des segments correspondant aux objets d'intérêt, autour des structures linéaires de la scène représentée.

Nous définissons par la suite, le modèle de région et le critère de fusion proposés dans ces travaux. Pour définir ces critères, il a été choisi de ne prendre en compte que les propriétés bas niveau des segments à fusionner, car nous considérons que les informations de haut niveau (sémantique) sont fournies dans cette approche par l'utilisateur, *via* ses exemples de segmentations.

B. Modèle de région

Un nœud (*i.e.*, un segment/une région de l'image) $S_i \in \mathcal{N}$ ($S_i \subseteq E$) est modélisé ici par un couple de valeurs

$$\begin{aligned} M_r(S_i) &= \langle (v_b^-(S_i), v_b^+(S_i)) \rangle_{b=1}^s \\ M_g(S_i) &= (e(S_i), a(S_i)) \end{aligned} \tag{5.1}$$

où v_b^* représente les valeurs extrêmes pour la bande spectrale b relativement à \mathcal{I} (*i.e.*, relativement à \mathcal{I}_b), tandis que e et a représentent respectivement les valeurs moyennes d'élongation et d'aire. Plus généralement, M_r et M_g représentent les informations radiométriques et géométriques des nœuds. Durant le processus de fusion, le modèle de région de deux nœuds fusionnées S_i et S_j est alors fourni par

$$\begin{aligned} M_r(S_i \cup S_j) &= \langle (\min\{v_b^-(S_i), v_b^-(S_j)\}, \max\{v_b^+(S_i), v_b^+(S_j)\}) \rangle_{b=1}^s \\ M_g(S_i \cup S_j) &= (e(S_i \cup S_j), a(S_i) + a(S_j)) \end{aligned} \quad (5.2)$$

Par opposition à M_r et a , dont le calcul est trivial, le calcul de l'élongation e nécessite le pré-calcul d'une carte d'élongation associée à \mathcal{I} (qui va permettre de mettre en évidence les structures linéaires) divisant ainsi E en (larges) zones. La détection de structures à fortes valeurs d'élongation a conduit à de nombreux travaux. Un exemple caractéristique d'approche est présenté dans [Valero et al., 2010a]. Ces travaux présentent une approche *ad hoc*, basée sur la morphologie mathématique, pour la détection de routes à partir d'images satellites (T)HRS.

Notre objectif, dans ces travaux, n'est pas d'obtenir les meilleurs résultats d'élongation possibles, mais d'obtenir une première information intuitive sur l'élongation des segments, représentant les objets géographiques de la scène, avec un faible coût de calcul. Pour ce faire, nous avons défini une heuristique pour générer la carte d'élongation e . Nous décrivons ci-dessous l'approche utilisée :

- (1.) pour chaque pixel $\mathbf{x} \in E$ (considéré comme une graine), une série de segmentations par croissance de régions (basées sur l'intensité radiométrique) est calculée avec un seuil de tolérance λ croissant $\lambda \in \llbracket 0, v_b^+(E) - v_b^-(E) \rrbracket$;
- (2.) pour chaque segment de chaque segmentation résultante, une valeur d'élongation est obtenue en calculant le rapport longueur/largeur de la meilleure boîte englobante du segment considéré (calculée pour différentes orientations discrètes) ;
- (3.) la meilleure valeur d'élongation (*i.e.*, la plus élevée) obtenue pour le segment est ensuite assignée à \mathbf{x} .

Cette approche présente un coût algorithmique borné, pour chaque pixel, par l'aire du voisinage où l'étape (1.) est réalisée (qui, en pratique, ne nécessite pas d'être élevée). Le calcul de la carte d'élongation est donc globalement linéaire par rapport à la taille de E . La figure 5.1 présente un exemple d'une carte d'élongation, obtenue grâce à l'heuristique proposée ci-dessus, et calculée sur une image HRS ayant une résolution spatiale de 2,4 m.

C. Critère de fusion

À chaque étape, l'algorithme détermine la paire de segments voisins les plus similaires¹. La paire de segments choisie est celle dont le segment potentiel résultant de cette fusion présente des propriétés minimisant l'augmentation des intervalles des valeurs d'intensités radiométriques (pour chaque bande spectrale) et de faibles caractéristiques d'élongation et d'aire. Cette définition conduit aux critères de fusion suivants

$$\begin{aligned} O_r(S_i, S_j) &= \frac{1}{s} \sum_{b=1}^s \frac{\max\{v_b^+(S_i), v_b^+(S_j)\} - \min\{v_b^-(S_i), v_b^-(S_j)\}}{v_b^+(E) - v_b^-(E)} \\ O_g(S_i, S_j) &= \frac{1}{2}(e(S_i \cup S_j) + a(S_i \cup S_j)) \end{aligned} \quad (5.3)$$

où les valeurs $e(S_i \cup S_j)$ et $a(S_i \cup S_j)$ sont normalisées en utilisant les valeurs maximales possibles de ces caractéristiques.

La valeur de similarité entre deux segments adjacents S_i et S_j peut ensuite être calculée comme

$$O(S_i, S_j) = \alpha.O_r(S_i, S_j) + (1 - \alpha).O_g(S_i, S_j) \quad (5.4)$$

où $\alpha \in [0, 1]$. Il est à noter que les critères de fusion $O_r(S_i, S_j)$ et $O_g(S_i, S_j)$ sont normalisés par construction.

En pratique, plus les nœuds sont proches de la racine, moins le critère de fusion O_r est pertinent. En effet, plus la taille d'un nœud est importante, et plus celui-ci est radiométriquement hétérogène (conduisant à une

1. On pourra également citer [Felzenszwalb et Huttenlocher, 2004] comme travaux parallèles sur ce type de mesures de similarité.



(a) Image satellite HRS (2,4 m).

(b) Carte d'élongation e .

FIGURE 5.1 – Calcul d'une carte d'élongation. (a) Image de référence. (b) Carte d'élongation e correspondante : les structures à forte (respectivement faible) élongation apparaissent en gris clair (respectivement gris foncé).

valeur de O_r ne permettant pas de caractériser correctement ce nœud). En conséquence, le poids α peut être défini comme une fonction dépendant directement de la valeur de O_r (décroissante quand O_r augmente). En particulier, nous avons observé expérimentalement qu'une formulation gaussienne standard

$$\alpha(O_r) = \exp(-O_r^2) \quad (5.5)$$

fournit un comportement satisfaisant pour le critère de fusion O . De plus, cette formulation simple permet à l'utilisateur d'adapter cette fonction par l'introduction de paramètres dans $\alpha(O_r)$ influant sur le comportement asymptotique de α et la valeur de O_r pour laquelle $\alpha = 1 - \alpha$.

Un BPT peut ainsi être produit pour une image *via* l'utilisation de ce modèle de région et de ce critère de fusion. La figure 5.2 illustre un exemple de construction d'un BPT.

5.1.3 Obtenir une segmentation à partir d'un arbre binaire de partitions

Une fois le BPT construit, il est nécessaire de définir une coupe au sein de cette hiérarchie pour obtenir une partition de l'image et ainsi un découpage des objets d'intérêt de la scène. Nous avons vu précédemment qu'un nombre conséquent de coupes peuvent être extraites de la hiérarchie de partitions résultante. Ainsi, il

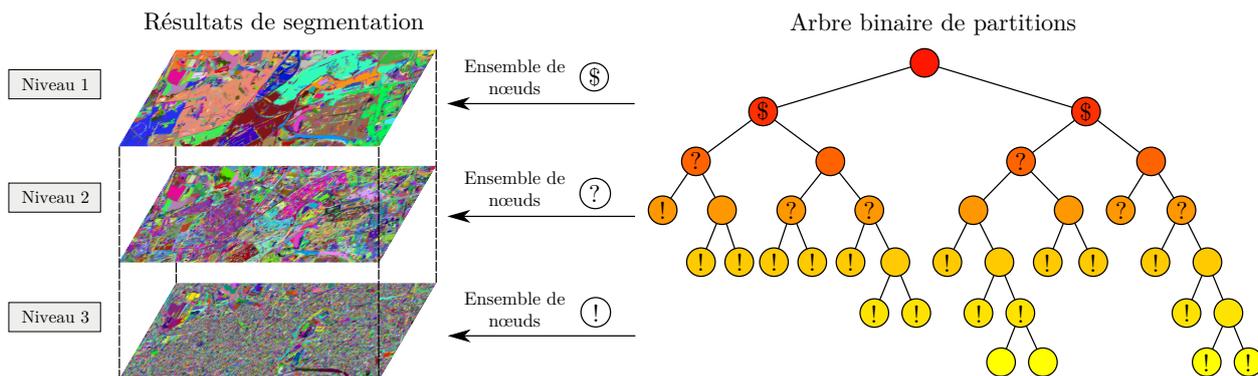


FIGURE 5.2 – Un exemple de BPT associé à l'image HRS présentée par la figure 5.1(a) (pour des raisons de lisibilité, le nombre de nœuds a été artificiellement réduit). Les nœuds de \mathcal{N} sont représentés par des disques colorés (la racine E correspond au nœud le plus haut). La fonction φ est modélisée par les couples d'arêtes (reliant deux nœuds N_1, N_2 à leur nœud parent commun $\varphi(N_1) = \varphi(N_2)$). Les couleurs des nœuds (du jaune au rouge) symbolisent la confiance accordée à la valeur de similarité O_r entre deux régions adjacentes (et ainsi, symbolisent la décroissance de la fonction α contrôlant la pondération entre O_r et O_g). Pour des raisons de lisibilité, trois partitions de E (associées à trois coupes du BPT) sont représentées.

est nécessaire de définir un processus permettant de choisir une coupe pertinente par rapport à l'application considérée.

De nombreuses approches ont été ainsi définies pour aider l'utilisateur à choisir une coupe pertinente. Parmi ces dernières on distingue des approches automatiques ou des approches impliquant une interaction de l'utilisateur. Nous décrivons par la suite certaines de ces approches.

A. Approches automatiques

Le principe des approches automatiques est de rechercher au sein d'une hiérarchie de partitions, une segmentation dont les caractéristiques minimisent un ou plusieurs critères associés aux segments considérés. Dans de nombreux travaux, ces critères sont basés sur des propriétés d'homogénéité [Akay et Aksoy, 2008; Serra, 2011] des segments issus des partitions. Dans [Plaza et Tilton, 2005], les auteurs proposent une approche recherchant une segmentation dans la hiérarchie dont les segments respectent des critères radiométriques et géométriques particuliers. Cette approche est basée sur un mécanisme qui permet d'analyser le comportement des différentes segmentations de la hiérarchie en fonction de leur niveau de détail.

Les approches automatiques sont généralement bien adaptées pour l'extraction de niveaux de segmentation relativement homogènes. Notre contexte applicatif étant principalement centré sur l'extraction d'environnements urbains à partir d'images (T)HRS, nous cherchons à extraire des segments correspondant à des objets d'intérêt complexes et fortement hétérogènes. Les premières expériences menées par notre équipe de recherche ont montré que l'utilisation de telles approches ne fournissait pas des résultats pertinents dans le cadre de l'extraction de ce type d'objets complexes car ces dernières sont essentiellement basées sur des hypothèses d'homogénéité radiométrique.

B. Approches interactives

Le principe des approches interactives est d'utiliser des connaissances ou une interaction de l'utilisateur pour définir une coupe au sein de la hiérarchie de partitions. La plupart de ces approches sont basées sur l'utilisation de marqueurs [McGuinness et O'Connor, 2010; Tarabalka et al., 2011; Passat et al., 2011]. Le principe est de demander à l'utilisateur de définir interactivement sur l'image un marqueur sur l'objet d'intérêt recherché. Ensuite, un processus est utilisé pour trouver la partition de la hiérarchie permettant de segmenter au mieux l'objet d'intérêt sélectionné au cours de l'étape précédente. Dans le cadre de ces travaux, nous recherchons à extraire de l'image considérée un nombre potentiellement important d'objets d'intérêt. Ainsi,

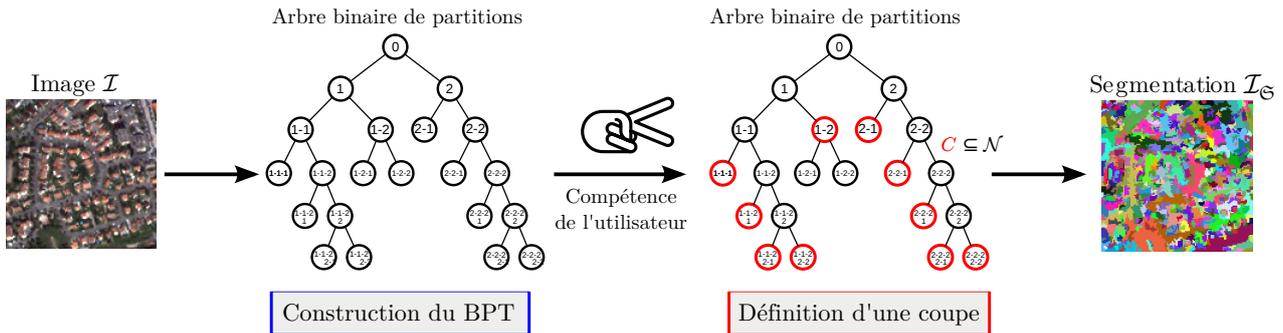


FIGURE 5.3 – Illustration de l'extraction interactive d'une coupe à partir d'un BPT.

de telles approches ne sont pas forcément pertinentes pour l'utilisateur.

Dans ces travaux, nous proposons d'utiliser une approche interactive pour aider l'utilisateur à choisir une coupe pertinente qui soit adaptée à son besoin applicatif. Nous avons ainsi développé une interface logicielle permettant de parcourir interactivement le contenu de l'arbre afin de trouver une partition pertinente. Une coupe dans un BPT peut être obtenue en réalisant un seuillage sur la mesure de similarité induite par la fonction O et attachée à chaque nœud de l'arbre (en sauvegardant pour un nœud $S_i \in \mathcal{N}$, la valeur O nécessaire pour construire ce dernier), à la valeur induite par l'utilisateur.

Le parcours de cet arbre peut être effectué de manière globale ou locale :

1. **Parcours global** L'interface du logiciel développé est munie d'un curseur permettant de se placer à un certain niveau d'hétérogénéité dans l'arbre. Pour ce faire, pour chaque valeur particulière de O , on recherche la coupe dont l'ensemble des nœuds possède une valeur O inférieure ou égale à cette valeur particulière. Cette première stratégie permet de parcourir rapidement un ensemble des coupes possibles de l'arbre.
2. **Parcours local** Une fois qu'un premier niveau de coupe a été choisi par l'utilisateur, un raffinement local des nœuds de cette dernière est alors possible. L'interface du logiciel développé permet de raffiner les branches de l'arbre afin de sélectionner des nœuds permettant une meilleure extraction des objets d'intérêt. Pour ce faire, à partir de la coupe courante sélectionnée *via* le parcours global, l'utilisateur peut interactivement parcourir les parents et enfants des nœuds (par le biais des relations de parenté induites par la fonction φ) influant ainsi d'une manière locale sur le résultat de segmentation obtenu. De telles opérations peuvent être réalisées en sélectionnant directement les segments produits ou en manipulant la structure de l'arbre affichée dans une fenêtre indépendante.

Les travaux effectués dans le cadre de ce logiciel sont présentés en Annexe D. La figure 5.3 illustre le principe de cette approche.

5.2 Le problème de l'échelle

5.2.1 Des objets d'intérêt à plusieurs échelles

Dans le cadre de l'analyse d'images (T)HRS, et en particulier dans le cadre de l'analyse d'environnements urbains, différents objets d'intérêt peuvent être extraits à différentes échelles. En effet, les objets d'intérêt relatifs à un même niveau d'analyse (*i.e.*, de sémantique) peuvent apparaître comme plus ou moins complexes et hétérogènes au sein d'une même image. Par exemple, quand on analyse les environnements urbains principaux, les territoires urbanisés vont apparaître comme plus hétérogènes que les zones rurales ou les forêts alors que ces différents objets d'intérêt font partie du même niveau d'analyse (*i.e.*, niveau des quartiers urbains).

Ce problème d'échelle va alors perturber l'étape de segmentation et la rendre plus complexe. Dans la plupart des cas, une segmentation unique ne permet pas d'extraire des segments correspondant à tous les objets d'intérêt relatifs à un même niveau d'analyse. Lors de l'utilisation d'une segmentation hiérarchique, ces différents objets d'intérêt peuvent alors émerger à différents niveaux de la hiérarchie obtenue. Il apparaît alors

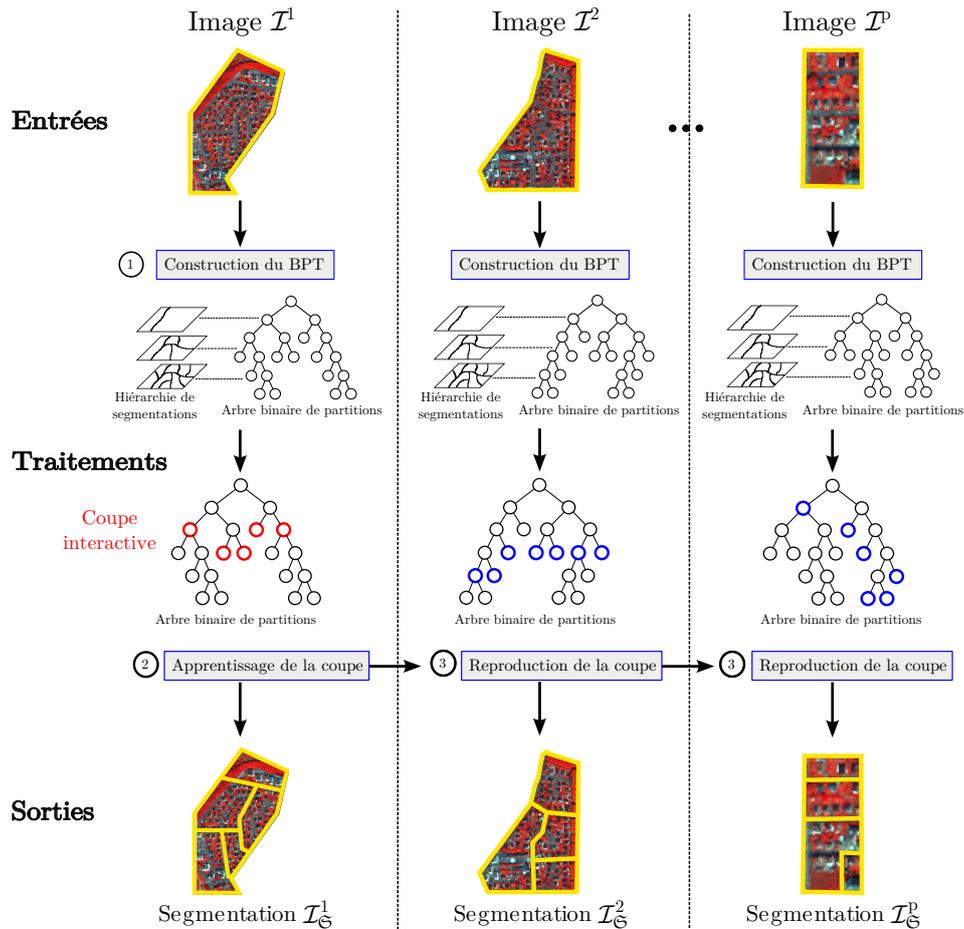


FIGURE 5.4 – Approche de segmentation « interactive », basée sur des exemples. En rouge : interactions de l'utilisateur. En bleu : traitements automatiques. En jaune : p imgettes représentant p zones différentes mais centrées sur le même type d'objets d'intérêt.

difficile de trouver une coupe unique dans une hiérarchie de segmentations, et en particulier dans un BPT, permettant d'extraire simultanément ces différents objets d'intérêt (liés à différents niveaux d'hétérogénéité).

Pour résoudre ce problème, une solution généralement employée consiste à utiliser plusieurs niveaux de la hiérarchie de segmentations afin d'extraire au mieux les différents objets d'intérêt recherchés. Par exemple, l'utilisateur va utiliser un premier niveau de la hiérarchie pour extraire les objets les plus homogènes (*e.g.*, zones de végétation urbaines) puis un second niveau pour extraire des objets plus hétérogènes (*e.g.*, blocs urbains, zones industrielles). Cependant, une telle approche pose deux problèmes majeurs. Le premier est lié au fait que l'utilisateur doit manuellement sélectionner, dans chaque niveau sélectionné de la hiérarchie, les différents segments qu'il souhaite en extraire. Ce processus est consommateur de temps et peut, de plus, engendrer des erreurs. Le second est lié au fait que les différents segments extraits de ces niveaux de la hiérarchie (correspondant aux objets d'intérêt recherchés), peuvent conduire à des phénomènes de superposition des partitions obtenues.

Nous proposons par la suite d'adapter la notion de segmentation par BPT pour pallier ce problème et aider l'utilisateur dans cette approche de segmentation multiniveau.

5.2.2 Diviser l'étape de segmentation

Pour prendre en compte ce phénomène d'échelle dans le processus de segmentation hiérarchique développé, nous proposons de diviser le processus de segmentation en un ensemble de sous-processus, chacun adapté à l'extraction d'une certaine catégorie d'objets d'intérêt (ayant un niveau d'échelle particulier).

Au lieu de segmenter l’intégralité d’une image en utilisant un seul niveau de la hiérarchie liée au BPT, l’idée principale est d’adapter l’approche de segmentation à des zones locales de l’image représentant le même type d’objets d’intérêt de même homogénéité radiométrique. Une approche similaire, proposée dans [Akçay et Aksoy, 2008], illustre cette stratégie. Une telle approche peut ainsi permettre de segmenter localement les différents objets d’intérêt de l’image (*i.e.*, en utilisant un niveau d’échelle locale), tout en évitant les problèmes liés à la superposition des résultats de segmentation obtenus.

Dans le cadre de ces travaux, nous proposons une approche permettant d’adapter la segmentation par BPT à ce paradigme. Le principe de cette approche est de construire plusieurs BPT pour une même image, chacun de ces BPT étant construit sur une zone particulière de l’image. Considérons une image $\mathcal{I} : E \rightarrow V$. Le support E de cette dernière peut être divisé en un ensemble de q imagettes E^i , centrées sur les différents objets d’intérêt que l’utilisateur souhaite extraire de la scène² (l’ensemble des q imagettes E^i forme ainsi une partition de E). Il est alors possible de construire un BPT pour chaque imagette ainsi définie (q BPT). Une fois ces arbres construits, on peut alors extraire une coupe spécifique par BPT, dans le but d’obtenir des niveaux de segmentation différents pour chacune des q imagettes. Comme nous l’avons énoncé précédemment, nous proposons d’utiliser une approche interactive pour aider l’utilisateur à choisir une coupe pertinente. Ainsi, une telle approche impliquerait de demander à l’utilisateur de choisir un niveau de coupe pour chacun des q BPT résultants (*i.e.*, q interactions avec l’utilisateur).

Pour faciliter cette tâche, nous proposons d’utiliser un mécanisme de reproduction automatique d’exemples de segmentations fournis par l’utilisateur. Pour ce faire, l’approche proposée implique un mécanisme interactif de définition d’exemples de coupes à travers des BPT définis pour des zones spécifiques de l’image, et un mécanisme automatique de reproduction de ces derniers à travers les BPT associés au reste de l’image. L’utilisateur peut ainsi choisir un niveau de coupe dans un BPT spécifique, puis ce niveau est automatiquement reproduit dans les BPT centrés sur des zones similaires de l’image. Il est alors possible d’utiliser les compétences de l’utilisateur pour choisir des niveaux d’échelle locaux (par le biais d’exemples de coupes), automatiquement reproduits dans le reste de l’image.

Cette approche de segmentation, basée sur des exemples, prend en entrée $p \geq 2$ imagettes (représentant p zones différentes mais centrées sur le même type d’objets d’intérêt, extraites des q zones définies précédemment) et retourne p segmentations (Figure 5.4). Pour l’une de ces p imagettes, un BPT est construit, puis l’utilisateur définit interactivement une coupe à travers cet arbre (produisant ainsi une segmentation). On fait alors l’hypothèse que cette segmentation caractérise le niveau d’échelle souhaité par l’utilisateur et qu’elle peut être utilisée comme exemple pour segmenter les $p - 1$ imagettes restantes. Ainsi, pour chacune de ces $p - 1$ imagettes, un BPT est construit, puis une coupe est automatiquement définie au sein de cet arbre, reproduisant alors la coupe définie précédemment par l’utilisateur sur l’imagette exemple.

Il est alors possible d’appliquer plusieurs fois cette approche, dans une même image, pour segmenter différents types d’objets ayant des sémantiques et des niveaux d’échelle différents, et pour traiter ainsi l’ensemble des q imagettes définies précédemment. En conséquence, toutes les imagettes centrées sur le même type d’objets d’intérêt sont segmentées d’une manière similaire. Une fois que toutes les q imagettes ont ainsi été décomposées, il est possible de créer une partition globale de l’image en regroupant tous les segments résultant de cette étape sans engendrer de problème de superposition.

5.3 Apprentissage et reproduction d’une coupe

La division du processus de segmentation en plusieurs sous-processus indépendants apporte une solution au problème de l’échelle, et permet ainsi d’éviter les problèmes liés à la complexité des images (T)HRS. La division du processus de segmentation en plusieurs sous-processus indépendants nécessite cependant la définition d’une approche d’apprentissage et de reproduction d’une coupe. Dans le cadre de ces travaux, deux approches de ce type ont été étudiées. La première est une approche naïve présentant un faible coût algorithmique. La seconde est une approche plus efficace et permet d’obtenir des résultats de segmentation plus pertinents. Cependant, elle présente un coût algorithmique plus élevé.

2. On considère dans ces travaux qu’une imagette E^i de E est une partie de E composée d’un ensemble de pixels connexes. Il est à noter que l’imagette E^i n’a pas forcément une forme rectangulaire.

5.3.1 Approche naïve

Dans une première approche (qualifiée de « naïve »), la coupe produite par l'utilisateur sur un BPT associé à une imagerie particulière (centrée sur un type particulier d'objets d'intérêt), est caractérisée par l'hétérogénéité radiométrique et géométrique qui lui est associée (induite directement par la valeur du critère de fusion O). Cette valeur d'hétérogénéité peut ensuite être utilisée pour automatiquement trouver une coupe de même niveau d'hétérogénéité dans les arbres liés aux images restantes.

L'approche d'apprentissage consiste ici à mémoriser la valeur O qui caractérise la coupe définie interactivement par l'utilisateur sur le BPT lié à l'imagerie utilisée comme exemple de segmentation. Le mécanisme de reproduction consiste alors à trouver, dans chacun des $p - 1$ BPT à segmenter, une coupe dont la valeur d'hétérogénéité est inférieure ou égale à la valeur de O mémorisée pendant l'étape d'apprentissage (*i.e.*, à seuiliser chacun des $p - 1$ BPT à la valeur de O mémorisée).

Cependant les résultats fournis par cette approche ne sont pas toujours pertinents. Ce problème est lié au fait que la valeur d'hétérogénéité d'une coupe O n'est pas suffisante pour décrire les caractéristiques de la coupe choisie par l'utilisateur. En particulier, une telle valeur « globale » d'hétérogénéité ne permet pas de considérer les gestes « locaux » réalisés par l'utilisateur lors du choix de la coupe conduisant à une segmentation pertinente.

Pour apporter une solution à ce problème, nous proposons une nouvelle approche permettant de mieux caractériser la coupe choisie par l'utilisateur et ainsi d'améliorer les étapes d'apprentissage et de reproduction de coupe. Cette nouvelle approche permet de reproduire, d'une manière plus réaliste, le geste de coupe produit par l'utilisateur.

5.3.2 Approche par apprentissage

Face aux différents problèmes de l'approche naïve présentée ci-dessus, nous proposons une approche basée sur un mécanisme d'apprentissage du geste de coupe de l'utilisateur. Pour ce faire, la coupe produite par ce dernier est *apprise* par le biais d'une classification non-supervisée, puis est reproduite sur l'ensemble des images restantes à segmenter.

A. Apprentissage du geste

Pour être capable de reproduire au mieux l'exemple de coupe défini par l'utilisateur au sein d'un BPT, dans les $p - 1$ BPT/images restantes³, il est dans un premier temps nécessaire d'apprendre cet exemple (*i.e.*, d'en extraire les propriétés qui le caractérisent, permettant ainsi sa reproduction). Nous détaillons par la suite la stratégie que nous avons développée pour apprendre le geste de coupe de l'utilisateur. La figure 5.5 illustre cette approche d'apprentissage.

Soit $C \subseteq \mathcal{N}$ la coupe exemple définie interactivement par l'utilisateur dans l'un des BPT (Figure 5.5-①). Dans le but d'extraire les nœuds les plus pertinents caractérisant cette coupe, il est tout d'abord nécessaire de séparer, au sein de cette dernière, les éléments linéaires des éléments qui correspondent à des zones « entourées » par ces derniers. Pour ce faire, C est partitionnée en deux sous-ensembles C_e et $C_{\bar{e}}$, correspondant aux nœuds/segments à fortes et faibles elongations, respectivement (Figure 5.5-②). Une telle partition peut être aisément obtenue par une classification non-supervisée binaire (*e.g.*, K -MEANS basé sur l'attribut e des nœuds). Les objets considérés par l'approche d'apprentissage proposée sont alors ceux de $C_{\bar{e}}$, qui correspondent aux zones de l'image « entourées » par les éléments linéaires de C_e . Les éléments de $C_{\bar{e}}$ correspondent généralement aux nœuds représentant les objets d'intérêt qui caractérisent le mieux une coupe choisie par un utilisateur (*e.g.*, blocs urbains, blocs industriels) tandis que les éléments de C_e correspondent généralement aux nœuds représentant des objets d'intérêt moins caractéristiques (*e.g.*, portions de routes, rivières).

Pour extraire les nœuds les plus pertinents caractérisant $C_{\bar{e}}$, un processus de partitionnement par classification non-supervisée est ensuite réalisé sur les éléments de $C_{\bar{e}}$ (Figure 5.5-③). Les segments liés aux éléments de $C_{\bar{e}}$ peuvent être caractérisés par le biais d'histogrammes colorimétriques (*i.e.*, chaque segment $S \in C_{\bar{e}}$ de $\mathcal{I} : E \rightarrow V$ (avec $S \subseteq E$) est caractérisé par un histogramme radiométrique normalisé $H(\mathcal{I}, S)$ construit par

3. Sans perte de généralité, l'approche permet aussi de définir plus qu'un seul exemple, dans le but d'améliorer l'étape d'apprentissage.

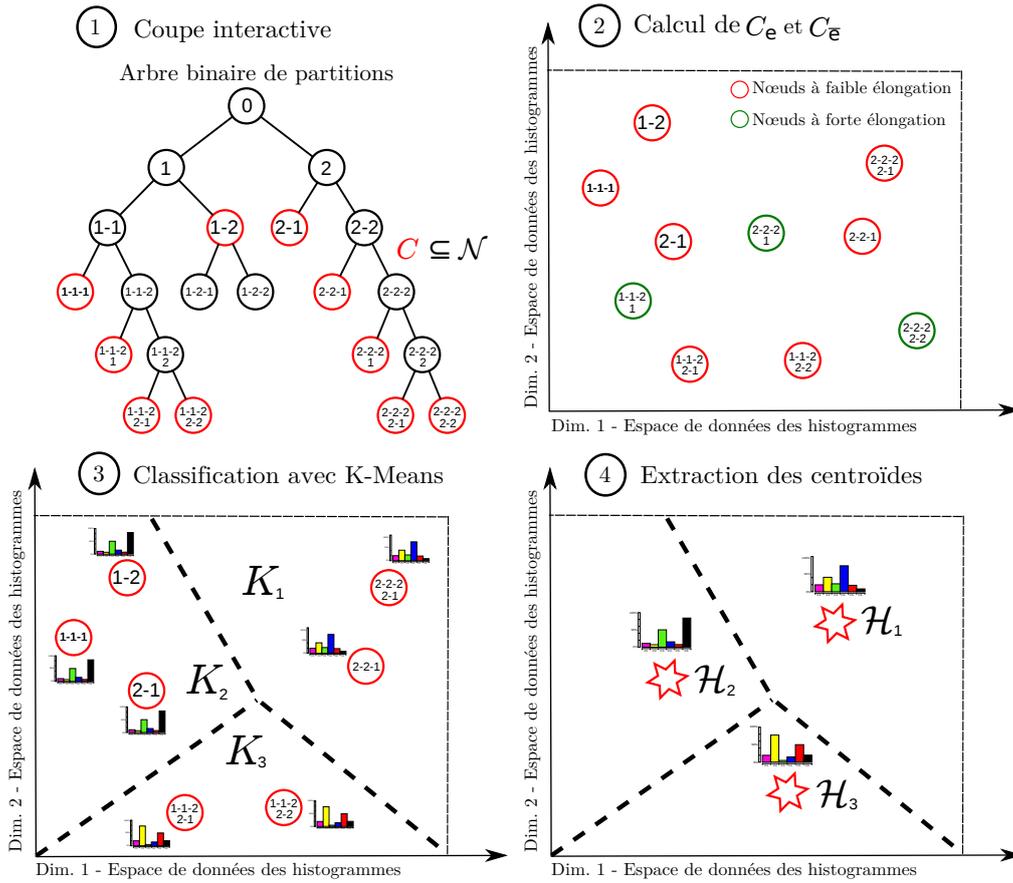


FIGURE 5.5 – Illustration de l'approche de segmentation basée sur des exemples : apprentissage de la coupe exemple.

le biais de la fonction $\mathcal{H}_{\mathcal{I}, S} : V \rightarrow \mathbb{N}$). Les approches de partitionnement par classification non-supervisée requièrent une mesure de distance (et une moyenne) pour comparer entre eux les objets à classifier. Par la suite, on notera \mathcal{D} la distance utilisée pour comparer ces histogrammes. Le nombre de clusters u , qui caractérise le nombre de familles de nœuds pertinentes composant $C_{\bar{e}}$ doit être fixé par l'utilisateur. Ce processus conduit à la création d'un ensemble de u clusters $\{K_i\}_{i=1}^u$, associé à un ensemble de u centroïdes. Chaque centroïde $i \in u$ est alors modélisé par l'histogramme moyen (normalisé) $H(\mathcal{I}, K_i)$ du cluster K_i (Figure 5.5-④).

Ces u centroïdes permettent ainsi de caractériser la coupe exemple définie par l'utilisateur au sein d'un BPT. Ils seront utilisés par la suite pour extraire automatiquement une coupe des $p - 1$ BPT restants.

B. Reproduire le geste

L'exemple de segmentation fourni par l'utilisateur est ainsi modélisé par les u centroïdes obtenus à partir de la coupe d'un BPT associé à l'une des p imagettes, ici l'imagette notée \mathcal{I} (Figure 5.5-④). Ces centroïdes peuvent ensuite être utilisés pour la segmentation automatique des $p - 1$ autres imagettes (Figure 5.6).

Ceci peut être réalisé en trouvant, pour chacune des $p - 1$ imagettes restantes \mathcal{I}^j ($j \in \llbracket 2, k \rrbracket$), une coupe \widehat{C}_j dans le BPT de \mathcal{I}^j , minimisant une mesure d'inertie calculée entre l'ensemble des nœuds centroïdes modélisés par $\{H(\mathcal{I}, K_i)\}_{i=1}^u$ et l'ensemble des nœuds de C_j (ou plus précisément, l'ensemble des histogrammes normalisés $\{H(\mathcal{I}^j, X)\}_{X \in C_j}$).

La mesure d'inertie $\zeta(C_j)$ associée à une coupe C_j , relativement à l'ensemble des centroïdes $\{H(\mathcal{I}, K_i)\}_{i=1}^u$, peut être définie comme

$$\zeta(C_j) = \sum_{i=1}^u \frac{|\cup_{X \in C_j^i} X|}{|\cup_{X \in C_j} X|} \cdot \mathcal{D}(H(\mathcal{I}^j, \cup_{X \in C_j^i} X), H(\mathcal{I}, K_i)) \quad (5.6)$$

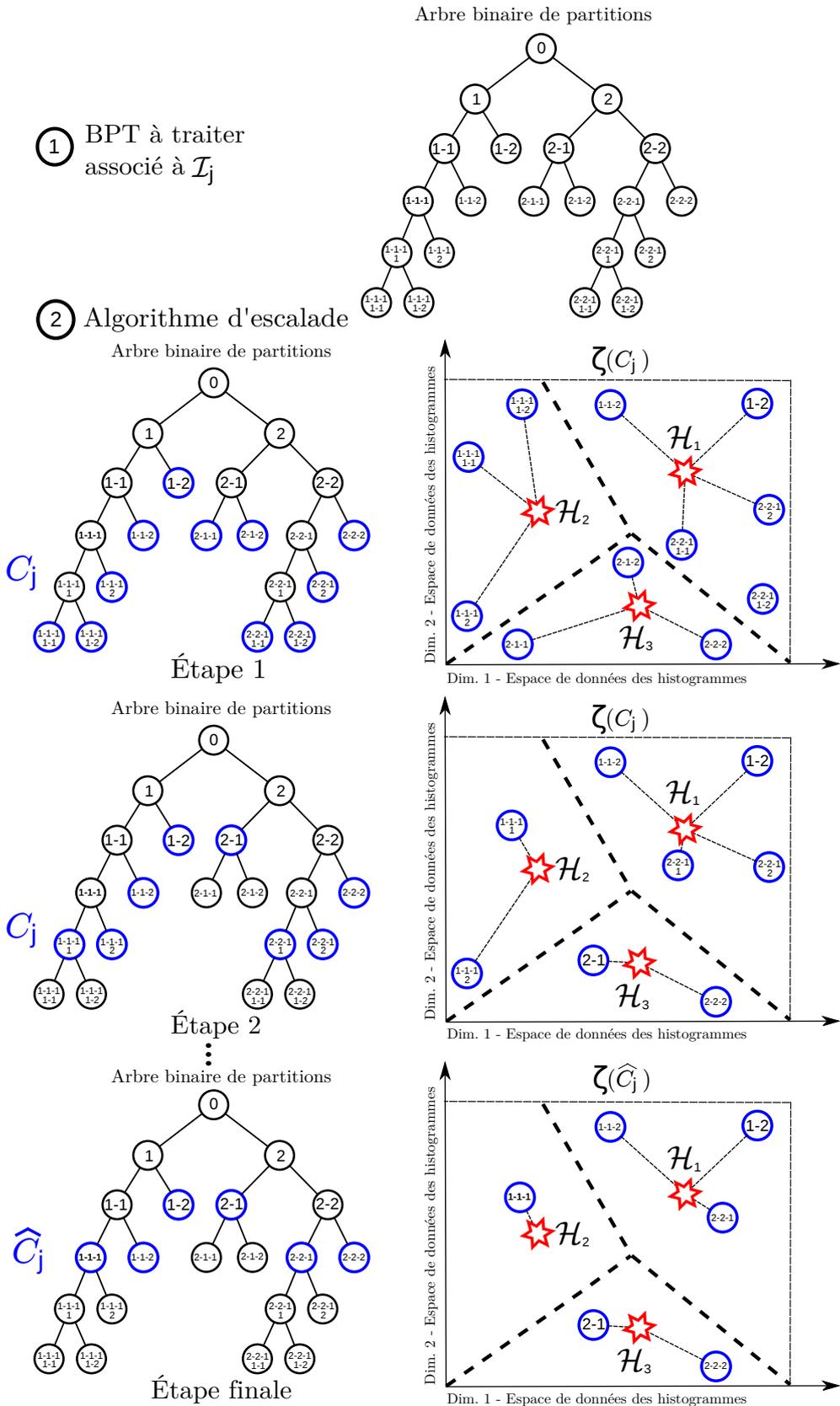


FIGURE 5.6 – Illustration de l’approche de segmentation basée sur des exemples : reproduction automatique de la coupe exemple.

où \mathcal{D} est la distance utilisée pour comparer deux histogrammes, et $C_j^i \subseteq C_j$ est l'ensemble des nœuds dont l'histogramme est le plus proche (relativement à \mathcal{D}) du centroïde modélisé par $H(\mathcal{I}, K_i)$ que des $u - 1$ autres centroïdes (il est à noter que $C_j = \bigsqcup_{i=1}^u C_j^i$), et $\overline{H(\mathcal{I}^j, \cup_{X \in C_j^i} X)}$ est l'histogramme moyen des nœuds $X \in C_j^i$, défini comme

$$\overline{H(\mathcal{I}^j, \cup_{X \in C_j^i} X)} = \sum_{X \in C_j^i} \frac{|X|}{|\cup_{Y \in C_j^i} Y|} \cdot H(\mathcal{I}^j, X) \quad (5.7)$$

Un algorithme d'escalade peut ensuite être utilisé pour trouver la meilleure coupe $\widehat{C}_j \subseteq \mathcal{N}_j$ parmi l'ensemble des nœuds \mathcal{N}_j du BPT associé à \mathcal{I}_j . Cet algorithme peut être formalisé de la manière suivante

$$\widehat{C}_j = \mathcal{F}(E) \quad (5.8)$$

où $\mathcal{F} : \mathcal{N}_j \rightarrow 2^{\mathcal{N}_j}$ est définie récursivement comme

$$\mathcal{F}(N) = \{N\} \quad (5.9)$$

si $N \notin \varphi(\mathcal{N}_j \setminus \{E\})$, *i.e.*, si N est une feuille du BPT, et comme

$$\mathcal{F}(N) = \begin{cases} \{N\} & \text{si } \zeta(\{N\}) \leq \sum_{N' \in \varphi^{-1}(\{N\})} \zeta(\mathcal{F}(N')) \\ \cup_{N' \in \varphi^{-1}(\{N\})} \mathcal{F}(N') & \text{sinon} \end{cases} \quad (5.10)$$

si $N \in \varphi(\mathcal{N}_j \setminus \{E\})$, *i.e.*, si N n'est pas une feuille du BPT.

En appliquant cet algorithme sur chacune des $p - 1$ imagettes restantes, on obtient automatiquement $p - 1$ segmentations avec un niveau de détail/d'hétérogénéité (*e.g.*, une échelle) similaire au niveau de détail de la segmentation exemple produite par l'utilisateur pour l'imagette \mathcal{I} .

La figure 5.6 illustre la reproduction automatique de la coupe exemple d'un BPT (associée à une image \mathcal{I}_j) sur les $p - 1$ BPT à couper restants. Le BPT associé à \mathcal{I}_j est illustré par la figure 5.6-①. L'application de cet algorithme d'escalade dans le but de couper ce BPT est illustrée par la figure 5.6-②.

Il est à noter que cet algorithme est adapté pour être appliqué sur une partie réduite du BPT de \mathcal{I}_j , correspondant à l'arbre induit par le sous-ensemble $\mathcal{N}_{j,\bar{e}} \subseteq \mathcal{N}_j$ des segments à faible élongation de \mathcal{I}_j . Ceci est justifié par le fait que les u centroïdes impliqués dans ce processus ont été obtenus par la classification non-supervisée des nœuds à faible élongation de $C_{\bar{e}}$. D'un point de vue pratique, la réduction du BPT n'influence pas le procédé algorithmique présenté ci-dessus. Les deux principales différences sont (i) le fait que l'arbre considéré n'est plus forcément un arbre binaire, étant donné qu'un nœud peut avoir 0, 1 ou 2 enfants, au lieu de 0 ou 2, et (ii) le fait que \widehat{C}_j ne forme plus forcément une partition de E . Cependant, les nœuds manquants nécessaires à la ré-obtention d'une partition peuvent être facilement retrouvés en réintroduisant \widehat{C}_j dans le BPT initial.

5.4 Bilan scientifique

5.4.1 Contributions

Pour permettre l'extraction et la segmentation d'objets complexes à partir d'images satellites (T)HRS, nous avons présenté une approche de segmentation basée sur les arbres binaires de partitions. Dans le cadre de l'analyse d'environnements urbains à partir d'images (T)HRS, différents objets d'intérêt peuvent être extraits à différentes échelles, rendant complexe la définition d'une coupe à partir d'un BPT. Nous nous sommes ainsi intéressés à définir un mécanisme permettant d'apporter une solution à ce problème. Pour ce faire, la méthodologie proposée implique une approche de segmentation basée sur les BPT, définie interactivement par l'utilisateur sur une partie de l'image (sous la forme d'exemples de segmentations), et automatiquement reproduite sur le reste des données. Il est alors possible d'utiliser les compétences de l'utilisateur pour choisir des niveaux d'échelle locaux (par le biais d'exemples de segmentations), automatiquement reproduits dans le reste de l'image.

Cette méthodologie de segmentation hiérarchique basée sur des exemples constitue la principale contribution de ce chapitre.

5.4.2 Valorisation scientifique

Ces travaux ont donné lieu à deux contributions dans des conférences internationales :

-  la première contribution [Kurtz et al., 2011c] est liée à un article applicatif dans la conférence internationale de télédétection *JURSE 2011, Joint Urban Remote Sensing Event*, publié par *IEEE Geoscience & Remote Sensing Society* ;
-  la seconde contribution [Kurtz et al., 2011a] est liée à un article méthodologique dans la conférence internationale de morphologie mathématique *ISMM 2011, International Symposium on Mathematical Morphology*, publié dans la série *Lecture Notes in Computer Sciences* de *Springer*.

Classification de segments multirésolutions

Sommaire

6.1	Caractérisation multirésolution des segments	71
6.1.1	Motivations	71
6.1.2	Obtenir une meilleure description des segments	72
6.1.3	Composition des segments à travers la résolution	72
6.2	Histogramme de composition	73
6.2.1	Définition	73
6.2.2	Construction d'un histogramme de composition	73
6.2.3	Préparation d'un histogramme de composition	74
6.3	Classification multirésolution	75
6.3.1	Classification de régions multirésolutions	75
6.3.2	Addendum : raffinement multirésolution	76
6.3.3	Addendum : extension à l'extraction de blocs urbains	77
6.4	Bilan scientifique	78
6.4.1	Contributions	78
6.4.2	Valorisation scientifique	78

Ce chapitre présente l'approche de classification multirésolution proposée au cours de cette thèse. Cette dernière permet de classifier des segments, obtenus par le biais d'un algorithme de segmentation, en considérant leurs compositions à travers des images à différentes résolutions spatiales. Pour ce faire, nous commençons par détailler comment caractériser des segments par l'intermédiaire d'attributs multirésolutions, modélisés sous la forme d'histogrammes de composition (Section 6.1). Nous montrons ensuite comment construire ces attributs (Section 6.2). Finalement, nous présentons une approche de classification non-supervisée pour classifier des segments caractérisés par le biais de ces attributs, ainsi qu'une extension particulière de cette approche pour le raffinement de clusters et pour l'extraction de blocs urbains (Section 6.3). Dans une dernière partie (Section 6.4), nous énonçons le bilan scientifique de ces travaux.

6.1 Caractérisation multirésolution des segments

6.1.1 Motivations

Comme nous l'avons présenté précédemment, les approches basées régions sont composées de deux étapes principales : la première est une étape de segmentation qui permet d'extraire de l'image des segments (transformés ensuite en régions par une étape de caractérisation) de l'image, pouvant potentiellement correspondre aux objets d'intérêt de la scène ; la seconde est une étape de classification qui a pour but d'étiqueter les régions extraites et présentant des caractéristiques similaires, avec des labels de classes (ou de clusters) communs. La notion de « caractéristique » prend ici toute son importance. En effet, plus la description utilisée pour caractériser les objets à classifier (ici les segments) sera pertinente (*i.e.*, reflétera au mieux les objets présents dans la scène à travers l'image), plus les régions associées à ces objets seront séparables dans l'espace des données, permettant ainsi d'obtenir de meilleurs résultats de classification.

Comme nous l’avons présenté dans le chapitre 2, l’étape de segmentation permet de caractériser les données à classifier par l’intermédiaire d’attributs basés régions. Ces attributs s’avèrent pertinents dans le cadre de la caractérisation de segments extraits d’images (T)HRS. Nous avons également montré, au cours du chapitre 3, que les données multirésolutions sont d’un intérêt considérable pour enrichir la description des éléments considérés. Nous nous sommes ainsi intéressés, dans ces travaux, à l’utilisation conjointe d’attributs régions et multirésolutions pour la caractérisation des segments à classifier. Nous présentons dans cette section comment caractériser des objets issus d’une segmentation par le biais de tels attributs.

6.1.2 Obtenir une meilleure description des segments

Dans les approches basées régions classiques, les segments à classifier (extraits d’une image à une résolution τ_t) sont généralement décrits par des attributs radiométriques (*e.g.*, moyenne, écart-type, indice de texture), par des attributs géométriques (*e.g.*, surface, périmètre) ou encore par des attributs contextuels (*e.g.*, type des segments voisins), calculés dans l’image à la résolution τ_t .

Des travaux menés au sein de notre équipe de recherche [Forestier et al., 2008] ont montré qu’il pouvait également être pertinent de caractériser un segment en fonction de son hétérogénéité radiométrique. Pour modéliser cette information, une solution consiste à stocker, pour chaque segment, la distribution des labels des clusters (issus d’une classification basée pixels) sous forme d’un histogramme dans lequel chaque barre représente la proportion de pixels ayant reçu un label spécifique par l’étape de classification. De cette manière, on résume les informations radiométriques tout en conservant celles relatives à la notion d’homogénéité ou d’hétérogénéité. Ces nouveaux attributs apparaissent ainsi beaucoup plus pertinents que de simples valeurs (ou moyennes) radiométriques, caractérisant les segments de l’image.

Parallèlement, les données multisources, et en particulier les données multirésolutions, permettent d’enrichir la description des segments considérés. Par exemple, caractériser des segments extraits à une résolution τ_t en prenant en compte des informations obtenues à une résolution τ_{t+1} peut permettre d’obtenir des informations ayant un niveau de détail plus élevé (*i.e.*, des caractéristiques plus spécialisées) pour décrire ces segments. Réciproquement, caractériser des segments extraits à une résolution τ_t en prenant en compte des informations obtenues à une résolution τ_{t-1} peut permettre d’obtenir des informations ayant un niveau de détail moins élevé (*i.e.*, des caractéristiques moins spécialisées) pour décrire ces segments. Ainsi, la caractérisation des segments à travers la résolution permet d’obtenir une description plus complète de ces derniers.

Dans le contexte de ces travaux, l’étape de segmentation utilisée exploite l’homogénéité radiométrique des images à faibles résolutions spatiales pour extraire des segments pouvant correspondre à des objets géographiques complexes, apparaissant comme hétérogènes dans les images à (très) hautes résolutions spatiales. Les expériences menées par notre équipe ont montré que les attributs basés régions classiques (*i.e.*, radiométriques et/ou géométriques) ne sont alors pas assez pertinents pour caractériser de manière efficace les segments considérés. Par exemple, les descriptions radiométriques de deux segments distincts, représentant deux blocs urbains, peuvent s’avérer similaires à une résolution τ_t alors que ces blocs urbains sont de deux types différents (*e.g.*, habitat et industrie). L’image utilisée pour extraire les segments (à la résolution τ_t) n’offre ainsi pas assez d’informations radiométriques et géométriques, permettant une description pertinente de ces objets, car ces derniers y apparaissent comme trop homogènes. De telles descriptions s’avèrent alors insuffisantes pour séparer ces objets dans l’espace des données et permettre ainsi une classification de ces éléments dans deux groupes disjoints. En conséquence, il semble plus pertinent d’utiliser une image offrant une description plus riche des segments considérés pour les caractériser.

Nous nous sommes ainsi intéressés à définir des moyens pour permettre une caractérisation de segments extraits à partir d’une image à une résolution τ_t en prenant en compte des informations obtenues à partir d’une image à une résolution τ_{t+1} .

6.1.3 Composition des segments à travers la résolution

Dans notre contexte applicatif, les objets d’intérêt complexes recherchés sont généralement composés, à un niveau d’analyse plus fin, d’ensembles d’objets simples. Si l’on considère l’exemple d’un bloc urbain, ce dernier peut généralement être décomposé en un ensemble d’objets urbains plus élémentaires comme des portions de routes, des toits de bâtiments ou encore des zones de végétation. Ainsi, de telles informations semblent

pertinentes pour caractériser ces objets complexes ; un bloc urbain d'habitations individuelles pourra par exemple être caractérisé par les informations suivantes :

- 15 toits en tuiles rouges ;
- 5 toits en tuiles grises ;
- 10 portions de routes ;
- 24 parcelles de végétation.

qui modélisent sa composition en termes d'objets simples. Cependant, de telles informations représentent des connaissances de haut niveau qui ne sont pas directement exploitables à partir des images considérées, car on ne dispose pas, à l'étape t , du contenu sémantique de l'image à la résolution \mathbf{r}_{t+1} .

Une solution plus réaliste pour caractériser un segment extrait d'une image à une résolution \mathbf{r}_t , en fonction de sa composition dans une image à une résolution \mathbf{r}_{t+1} , consiste à considérer la distribution de sa composition en termes de classes/clusters dans cette résolution plus fine. De telles informations peuvent naturellement être modélisées par des histogrammes similaires à ceux énoncés dans la section précédente, mais calculés à travers la résolution. Nous nommerons par la suite ces structures des histogrammes de composition. Nous proposons ainsi de caractériser les segments extraits par l'approche de segmentation hiérarchique définie précédemment *via* des attributs reposant sur de tels histogrammes. Dans le cadre de ces travaux, nous avons notamment étendu la notion d'histogramme de composition au paradigme basé régions et à la caractérisation de données à travers la résolution.

6.2 Histogramme de composition

Pour modéliser la composition des segments extraits d'une image à une résolution \mathbf{r}_t dans une image à une résolution \mathbf{r}_{t+1} , nous proposons d'utiliser des histogrammes de composition. Nous présentons dans cette section comment étendre cette notion au paradigme multirésolution. Nous commençons par décrire comment construire ces histogrammes ; nous détaillons ensuite le fonctionnement d'un processus permettant de préparer ces attributs pour les rendre exploitables par un processus de classification.

6.2.1 Définition

L'histogramme de composition $H(\mathcal{I}_C, E)$ représentant la distribution des valeurs d'une image classifiée $\mathcal{I}_C : E \rightarrow \llbracket 1, k \rrbracket$, peut être construit par le biais de la fonction

$$\mathcal{H}_{\mathcal{I}_C} : \llbracket 1, k \rrbracket \rightarrow \mathbb{N} \\ v \mapsto |\mathcal{I}_C^{-1}(\{v\})| \quad (6.1)$$

qui associe à chaque valeur $v \in \llbracket 1, k \rrbracket$, le nombre de pixels de \mathcal{I}_C possédant la valeur v . On considère ici v comme une mesure pouvant prendre k valeurs distinctes. L'histogramme $H(\mathcal{I}_C, E)$ peut alors être défini comme $H(\mathcal{I}_C, E) = (H_1(\mathcal{I}_C, E), H_2(\mathcal{I}_C, E), \dots, H_k(\mathcal{I}_C, E))$ où $H_i(\mathcal{I}_C, E) = \mathcal{H}_{\mathcal{I}_C}(i)$, $i \in \llbracket 1, k \rrbracket$.

L'histogramme de composition $H(\mathcal{I}_C, X)$ associé au sous-ensemble $X \subseteq E$ peut être construit d'une manière similaire par le biais de la fonction

$$\mathcal{H}_{\mathcal{I}_C, X} : \llbracket 1, k \rrbracket \rightarrow \mathbb{N} \\ v \mapsto |\mathcal{I}_C^{-1}(\{v\}) \cap X| \quad (6.2)$$

correspondant à l'histogramme de composition restreint de \mathcal{I}_C à X . Une telle structure pourra être employée par la suite pour caractériser un région X pendant une étape de classification.

6.2.2 Construction d'un histogramme de composition

Pour pouvoir construire des histogrammes de composition à travers la résolution, il est nécessaire d'étendre la définition de ces derniers au paradigme multirésolution.

Pour ce faire, on considère deux images $\mathcal{I}^t : E^t \rightarrow V^t$ et $\mathcal{I}^{t+1} : E^{t+1} \rightarrow V^{t+1}$ avec $\mathbf{r}_t \leq \mathbf{r}_{t+1}$ ainsi qu'une fonction de correspondance $\lambda_{t \rightarrow t+1} : E^t \rightarrow E^{t+1}$ permettant de mettre en correspondance les pixels de E^t avec

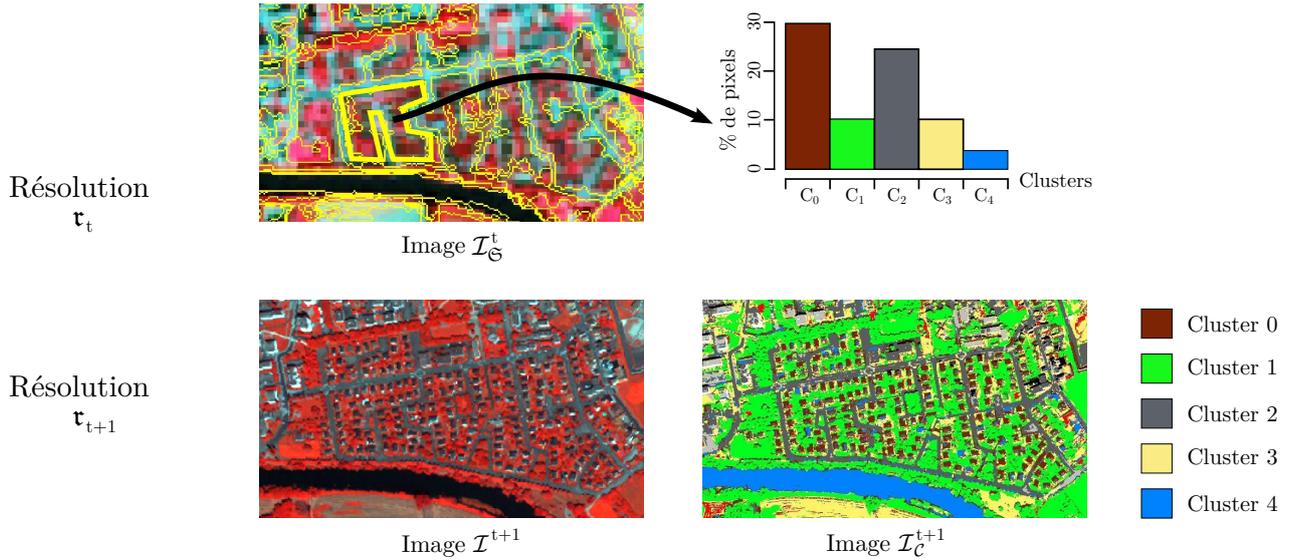


FIGURE 6.1 – Illustration d'un histogramme de composition : un des segments de la partition, délimités en jaune sur l'image \mathcal{I}^t , est caractérisé par un histogramme de composition modélisant la répartition de ses pixels en termes de labels de clusters dans l'image classifiée \mathcal{I}_C^{t+1} .

ceux de E^{t+1} (voir la section 3.1.4 B.). On considère également une image classifiée $\mathcal{I}_C^{t+1} : E^{t+1} \rightarrow \llbracket t, k \rrbracket$ à la résolution \mathbf{r}_{t+1} .

Il est alors possible de construire un histogramme de composition $H(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X))$ pour modéliser la distribution des pixels d'un ensemble $X \subseteq E^t$ en fonction des labels des classes composant l'image classifiée \mathcal{I}_C^{t+1} . Un tel histogramme peut être construit par le biais de la fonction suivante

$$\mathcal{H}_{\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)} : \llbracket 1, k \rrbracket \rightarrow \mathbb{N} \quad (6.3)$$

$$i \mapsto \left| \bigcup_{\mathbf{x} \in X} \lambda_{t \rightarrow t+1}(\mathbf{x}) \cap (\mathcal{I}_C^{t+1})^{-1}(\{i\}) \right|$$

Cet histogramme de composition permet d'associer à chaque label i de l'image classifiée \mathcal{I}_C^{t+1} , le nombre de pixels ayant le label i et correspondant à un pixel de l'ensemble X dans l'image \mathcal{I}^t . La figure 6.1 illustre la construction d'un histogramme de composition.

6.2.3 Préparation d'un histogramme de composition

Les algorithmes de classification sont souvent sensibles au nombre d'attributs caractérisant les données à classifier. En particulier, les algorithmes de classification basés sur des distances nécessitent de définir une mesure de similarité pour comparer les données à classifier. Généralement, plus le nombre d'attributs caractérisant ces données est important et moins la mesure utilisée est significative pour comparer ces données [Lee et Verleysen, 2007].

Dans notre contexte, le nombre d'attributs considérés correspond au nombre de barres dans les histogrammes de composition, c'est-à-dire au nombre de classes/clusters dans l'image de classification utilisée pour construire ces histogrammes. Notre approche étant dédiée à l'analyse d'image satellites à (très) hautes résolutions spatiales représentant des environnements urbains, le nombre de classes considérées peut être élevé (*e.g.*, 25 classes). Dans le but d'obtenir des résultats de classification plus pertinents, nous proposons de réduire la taille des histogrammes de composition considérés avant l'étape de classification.

Dans le but de simplifier un histogramme de composition, et en particulier dans le but d'éliminer un bruit potentiellement induit par les données, un seuil Λ (défini comme $\Lambda_q(v) = v$ si $v \geq q$ et 0 sinon) est appliqué sur ce dernier. L'idée générale est d'éliminer les valeurs correspondant à des classes de l'image classifiée \mathcal{I}_C^{t+1} n'apportant pas une contribution pertinente à la caractérisation des éléments à classifier.

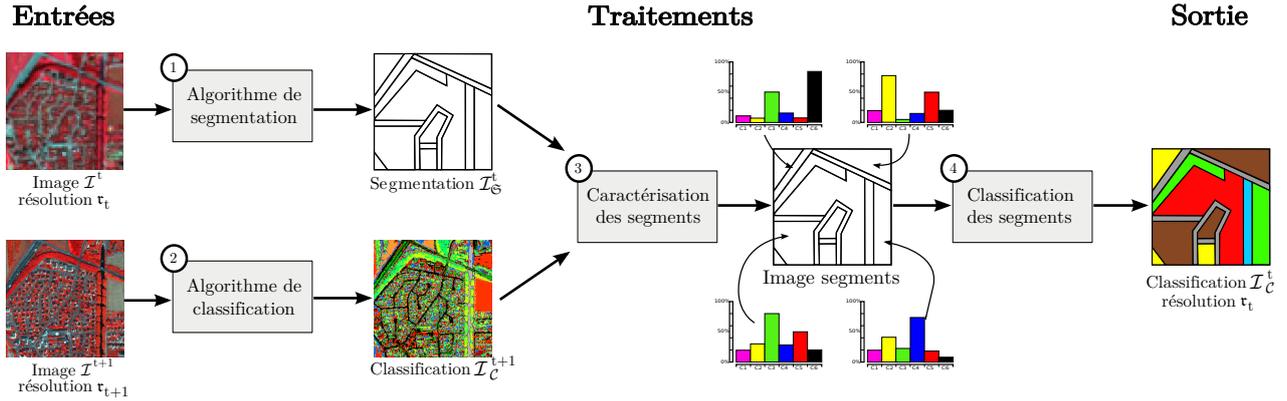


FIGURE 6.2 – Méthodologie de classification multirésolution.

Un histogramme de composition élagué, noté $H^*(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X))$, est alors défini comme

$$H^*(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)) = (H_1^*(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)), \dots, (H_k^*(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)))) \quad (6.4)$$

où $H_i^*(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)) = \Lambda_q(H_i(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)))$ avec $i \in \llbracket t, k \rrbracket$. Nous avons observé expérimentalement que choisir une valeur de q proche de la moyenne des valeurs de l'histogramme (*i.e.*, $|\cup_{\mathbf{x} \in X} \lambda_{t \rightarrow t+1}(\mathbf{x})|/k$) fournissait des résultats satisfaisants.

Les algorithmes de classification peuvent aussi être sensibles au fait que les attributs décrivant les données soient comparables entre eux. En conséquence, il est nécessaire de normaliser les histogrammes de composition utilisés pour caractériser les données. Si $P_i(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X))$ représente la fréquence d'occurrence des échantillons relativement à la i -ème barre, alors

$$P(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)) = \frac{H_i(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X))}{|\cup_{\mathbf{x} \in X} \lambda_{t \rightarrow t+1}(\mathbf{x})|} \quad (6.5)$$

Ainsi, $P(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X))$ peut être considéré comme un histogramme normalisé.

6.3 Classification multirésolution

6.3.1 Classification de régions multirésolutions

L'approche de classification multirésolution proposée permet de classifier des segments obtenus par le biais d'une segmentation d'une image à une résolution r_t , en considérant leur composition à travers une image classifiée à une résolution r_{t+1} *via* des histogrammes de composition.

A. Entrées et sorties

L'approche proposée requiert en entrée une image $\mathcal{I}^t : E^t \rightarrow V^t$ (*i.e.*, l'image à classifier), une segmentation \mathfrak{S} de cette image fournissant une partition de E^t (et une image segmentée \mathcal{I}_S^t), ainsi qu'une image $\mathcal{I}^{t+1} : E^{t+1} \rightarrow V^{t+1}$ avec $r_t \leq r_{t+1}$. On considère également qu'une fonction de correspondance $\lambda_{t \rightarrow t+1} : E^t \rightarrow E^{t+1}$, permettant de mettre en correspondance les pixels de E^t avec ceux de E^{t+1} , est disponible.

Le principe général consiste à combiner les informations extraites de l'image à la plus faible résolution (par le biais de la partition \mathfrak{S} , voir la figure 6.2-①) et les informations extraites de l'image à la plus haute résolution (par le biais d'une classification de l'image \mathcal{I}^{t+1} , voir la figure 6.2-②), pour obtenir une classification plus pertinente des segments de l'image \mathcal{I}_S^t .

B. Caractérisation des segments *via* des histogrammes de composition

Pour pouvoir caractériser les segments de l'image \mathcal{I}_S^t *via* leur composition en fonction de la distribution de clusters/classes formé(e)s dans l'image \mathcal{I}^{t+1} , il est nécessaire, au préalable, de classifier cette dernière.

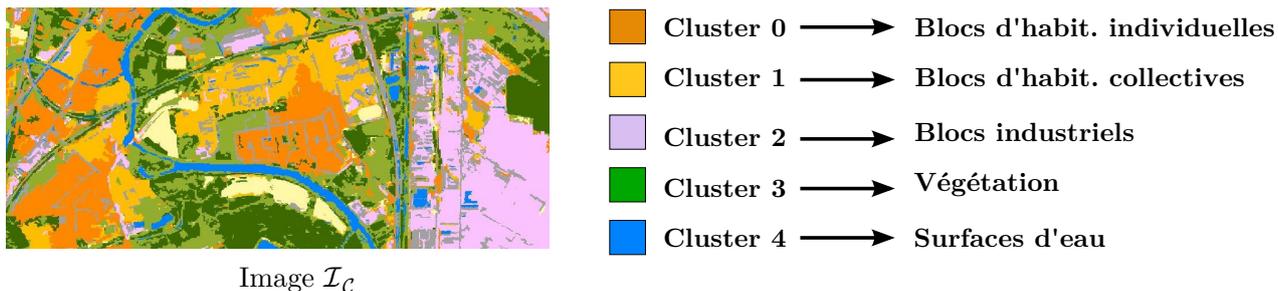


FIGURE 6.3 – Étiquetage des clusters obtenus.

Pour réaliser cette classification, il est possible d'employer une approche basée pixels ou encore une approche basée régions (supervisée ou non). Dans ces travaux nous avons choisi d'instancier cette approche en utilisant une classification non-supervisée basée pixels de l'image \mathcal{I}^{t+1} (via l'algorithme K -MEANS) fournissant l'image classifiée $\mathcal{I}_C^{t+1} : E^{t+1} \rightarrow \llbracket 1, k_{t+1} \rrbracket$.

Une fois cette image classifiée construite, il devient possible de déterminer la composition de chaque segment S de l'image \mathcal{I}_C^t (i.e., chaque $S \in \mathfrak{S}$) relativement à l'image classifiée \mathcal{I}_C^{t+1} . Cette composition peut être représentée via la notion d'histogramme de composition définie dans la section précédente.

Pour chaque segment $S \in \mathfrak{S}$, un histogramme de composition est ainsi calculé, prenant en compte la distribution des labels de clusters associés aux pixels composant S par l'image à haute résolution classifiée \mathcal{I}_C^{t+1} (voir la figure 6.2-③). Ces segments sont alors caractérisés par des attributs formés à partir de ces histogrammes, produisant un ensemble de régions qui vont pouvoir être traitées par un algorithme de classification.

C. Classification non-supervisée des régions

Les régions produites par l'étape précédente peuvent alors être classifiées (voir la figure 6.2-④). Cette classification non-supervisée est basée sur les histogrammes de composition normalisés (et élagués) des régions. Pour classifier ces régions, nous proposons dans ces travaux d'utiliser l'algorithme de classification non-supervisée K -MEANS. Cependant, l'approche autorise l'utilisation de n'importe quel algorithme de classification capable de traiter des données décrites par des valeurs numériques (structurées sous la forme d'histogrammes).

Les approches de partitionnement par classification non-supervisée requièrent une mesure de distance (et une moyenne) pour comparer chaque objet à classifier. Par la suite, on notera \mathcal{D} la distance utilisée pour comparer ces histogrammes. Le nombre de clusters k_t , qui caractérise le nombre de groupes à extraire par l'algorithme doit être fixé par l'utilisateur.

Ce processus conduit à la création d'un ensemble de k_t clusters $\{K_i\}_{i=1}^{k_t}$, associé à un ensemble de k_t centroïdes. Chaque centroïde u_i est représenté par l'histogramme de composition moyen (normalisé) $H(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(K_i))$ du cluster K_i .

D. Étiquetage des clusters obtenus

Une fois les différents clusters obtenus, il convient (ou non) à l'utilisateur d'étiqueter les clusters reconnus par des labels portant une sémantique (e.g., associer le cluster 2 à la classe thématique des blocs urbains industriels). Cette étape est illustrée par la figure 6.3.

Cependant, cette étape n'est pas nécessaire pour le bon déroulement de cette approche. Néanmoins, nous verrons dans la suite de ce manuscrit que les connaissances ajoutées dans ce processus par l'utilisateur peuvent être utilisées dans une étape de post-classification pour corriger et remettre en cause les résultats fournis par cette approche non-supervisée.

6.3.2 Addendum : raffinement multirésolution

Un autre avantage offert par la caractérisation multirésolution est la possibilité de pouvoir réaliser un raffinement des clusters formés à la plus haute résolution.

En effet, la composition des clusters formés dans l'image \mathcal{I}_C^t (à la résolution la plus basse) relativement à la distribution des clusters formés dans l'image \mathcal{I}_C^{t+1} (à la résolution la plus haute), permet de scinder des clusters de l'image haute résolution en sous-clusters plus spécifiques. Par exemple, un cluster de *Végétation* extrait de l'image \mathcal{I}_C^{t+1} et entrant dans la composition de plusieurs clusters de l'image classifiée \mathcal{I}_C^t (*Zones industrielles*, *Zones résidentielles à hautes densités*) peut être scindé en deux sous-clusters distincts à la résolution r_{t+1} (*Végétation industrielle* et *Végétation urbaine*). L'analyse de l'image \mathcal{I}^t fournit des informations contextuelles durant l'analyse de l'image \mathcal{I}^{t+1} .

Les quatre étapes nécessaires pour réaliser ce processus de raffinement multirésolution de clusters sont décrites ci-après. Ce processus est dédié à l'analyse simultanée de deux images (une image \mathcal{I}^1 à basse résolution spatiale et une image \mathcal{I}^2 à plus haute résolution spatiale) représentant la même scène urbaine.

- **Étape A – Segmentation** Les deux images sont dans un premier temps segmentées (indépendamment) par la méthode de segmentation proposée dans le chapitre précédent (ou une méthode de segmentation quelconque produisant des partitions complètes des images). Soit \mathcal{I}_G^1 et \mathcal{I}_G^2 les images régions obtenues.
- **Étape B – Classification non-supervisée initiale** L'image segmentée \mathcal{I}_G^1 (respectivement \mathcal{I}_G^2) est ensuite classifiée. Les régions étant caractérisées par les valeurs radiométriques des pixels de E^1 (respectivement E^2), cette classification permet d'obtenir un résultat lié à un niveau d'analyse spécifique (e.g., niveau des quartiers urbains *vs.* niveau des objets urbains simples). L'ensemble des k_1 (respectivement k_2) clusters résultants est noté \mathfrak{R}_1 (respectivement \mathfrak{R}_2). Les images classifiées résultantes sont notées \mathcal{I}_C^1 et \mathcal{I}_C^2 .
- **Étape C – Calcul de la composition des clusters de l'image basse résolution** Une fois que les images régions \mathcal{I}_G^1 et \mathcal{I}_G^2 ont été classifiées, il devient possible de déterminer la composition de chaque cluster K composant \mathcal{I}_C^1 relativement à l'image classifiée \mathcal{I}_C^2 par le biais d'un histogramme de composition.
- **Étape D – Découverte de nouveaux clusters à haute résolution** Finalement, en considérant la composition des clusters à basse résolution, des clusters à haute résolution peuvent être scindés en clusters moins spécialisés : si un cluster à haute résolution rentre dans la composition d'au moins un histogramme de composition d'un cluster à basse résolution, ce cluster à haute résolution est scindé en plusieurs sous-clusters à haute résolution. Plus formellement, on peut scinder $K_i \in \mathfrak{R}_2$ si $\exists j_1 \neq j_2$ avec $H(\mathcal{I}_C^2, \lambda_{1 \rightarrow 2}((\mathcal{I}_C^1)^{-1}(\{j_1\}))) (i) > s$ et $H(\mathcal{I}_C^2, \lambda_{1 \rightarrow 2}((\mathcal{I}_C^1)^{-1}(\{j_2\}))) (i) > s$ où s est un paramètre fixé par l'utilisateur. Ces nouveaux sous-clusters sont notés $K_{i_{j_1}}$ et $K_{i_{j_2}}$.

Cette méthode de raffinement de clusters à haute résolution fournit en sortie un nouveau résultat de classification de l'image segmentée à haute résolution \mathcal{I}_G^2 .

6.3.3 Addendum : extension à l'extraction de blocs urbains

Dans des travaux préliminaires à cette thèse (stage de Master 2 Recherche, janvier–juillet 2009), nous avons montré l'intérêt de l'approche de classification multirésolution proposée dans le contexte spécifique de l'extraction et de la classification de blocs urbains. Pour ce faire, l'approche de classification multirésolution a été modifiée pour la spécialiser dans l'analyse de ces structures d'intérêt. Ces travaux étendent ceux proposés dans [Wemmert et al., 2009], mais dans une optique totalement basée régions.

Ce processus est dédié à l'analyse simultanée de deux images (une image MRS \mathcal{I}^1 et une image HRS \mathcal{I}^2) représentant la même scène urbaine. L'idée principale, sous-jacente à cette approche, est de fusionner les informations fournies par l'analyse des régions de l'image HRS avec le résultat de classification non-supervisée de l'image MRS, afin d'obtenir un résultat final de classification permettant une analyse du territoire à un niveau intermédiaire (dans le cas présent, le niveau sémantique des blocs urbains).

La méthode proposée est divisée en quatre étapes principales décrites ci-dessous :

- **Étape A – Segmentation indépendante des images et classification des régions HRS** Dans un premier temps, les deux images sont segmentées indépendamment par un processus de fusion de zones plates. Soit \mathcal{I}_G^1 et \mathcal{I}_G^2 les images régions obtenues. Les régions de l'image HRS segmentée \mathcal{I}_G^2 sont alors classifiées d'une manière non-supervisée en fonction des moyennes des valeurs radiométriques des

pixels composant ces régions, fournissant ainsi un résultat de classification \mathcal{I}_C^2 pour l'image HRS. Pour ce faire, l'algorithme *K-MEANS* a été utilisé.

- **Étape B – Caractérisation des régions MRS** La composition de chacune des régions de l'image MRS segmentée \mathcal{I}_C^1 est alors calculée en fonction des clusters composant l'image classifiée HRS \mathcal{I}_C^2 par le biais d'un histogramme de composition.
- **Étape C – Classification des régions MRS** Basée sur ces compositions, une classification non-supervisée (*via* l'algorithme *K-MEANS*) des régions de l'image MRS segmentée \mathcal{I}_C^1 est réalisée : par opposition à une classification « classique », elle a pour but de former des clusters plus « sémantiques » que « radiométriques ». Ces clusters regroupent ainsi des objets d'intérêt relatifs à un niveau de sémantique intermédiaire, qui peut correspondre à celui des blocs urbains.
- **Étape D – Projection des régions HRS dans l'espace des clusters MRS** Pour chacun des clusters ainsi formés, sa composition moyenne en fonction des clusters formés dans le résultat de classification HRS, est ensuite calculée. Finalement, les régions de l'image HRS segmentée sont projetées dans l'espace des données de l'image MRS afin d'attribuer, à chacune, un label de cluster de niveau intermédiaire.

La méthode a été appliquée sur plusieurs jeux de données, chacun composé d'une image MRS et d'une image HRS, afin d'en extraire des régions pouvant correspondre à des objets d'intérêt complexes, permettant une analyse au niveau sémantique des blocs urbains. Les résultats ainsi obtenus se sont montrés encourageants [Kurtz et al., 2010].

6.4 Bilan scientifique

6.4.1 Contributions

Pour enrichir la caractérisation des segments extraits d'une image par le biais d'une étape de segmentation, nous avons proposé dans ces travaux de caractériser ces segments *via* des caractéristiques multirésolutions. De telles propriétés peuvent permettre de caractériser des segments extraits d'une image à une résolution τ_t (n'offrant pas d'informations radiométriques et géométriques suffisantes pour correctement décrire ces segments), en fonction de leur composition dans une image à une résolution τ_{t+1} , par la prise en compte de la distribution de leur composition en termes de classes/clusters dans cette résolution plus fine. Le principal avantage de cette approche est de pouvoir caractériser des segments homogènes (extraits d'une image à faible résolution) pouvant correspondre à des objets d'intérêt complexes, en considérant leurs descriptions dans une image à plus haute résolution, fournissant plus de détails sur ces segments. Pour modéliser une telle information, nous avons étendu la notion d'histogramme de composition au paradigme multirésolution. Nous avons ensuite proposé de caractériser les segments à classifier par des attributs reposant sur ces histogrammes particuliers. Finalement, nous avons présenté une approche de classification non-supervisée permettant de classifier des données caractérisées par le biais de ces attributs.

Cette méthodologie de classification multirésolution constitue la principale contribution de ce chapitre.

6.4.2 Valorisation scientifique

Ces travaux ont donné lieu à deux contributions scientifiques :

- 📌 la première [Kurtz et al., 2010] est liée à un numéro spécial (édité par Jean Serra et Daya Sagar) de la revue internationale *International Journal of Remote Sensing*, publiée par *Taylor & Francis* ;
- 📌 la seconde [Kurtz, 2011b] est liée à un article méthodologique dans les actes de l'atelier *Fouille de Données Complexes (FDC 2011)*, des journées francophones sur l'*Extraction et la Gestion des Connaissances (EGC 2011)*.

PCSC, un processus d'analyse collaboratif segmentation–classification

Sommaire

7.1 Schéma méthodologique de PCSC	79
7.1.1 Entrées et sorties	79
7.1.2 Chaîne de traitement	80
7.2 Processus collaboratif segmentation–classification à travers la résolution	80
7.2.1 Guider la segmentation par la classification	80
7.2.2 Paradigme multirésolution	81
7.3 Complexité algorithmique de PCSC	82
7.3.1 Complexité en temps	82
7.3.2 Complexité en espace	84
7.4 Bilan scientifique	84
7.4.1 Contributions	84
7.4.2 Valorisation scientifique	84

Ce chapitre présente les détails méthodologiques du processus collaboratif basé régions PCSC, proposé au cours de cette thèse. Le principe de PCSC repose sur une collaboration entre les approches de segmentation hiérarchique et de classification multirésolution, proposées dans les deux chapitres précédents. Nous commençons par présenter, de façon détaillée, le schéma méthodologique de ce processus (Section 7.1). Nous montrons ensuite comment l'approche de segmentation hiérarchique peut s'articuler avec l'approche de classification multirésolution pour traiter des ensembles d'images multirésolutions (Section 7.2). Nous étudions par la suite la complexité algorithmique liée à PCSC (Section 7.3). Finalement, nous énonçons le bilan scientifique de ces travaux (Section 7.4).

7.1 Schéma méthodologique de PCSC

Le processus PCSC permet d'obtenir une analyse basée régions (segmentation/classification) de n images d'une même scène à différentes résolutions spatiales, en partant de la plus faible pour aller jusqu'à la plus élevée, autorisant ainsi n différents niveaux d'interprétation. Le principe consiste à extraire des segments grossiers des images aux plus faibles résolutions (pour analyser les grandes zones de la scène), puis à les affiner successivement en utilisant les images aux résolutions les plus fines (pour analyser des structures plus élémentaires). Nous décrivons ci-après les entrées/sorties de ce processus ainsi que la chaîne de traitement sur laquelle ce dernier est basé.

7.1.1 Entrées et sorties

Le processus proposé requiert en entrée :

- un ensemble multirésolution $\{\mathcal{I}^t : E^t \rightarrow V^t\}_{t=1}^n$ de $n \geq 2$ images ($n = 3$ dans le cas courant) d'une même scène, ayant des résolutions spatiales croissantes $\mathbf{r}_1 \leq \dots \leq \mathbf{r}_n$, et pouvant avoir été produites par des capteurs différents (et posséder ainsi des bandes spectrales différentes);

et fournit en sortie :

- un ensemble $\{\mathcal{I}_{\mathcal{S}}^t\}_{t=1}^n$ de n images segmentées, associé à un ensemble $\{\mathcal{I}_{\mathcal{C}}^t\}_{t=1}^n$ de n images classifiées (une par résolution considérée), organisées hiérarchiquement, permettant différents niveaux d'interprétation.

7.1.2 Chaîne de traitement

Ce processus fonctionne en n étapes successives, chaque étape étant réalisée dans le but d'analyser une image \mathcal{I}^t parmi les n images (en partant de l'image \mathcal{I}^1 à la plus faible résolution, jusqu'à l'image \mathcal{I}^n à la résolution la plus élevée). À l'étape $t \in n$, l'image \mathcal{I}^t est traitée (on fait de plus l'hypothèse que toutes les images \mathcal{I}^i ($i \in \llbracket 1, t-1 \rrbracket$) ont déjà été traitées). Finalement, n niveaux de segmentation et de classification sont produits.

Ce processus étant basé sur le paradigme régions, chacune des n étapes est composée, itérativement, de :

1. une approche de segmentation hiérarchique (basée sur des exemples) ;
2. une approche de classification multirésolution ;
3. une projection des clusters obtenus dans la résolution suivante.

Pour traiter des ensembles d'images multirésolutions, ces trois sous-étapes sont appliquées successivement à travers la résolution. À chaque étape $t \in n$, le résultat de l'étape précédente $t-1$ (un ensemble de régions labellisées en différentes catégories) est projeté dans l'image à la résolution courante \mathcal{I}^t et traité comme entrée de la méthode. Le principe est alors de décomposer (c'est-à-dire de segmenter puis de classifier) ces différentes catégories de régions dans la résolution courante afin de les affiner. Pour ce faire, à chaque étape t , l'approche de segmentation (basée sur des exemples) est appliquée afin de partitionner l'ensemble des catégories de régions fournies par l'étape $t-1$. Une fois que toutes les régions, composant ces différentes catégories, ont été décomposées, il est possible de créer une partition globale de l'image \mathcal{I}^t en regroupant tous les segments résultant de cette étape. Ces segments sont ensuite caractérisés par des attributs multirésolutions, puis sont classifiés en différents ensembles homogènes par l'approche de classification multirésolution non-supervisée proposée, formant ainsi de nouvelles catégories de régions autorisant un niveau d'analyse différent. Ce résultat est alors projeté dans l'image \mathcal{I}^{t+1} , de résolution plus fine, pour que le processus y soit ré-appliqué. L'idée sous-jacente est d'affiner ces régions afin d'obtenir un niveau d'analyse encore plus fin.

La figure 7.1 illustre ce principe de collaboration descendante. Nous détaillons ci-après, l'ensemble des étapes composant ce processus.

7.2 Processus collaboratif segmentation–classification à travers la résolution

7.2.1 Guider la segmentation par la classification

Pour pouvoir segmenter une image $\mathcal{I}^t : E \rightarrow V$ via l'approche de segmentation hiérarchique définie précédemment (Chapitre 5), il est nécessaire de pré-diviser l'image \mathcal{I}^t en q imagerettes dans lesquelles des processus indépendants de segmentation pourront être appliqués.

Pour obtenir ces q imagerettes (ainsi que leur regroupement en k_{t-1} ensembles homogènes d'imagerettes) à partir d'une image \mathcal{I}^t , nous proposons d'utiliser le résultat d'analyse (segmentation/classification) obtenu via le traitement de l'image à la résolution précédente \mathcal{I}^{t-1} . Cette stratégie permet de guider l'approche de segmentation hiérarchique d'une image à une résolution \mathbf{r}_t par un résultat de classification obtenu à une résolution \mathbf{r}_{t-1} . De plus, cette stratégie permet de segmenter d'une façon similaire à un niveau t (*i.e.*, avec le même niveau d'échelle), des régions représentant le même type d'objets d'intérêt. L'approche de segmentation basée sur des exemples peut ainsi être appliquée k_{t-1} fois au niveau t pour segmenter le contenu des k_{t-1} catégories obtenues à l'étape précédente.

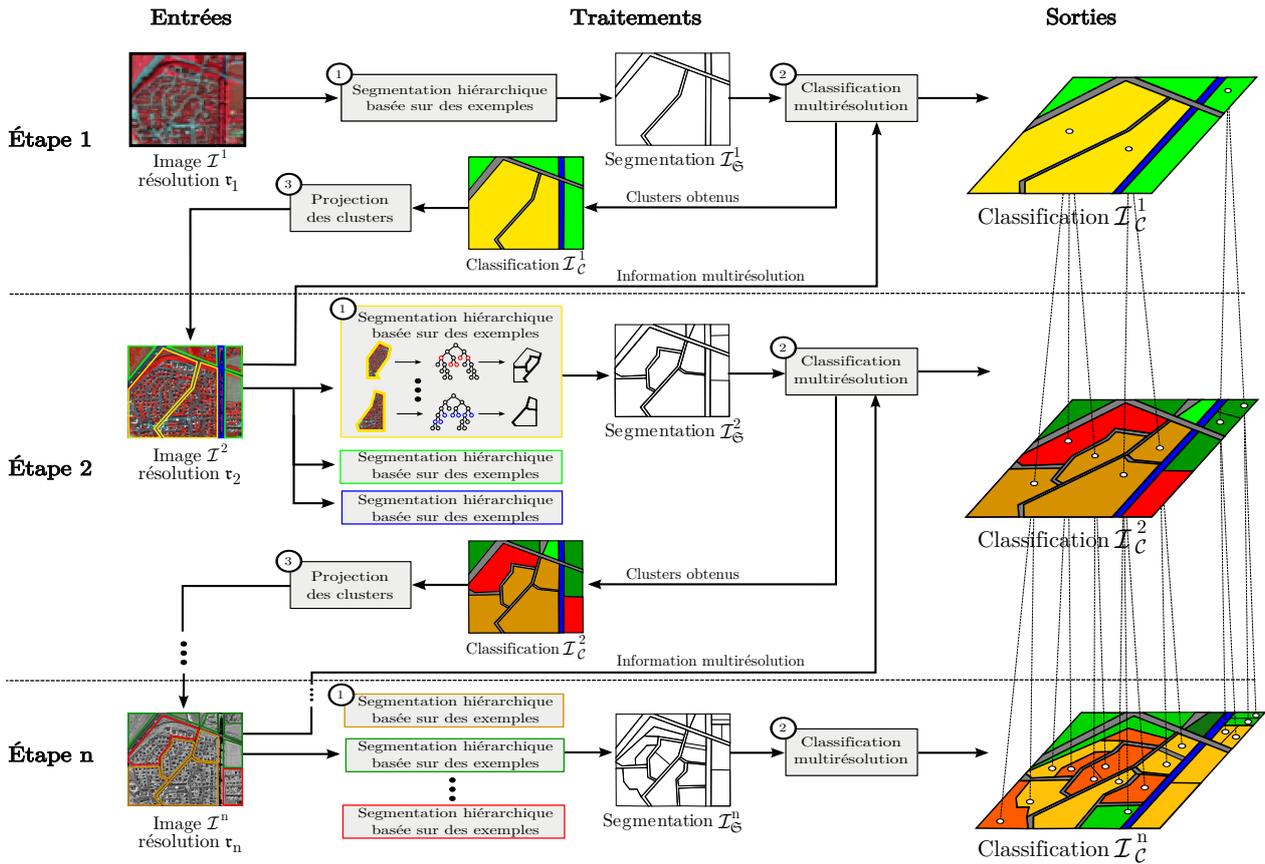


FIGURE 7.1 – Illustration du processus PCSC appliqué au traitement d’un ensemble d’images multirésolution.

7.2.2 Paradigme multirésolution

Pour traiter un ensemble multirésolution d’images $\{\mathcal{I}^t : E^t \rightarrow V^t\}_{t=1}^n$ par l’intermédiaire de cette approche, on applique alors la stratégie définie ci-après.

Grâce au traitement réalisé à l’étape précédente de l’image \mathcal{I}^{t-1} , une classification de l’image $\mathcal{I}^t : E^t \rightarrow V^t$ en k_{t-1} clusters est disponible, donnant lieu à une décomposition¹ de E^t . Par exemple, dans la figure 7.1, une classification de l’image \mathcal{I}^2 en trois clusters (*i.e.*, bleu, vert et jaune) est disponible. Ces clusters permettent de diviser l’image E^t en k_{t-1} catégories de segments au niveau de détail de l’image à plus faible résolution \mathcal{I}^{t-1} . Le nombre de clusters k_{t-1} doit être fixé par l’utilisateur. Il correspond au nombre de catégories d’objets d’intérêt que l’utilisateur veut extraire (et affiner) de l’image \mathcal{I}^{t-1} .

Chacune de ces catégories peut alors être décomposée en sous-catégories correspondant au niveau de détail offert par l’image \mathcal{I}^t (Figure 7.1-①). Pour ce faire, il est nécessaire de réaliser une segmentation de la partie de l’image correspondant à chacune des k_{t-1} catégories, *i.e.*, de segmenter l’imagerie $\mathcal{I}^{t,i} : K_i \rightarrow V^t$ de \mathcal{I}^t définie pour le cluster $K_i \subseteq E^t$ pour chaque $i \in \llbracket 1, k_{t-1} \rrbracket$. Il est à noter que l’utilisateur peut cependant avoir choisi de restreindre son étude à certaines de ces catégories, conduisant ainsi à une analyse partielle de l’ensemble d’images considéré. Pour chaque catégorie considérée $i \in \llbracket 1, k_{t-1} \rrbracket$, la segmentation de $\mathcal{I}^{t,i} : K_i \rightarrow V^t$ est réalisée grâce à l’approche de segmentation hiérarchique basée sur des exemples, présentée précédemment. En effet, un cluster K_i peut être décomposé en plusieurs segments (non connexes), induisant plusieurs imagerie de $\mathcal{I}^{t,i}$ de même résolution et représentant, au sein de l’image, les mêmes types d’objets d’intérêt. Ces imagerie peuvent ensuite être utilisées comme entrées de l’approche de segmentation hiérarchique basée sur des exemples. L’utilisateur peut alors réaliser la segmentation de l’une de ces imagerie, et cet exemple de segmentation est ensuite automatiquement reproduit dans toutes les autres imagerie composant le cluster K_i . L’image

1. Dans le cas de \mathcal{I}^1 , on considère, sans perte de généralité, qu’un seul cluster est disponible, dont la sémantique est liée au contenu de l’image (de la scène) entière.

segmentée $\mathcal{I}_{\mathcal{C}}^t$, obtenue en regroupant les k_{t-1} imageries segmentées (correspondant aux k_{t-1} catégories), constitue le résultat de cette sous-étape (et un résultat partiel de la méthodologie globale).

Comme indiqué ci dessus, à chaque étape t , la segmentation de l'image \mathcal{I}^t requiert une classification réalisée à l'étape $t-1$. Dans le but de pouvoir réaliser l'étape $t+1$, et d'obtenir un résultat permettant une analyse du niveau t , il est ainsi nécessaire de réaliser une classification de l'image segmentée $\mathcal{I}_{\mathcal{C}}^t$ à l'étape courante t . L'approche de classification employée ici est l'approche de classification multirésolution décrite précédemment. L'application de cette approche sur l'image segmentée $\mathcal{I}_{\mathcal{C}}^t$ produit un résultat de classification $\mathcal{I}_{\mathcal{C}}^t$ (Figure 7.1-②). Finalement, ces régions classifiées sont projetées dans la prochaine résolution, formant ainsi, pour chaque cluster résultant, une nouvelle catégorie d'imageries qui peut être décomposée dans l'image \mathcal{I}^{t+1} en appliquant la même stratégie (Figure 7.1-③).

7.3 Complexité algorithmique de PCSC

Nous détaillons ci-après la complexité algorithmique liée au processus collaboratif PCSC. Nous commençons par présenter la complexité en temps, permettant d'évaluer le nombre d'opérations nécessaires pour traiter un jeu de données multirésolution de taille ω . Nous présentons ensuite la complexité en espace, permettant d'évaluer les ressources mémoires requises au traitement.

7.3.1 Complexité en temps

Soit un jeu de données multirésolution composé de n images. Le nombre de pixels de chaque image est noté ω_i avec $i \in \llbracket 1, n \rrbracket$. Le nombre de pixels total est ainsi égal à $\omega = \sum_{i=1}^n \omega_i$ et correspond à la taille du jeu de données à traiter.

Le processus PCSC étant composé d'une étape de segmentation hiérarchique suivie d'une étape de classification multirésolution, appliquées successivement à travers la résolution (n fois), nous détaillons séparément la complexité en temps associée à chacune de ces deux sous-étapes.

A. Segmentation hiérarchique basée sur des exemples

L'approche de segmentation hiérarchique repose sur une stratégie impliquant la construction de BPT et l'apprentissage d'exemples de coupes. La complexité en temps, nécessaire à la construction d'un BPT dépend directement du contenu des données à segmenter et du comportement de l'algorithme de construction de l'arbre. Ainsi, le nombre d'opérations à effectuer pour construire un BPT est lié au nombre de nœuds composant cet arbre.

La taille d'un BPT peut être estimée généralement par : $\#noeuds = f + \frac{f}{2} + \frac{f}{4} + \frac{f}{8} + \dots + 2 + 1 = 2f - 1$ où f représente le nombre de feuilles d'un BPT. En considérant que chaque feuille de l'arbre représente un pixel d'une image ou une zone plate de cette dernière, le nombre de nœuds peut ainsi être estimé à $\#noeuds = (2 \times \omega_i) - 1$ pour une image \mathcal{I}^i . On considère ici que le nombre d'opérations requises pour la création d'un nœud de l'arbre est quasi-constant. Ainsi le nombre d'opérations à exécuter pour créer un BPT pour un ensemble de données de taille ω_i peut être donné par $\Theta((2 \times \omega_i) - 1)$. Cependant, le processus algorithmique de fusion de segments, sur lequel repose l'algorithme de construction de l'arbre, nécessite de maintenir un graphe modélisant l'adjacence des segments à fusionner ainsi qu'une liste triée permettant de récupérer la prochaine adjacence à considérer (*i.e.*, le prochain couple de segments à fusionner). Ceci nécessite $\log(\omega_i)$ opérations supplémentaires à effectuer après chaque fusion de segments, conduisant ainsi à un coût opérationnel total, pour la construction du BPT, égal à $\Theta(((2 \times \omega_i) - 1) \times \log(\omega_i)) = \Theta(\omega_i \times \log(\omega_i))$.

Pour pouvoir extraire des segments d'une image dans laquelle les objets d'intérêt peuvent émerger à différents niveaux d'échelle, cette approche de segmentation hiérarchique est basée sur une stratégie visant à découper l'image considérée en un ensemble d'imageries nécessitant chacune la construction d'un BPT. Soit q_i le nombre d'imageries considérées pour une image \mathcal{I}^i . Chacune de ces q_i imageries est composée de ω_i^j pixels avec $j \in \llbracket 1, q_i \rrbracket$. Il sera ainsi nécessaire de construire q_i BPT pour segmenter une image \mathcal{I}^i . Cependant, le nombre total de nœuds composant ces q_i BPT sera égal au nombre de nœuds résultant de la création

d'un seul BPT pour toute l'image. Le nombre d'opérations nécessaires à la création de ces q_i BPT sera donc exprimé par $\Theta(q_i \times (\omega_i^j \times \log(\omega_i^j))) = \Theta(\omega_i \times \log(\omega_i))$.

Le choix des coupes exemples est une étape interactive qui consiste à demander à l'utilisateur de parcourir k_i BPT afin d'y définir des coupes permettant de segmenter les q_i imagerie. Le parcours d'un arbre peut être effectué de manière linéaire avec le nombre de nœuds composant cet arbre. La définition des k_i exemples de coupes requiert ainsi un nombre d'opérations qu'on peut exprimer par $\Theta(k_i \times \omega_i^j) \leq \Theta(\omega_i)$.

L'apprentissage des coupes exemples est effectué par l'intermédiaire de l'algorithme K -MEANS, qui nécessite $c \times q \times d \times \Psi(\mathcal{D})$ opérations où c représente le nombre de clusters, q le nombre d'itérations, d le nombre de données à traiter et $\Psi(\mathcal{D})$ est le coût de la distance utilisée pour traiter ces données. Les expériences menées ont montré qu'en général, $c = 15$ clusters et $q = 15$ itérations étaient des valeurs suffisantes pour caractériser au mieux les coupes exemples interactivement définies par l'utilisateur. La complexité engendrée par cette étape d'apprentissage peut donc être exprimée par $\Theta(k_i \times 15 \times 15 \times \Psi(\mathcal{D}) \times \omega_i^j) = \Theta(k_i \times \omega_i^j) \leq \Theta(\omega_i)$, en considérant que le coût de calcul de \mathcal{D} est quasi-constant (*i.e.*, linéaire par rapport au nombre d'attributs utilisés pour caractériser les données).

Finalement, une fois ces k_i exemples de coupes définis, il est possible de les employer pour segmenter l'ensemble des q_i imagerie définies précédemment, par le biais d'un algorithme d'escalade appliqué à chaque arbre. Cet algorithme d'escalade est, de plus, linéaire avec le nombre de nœuds composant l'arbre. La complexité d'opérations nécessaires à la création de ces q_i coupes peut alors être exprimé par $\Theta(q_i \times \omega_i^j)$.

Le nombre d'opérations total résultant de l'application de cette approche de segmentation hiérarchique pour traiter une image \mathcal{I}^i peut alors être représenté par $\Theta((\omega_i \times \log(\omega_i)) + \omega_i + \omega_i + (q_i \times \omega_i^j)) = \Theta(\omega_i \times \log(\omega_i))$.

B. Classification multirésolution

L'approche de classification multirésolution permet de classifier les segments obtenus par le biais de l'étape de segmentation hiérarchique en différents groupes de segments homogènes. On notera r_i le nombre de segments résultant de la segmentation de l'image \mathcal{I}^i . Pour pouvoir classifier ces segments, cette étape nécessite une pré-étape de classification dans laquelle des attributs multirésolutions sont calculés, puis affectés à chacun des r_i segments considérés. Cette caractérisation peut être effectuée avec un nombre d'opérations que l'on peut exprimer par $\Theta(\omega_i)$, car il est nécessaire de parcourir l'ensemble des pixels composant les segments pour le calcul des histogrammes de composition.

Une fois les segments caractérisés, ces derniers sont classifiés en k_i clusters par le biais de l'algorithme K -MEANS qui nécessite $k_i \times 15 \times r_i \times \Psi(\mathcal{D})$ opérations. La complexité de cette opération peut alors être exprimée comme $\Theta(k_i \times 15 \times r_i \times \Psi(\mathcal{D})) = \Theta(r_i)$ en considérant que le nombre de clusters k_i choisi par l'utilisateur est négligeable par rapport au nombre r_i de régions à classifier.

La complexité d'opérations total résultant de l'application de cette approche de classification multirésolution pour traiter une image \mathcal{I}^i peut alors être représenté par $\Theta(\omega_i + r_i) = \Theta(\omega_i)$.

C. Complexité globale

Le nombre d'opérations pour traiter un ensemble multirésolution de n images \mathcal{I}^i avec $i \in \llbracket 1, n \rrbracket$ en utilisant successivement les approches de segmentation hiérarchique et de classification multirésolution à travers la résolution peut alors être exprimé par

$$\begin{aligned} \Theta_{temps}(\text{PCSC}) &= \Theta((\omega_1 \times \log(\omega_1) + \omega_1) + \dots + (\omega_n \times \log(\omega_n) + \omega_n)) \\ &= \Theta(\omega \times \log(\omega)) \end{aligned} \tag{7.1}$$

La complexité en temps (*i.e.*, en nombre d'opérations) requise par le processus PCSC peut ainsi être qualifiée de quasi-linéaire par rapport au nombre de pixels ω composant l'ensemble multirésolution d'images traitées. Par ailleurs, nous proposons dans le chapitre suivant un ensemble de validation permettant de confirmer expérimentalement ces résultats.

7.3.2 Complexité en espace

La complexité en espace (*i.e.*, en mémoire) requise par le processus PCSC est principalement liée à la construction des BPT lors de l'étape de segmentation hiérarchique.

La taille d'un BPT peut être estimée comme $\#noeuds = 2f - 1$ où f représente le nombre de feuilles du BPT considéré. La taille maximale d'un BPT est obtenue lorsque chaque pixel de l'image correspond à une feuille de l'arbre. En conséquence, le traitement d'images (T)HRS possédant un nombre très élevé de pixels (*e.g.*, $10\,000 \times 10\,000$ pixels) *via* l'utilisation d'une structure de BPT classique peut impliquer un nombre de nœuds conséquent (*e.g.*, jusqu'à 10^8 nœuds). Un tel ensemble de nœuds est difficile à stocker en mémoire vive par les configurations matérielles actuelles d'une station de travail. Cependant l'approche de segmentation hiérarchique proposée permet d'apporter une solution à ce problème de mémoire.

En effet, dans le cadre de cette approche les images sont traitées d'une façon descendante à travers la résolution et sont divisées en ensembles d'imagettes de tailles réduites, pour lesquelles des BPT sont construits indépendamment. En conséquence, il n'est jamais nécessaire de stocker en mémoire simultanément l'intégralité des nœuds résultants de la création des différents BPT. Au maximum, l'espace mémoire requis par le traitement d'un BPT pour une imagette (associée à l'image \mathcal{I}^i) composée de ω_i^j pixels sera égal à $2 \times \omega_i^j - 1$. La complexité en espace requise par le processus PCSC peut ainsi être représentée par

$$\Theta_{espace}(\text{PCSC}) = \Theta(2 \times \omega_i^j - 1) \ll \Theta(2 \times \omega - 1) \quad (7.2)$$

Ces résultats sont confirmés expérimentalement dans le chapitre suivant.

7.4 Bilan scientifique

7.4.1 Contributions

Ce chapitre a présenté un processus collaboratif basé régions, impliquant des mécanismes de segmentation et de classification. Il permet d'extraire des objets d'intérêt complexes à partir d'images satellites à (très) hautes résolutions spatiales. Pour s'abstraire de la complexité de ces images, la solution proposée consiste à appliquer ce processus à l'analyse d'images multirésolutions en adoptant une stratégie descendante à travers la résolution. Ce processus combine les avantages des stratégies multirésolutions et l'efficacité des approches basées régions, et permet d'extraire, à partir de ces images, des hiérarchies d'objets d'intérêt complexes autorisant une analyse de la scène observée à différents niveaux.

Ce processus constitue ainsi la contribution principale du premier axe de recherche de cette thèse.

7.4.2 Valorisation scientifique

Ces travaux ont donné lieu à deux contributions scientifiques :

-  la première [Kurtz et al., 2012] est liée à un article méthodologique paru dans la revue internationale *Pattern Recognition* publiée par *Elsevier* ;
-  la seconde [Kurtz, 2011a] est liée à un article dans les actes de la conférence nationale des rencontres de la *Société Francophone de Classification (SFC 2011)*.

Validations expérimentales

Sommaire

8.1	Protocole de validation	85
8.1.1	Contexte : cartographie des milieux urbains	85
8.1.2	Analyse de différents niveaux sémantiques	86
8.1.3	Description du protocole	86
8.2	Matériel	87
8.2.1	Jeux de données	87
8.2.2	Logiciel	87
8.3	Expérimentations, résultats et analyses	90
8.3.1	Efficacité	90
8.3.2	Précision	93
8.3.3	Robustesse	104
8.3.4	Complexité de calcul	107
8.4	Bilan scientifique	107
8.4.1	Contributions	107
8.4.2	Valorisation scientifique	108

Ce chapitre présente les expérimentations menées pour valider le processus PCSC, dans le contexte de l'analyse de scènes urbaines à partir d'images satellites multirésolutions. Nous commençons par décrire le protocole de validation proposé (Section 8.1). Nous détaillons ensuite les jeux de données utilisés, ainsi que le logiciel développé pour réaliser ces expérimentations (Section 8.2). Nous présentons par la suite les différentes expériences réalisées ainsi que les résultats obtenus, suivis d'une discussion sur ces derniers (Section 8.3). Dans une dernière section (Section 8.4), nous énonçons le bilan scientifique de ces travaux.

8.1 Protocole de validation

8.1.1 Contexte : cartographie des milieux urbains

Les collectivités chargées de la gestion et de la planification du développement urbain, ou encore les agences de gestion de crises environnementales et de catastrophes naturelles doivent pouvoir gérer et suivre l'évolution de l'urbanisation de façon précise, et en particulier sur les zones à hauts risques pour les populations humaines. Pour ce faire, il est nécessaire de cartographier les zones urbaines, pour obtenir des informations utiles, pouvant être directement exploitables par ces organisations (*e.g.*, quantifier le développement urbain, obtenir des cartes des dommages liés à une catastrophe naturelle). Nous avons présenté précédemment les problèmes et difficultés engendrés par le fait d'extraire manuellement (*i.e.*, par photo-interprétation) les objets d'intérêt relatifs aux différents niveaux d'analyse.

Parallèlement, les utilisateurs ont généralement besoin d'analyser les scènes urbaines à différents niveaux d'échelle. Les objets d'intérêt urbains, relatifs à ces niveaux d'analyse spécifiques, représentent des structures complexes et s'organisent de façon hiérarchique. Pour évaluer l'intérêt et l'efficacité de PCSC, nous proposons d'appliquer ce processus à une tâche d'analyse de scènes urbaines, à partir d'images satellites multirésolutions. L'idée sous-jacente à ces expérimentations est de montrer qu'il est possible d'utiliser un processus semi-automatique (1) pour aider les utilisateurs et faciliter leurs tâches d'intervention sur le territoire aux différents

Niveau	1. Quartiers urbains	2. Blocs urbains	3. Objets urbains simples
Échelle	1 : 100 000 ↔ 1 : 25 000	1 : 10 000	1 : 5 000
Objets d'intérêt			* Bâtiments/toits
			- Toits en tuiles rouges
		* Blocs urbains continus	- Toits en tuiles grises
		* Blocs urbains discontinus	- Toits réfléchissants
	* Quartiers à forte densité	- Blocs d'habitations individuelles	* Végétation
	* Quartiers à faible densité	- Blocs d'habitations collectives	- Végétation chloroph.
	* Zones industrielles	* Blocs urbains industriels	- Végétation non-chloroph.
	* Zones boisées	* Zones de végétation urbaine	* Voies de communication
* Zones agricoles	* Forêts	- Rues	
* Surfaces d'eau	* Zones agricoles	- Parkings	
* Sols nus	* Surfaces d'eau	* Surfaces d'eau	
	* Routes	- Rivières	
		- Étangs	
		* Sols nus	
		* Ombres portées	

TABLE 8.1 – Typologies et niveaux d'analyse utilisés pour analyser les territoires urbains à différentes échelles.

niveaux d'organisation, et (2) pour améliorer la précision des cartes produites, afin de permettre une meilleure compréhension de l'organisation spatiale de ces milieux.

8.1.2 Analyse de différents niveaux sémantiques

On trouve dans la littérature [Collet et al., 2001] de nombreuses typologies d'objets urbains, permettant une représentation cartographique à une échelle du 1 : 100 000 au 1 : 5 000. Nous présentons ci-après une nomenclature hiérarchique à trois niveaux de sémantique, permettant de cartographier des surfaces urbanisées :

1. le premier niveau de nomenclature correspond au niveau des quartiers urbains (Table 8.1, niveau 1). Il permet de cartographier le territoire d'une échelle au 1 : 100 000 à une échelle au 1 : 25 000, à partir d'images MRS (fournies, par exemple, par la constellation SPOT), avec un faible niveau de détail. Ce premier niveau de sémantique est employé pour étudier les grandes zones urbaines et leur densité ;
2. le deuxième niveau de nomenclature correspond au niveau des blocs urbains (Table 8.1, niveau 2) qui permet une analyse du territoire à l'échelle du 1 : 10 000. Ce niveau de sémantique a été proposé pour cartographier les blocs délimités par des axes de communication. Il est généralement utilisé pour cartographier le territoire à partir d'images HRS et/ou MRS ;
3. le troisième niveau de nomenclature correspond au niveau des objets urbains simples (Table 8.1, niveau 3). Il permet de cartographier le territoire à l'échelle du 1 : 5 000 afin d'analyser les objets urbains individuels (*e.g.*, bâtiments, routes) en fonction de leurs matériaux (*e.g.*, toits en tuiles de couleur rouge). Il peut être employé pour cartographier le territoire à partir d'images THRS. Dans les prochaines années, de telles images vont devenir de plus en plus accessibles, notamment grâce au programme spatial européen PLÉIADES [de Lussy et al., 2005].

Il est à noter que cette nomenclature est adaptée pour la cartographie de territoires urbains, situés principalement en Europe ou en Amérique du Nord.

Pour permettre une analyse du territoire à ces trois niveaux différents, nous proposons d'utiliser PCSC afin d'extraire des hiérarchies de régions pouvant correspondre à des hiérarchies d'objets urbains complexes (*e.g.*, quartiers urbains, blocs urbains, objets urbains simples), à partir d'ensembles d'images multirésolutions.

8.1.3 Description du protocole

Pour évaluer PCSC, nous avons développé un protocole expérimental reposant sur trois critères principaux :

1. **efficacité** : les étapes à suivre pour paramétrer le processus et obtenir des résultats de segmentation et de classification pertinents pour l'utilisateur. Les expériences réalisées pour évaluer ce facteur sont présentées dans la section 8.3.1 ;
2. **précision** : le degré de concordance entre les objets d'intérêt, extraits des images par PCSC, et des cartes de vérité terrain fournies par les experts en cartographie urbaine. Les expériences réalisées pour évaluer quantitativement les résultats offerts par PCSC sont présentées dans la section 8.3.2 ;
3. **robustesse** : le degré de déterminisme du processus. Pour ce faire, nous avons dans un premier temps évalué l'impact du choix des exemples de coupes fournis par l'utilisateur lors de l'étape de segmentation. Dans un second temps nous avons évalué la stabilité des modèles d'apprentissage issus de ces expériences, dans le cadre d'un tâche d'adaptation au domaine. Les résultats de ces expérimentations sont présentés dans la section 8.3.3.

Finalement, nous présentons dans la section 8.3.4 une étude expérimentale relative à la complexité algorithmique de PCSC, ainsi que les temps de calcul et les consommations mémoire liés à ces expériences.

8.2 Matériel

8.2.1 Jeux de données

Les expériences ont été réalisées sur la zone urbaine de Strasbourg (France), sur trois sites nommés **ELSAU**, **ILLKIRCH** et **HAUTEPIERRE** (Figure 8.1). Ces sites présentent des environnements urbains typiques, composés d'objets d'intérêt urbains classiques (*e.g.*, surfaces d'eau, zones de forêt, zones industrielles, blocs urbains d'habitations individuelles et collectives, zones agricoles). Le site **ELSAU** (6 576 m × 2 793 m) correspond à une zone péri-centrale tandis que le site **ILLKIRCH** (6 576 m × 2 803 m) est localisé dans la première couronne urbaine, au sud de la ville de Strasbourg. Le site **HAUTEPIERRE** (3 293 m × 2 246 m) est localisé dans la seconde couronne urbaine, au nord-ouest de l'agglomération de Strasbourg.

Chaque site est associé à un jeu de données multirésolution. Chaque jeu de données est composé de :

- une image SPOT-5 MRS multispectrale, composée de quatre bandes spectrales (© CNES). La résolution spatiale originale de l'image était de 10 m. Cette dernière a été ré-échantillonnée à 9,6 m dans le but d'obtenir un facteur entier entre les images ;
- un couple d'images QUICKBIRD (© DigitalGlobe Inc.), composé d'une image multispectrale HRS (2,4 m) à quatre bandes spectrales, et d'une image panchromatique THRS (60 cm).

Toutes les images ont été acquises au cours de l'année 2002 (à la même saison). Ces dernières sont illustrées par la figure 8.2, et les figures B.1 et B.2 (présentées en Annexe B). De plus, une synthèse de toutes les données utilisées dans ces expériences est présentée dans la table 8.2.

8.2.2 Logiciel

Pour permettre de tester expérimentalement le processus multirésolution proposé, un logiciel implantant le processus PCSC a été développé.

Dans le cadre de ces travaux, l'approche de segmentation hiérarchique basée sur des exemples a été intégrée dans la bibliothèque logicielle de segmentation JSL, développée en Java tandis que l'approche de classification multirésolution a été implantée dans la bibliothèque logicielle de clustering JCL, elle aussi développée en Java.

Finalement, une interface graphique permettant à l'utilisateur de paramétrer les différents algorithmes liés à ces deux approches (et d'explorer les résultats), a été développée et intégrée dans la plate-forme Java MUSTIC librement téléchargeable¹. Les bibliothèques JSL et JCL, ainsi que la plate-forme Java MUSTIC sont détaillées en annexe de ce manuscrit (Annexe D). Finalement, nous prévoyons d'intégrer totalement le processus PCSC au sein de ces bibliothèques et de distribuer ce logiciel sous une licence libre.

1. <http://lsiit-cnrs.unistra.fr/fdbt-fr/index.php/Logiciels>

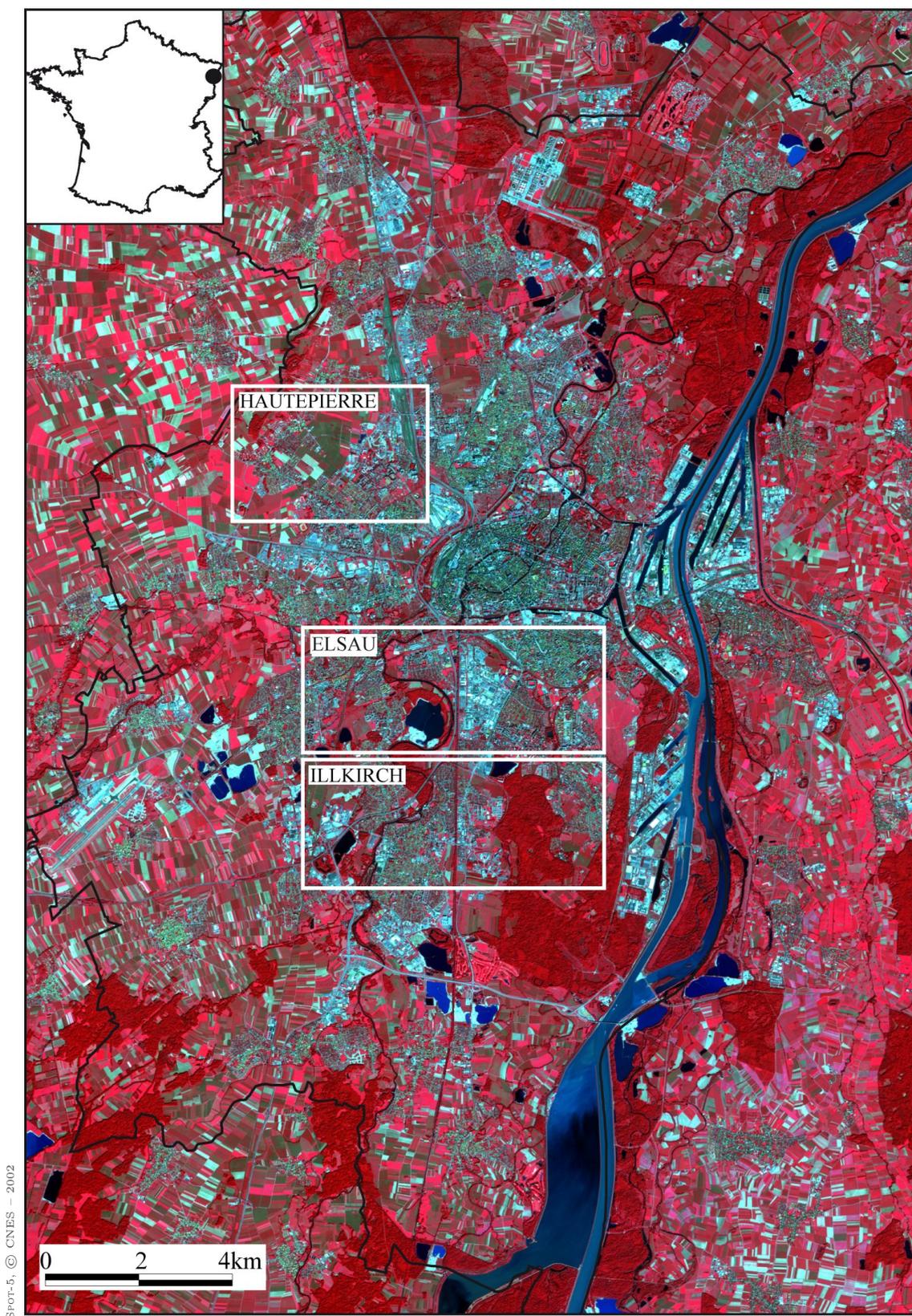


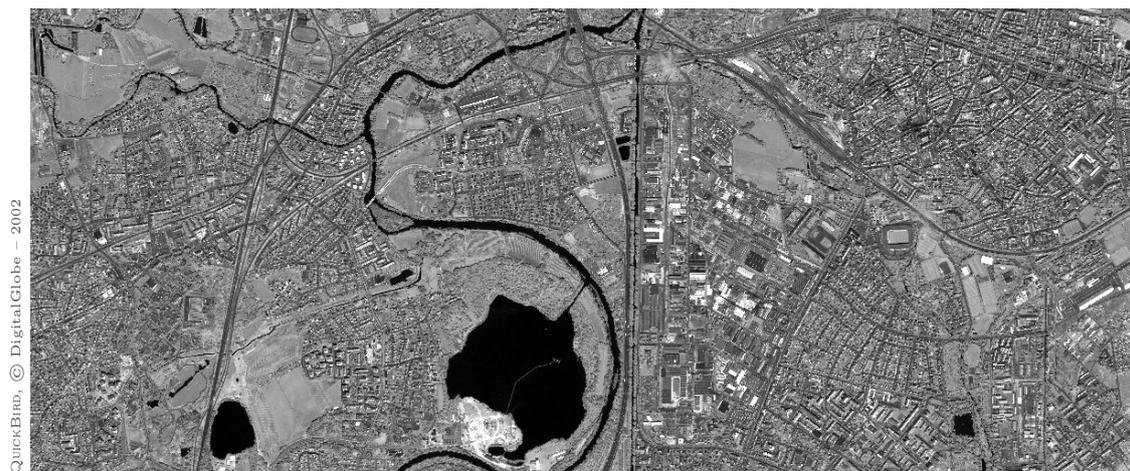
FIGURE 8.1 – Illustration de la zone urbaine de Strasbourg (France). Les sites d'étude (**ELSAU**, **ILLKIRCH**, **HAUTEPIERRE**) sont représentés sur une image MRS (10 m) et sont délimités par un rectangle blanc.



(a) MRS – 685 × 291 pixels – quatre bandes spectrales.



(b) HRS – 2 740 × 1 164 pixels – quatre bandes spectrales.



(c) THRS – 10 960 × 4 656 pixels – une bande spectrale.

FIGURE 8.2 – Présentation du jeu de données **ELSAU** (6 576 m × 2 793 m). (a) image MRS (9,6 m). (b) image HRS (2,4 m). (c) image THRS (60 cm).

Jeu de données	Image	Nombre de bandes	Résolution spatiale	Nombre de pixels	Surface	Mémoire
ELSAU	MRS	4 bandes	9,6 m	685 × 291	6 576 m × 2 793 m	1,5 Mo
	HRS	4 bandes	2,4 m	2 740 × 1 164		26 Mo
	THRS	1 bande	60 cm	10 960 × 4 656		98 Mo
ILLKIRCH	MRS	4 bandes	9,6 m	685 × 292	6 576 m × 2 803 m	1,5 Mo
	HRS	4 bandes	2,4 m	2 740 × 1 168		27 Mo
	THRS	1 bande	60 cm	10 960 × 4 672		101 Mo
HAUTEPIERRE	MRS	4 bandes	9,6 m	343 × 234	3 293 m × 2 246 m	600 Ko
	HRS	4 bandes	2,4 m	1 372 × 936		3 Mo
	THRS	1 bande	60 cm	5 488 × 3 744		24,5 Mo

TABLE 8.2 – Synthèse des jeux de données utilisés pour les expériences.

8.3 Expérimentations, résultats et analyses

8.3.1 Efficacité

Pour évaluer l'efficacité de PCSC, plusieurs expériences ont été réalisées afin d'extraire des hiérarchies d'objets urbains complexes à partir des trois jeux de données présentés. Le processus PCSC a été employé avec trois niveaux de résolution ($n = 3$), dans le but d'extraire trois niveaux de détail : les quartiers urbains à partir des images MRS \mathcal{I}_1 (Étape 1), les blocs urbains à partir des images HRS \mathcal{I}_2 (Étape 2) et les objets urbains simples à partir des images THRS \mathcal{I}_3 (Étape 3).

Pour appliquer PCSC, il est nécessaire d'ajuster deux groupes de paramètres : (1) ceux requis par l'approche de segmentation hiérarchique basée sur des exemples (*i.e.*, la paramétrisation du calcul des BPT, le nombre de centroïdes et d'exemples de coupes dans l'étape d'apprentissage) et (2) ceux requis par l'approche de classification multirésolution (*i.e.*, le nombre de clusters k_t pour chaque étape $t \in \llbracket 1, 3 \rrbracket$). Pour permettre une reproductibilité de ces expériences, nous fournissons ci-dessous les détails des paramètres expérimentaux.

A. Segmentation hiérarchique basée sur des exemples

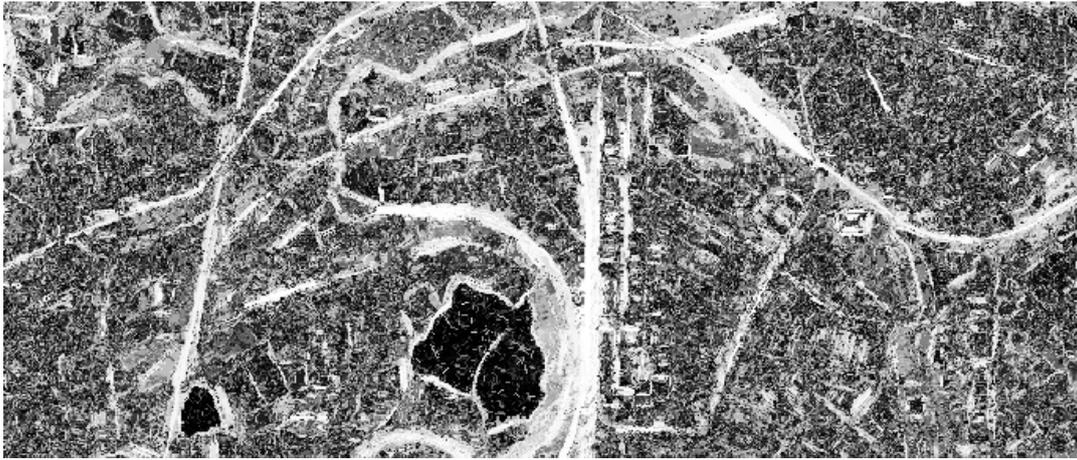
À chaque étape $t \in \llbracket 1, 3 \rrbracket$, l'approche de segmentation hiérarchique basée sur des exemples a été appliquée pour chaque catégorie K_i (composée de p_i régions) issue du traitement de la résolution précédente \mathfrak{r}_{t-1} . Nous décrivons ci-après la paramétrisation des trois étapes de cette approche.

Construction des BPT La construction des BPT nécessite la paramétrisation (1) des éléments relatifs au calcul de la carte d'élongation, et (2) des éléments relatifs à la fonction α .

Comme nous l'avons vu précédemment, le but de ces travaux n'est pas d'obtenir les meilleures cartes d'élongation possibles, mais d'être capable de calculer ces dernières avec un faible coût de calcul. Nous avons ainsi employé l'algorithme de segmentation hiérarchique en utilisant différentes cartes d'élongation. Ces dernières ont été obtenues en faisant varier le paramètre $\lambda \in \llbracket 0, v_b^+(E) - v_b^-(E) \rrbracket$ (*i.e.*, le seuil de tolérance). Pour chaque expérience, la boîte englobante a été calculée dans huit directions principales (*i.e.*, chaque $\pi/8$). Un compromis pertinent entre le temps de calcul requis pour cette étape et la qualité des résultats a été trouvé pour $\lambda = \frac{1}{10} \cdot (v_b^+(E) - v_b^-(E))$. La figure 8.3 illustre deux cartes d'élongation obtenues pour le jeu de données **ELSAU**.

La pondération $\alpha \in [0, 1]$ affecte la mesure de similarité calculée entre deux segments voisins S_i et S_j . Pour permettre à l'utilisateur d'ajuster cette pondération, nous proposons de modifier la fonction définie dans l'équation (5.5), en introduisant des paramètres dans $\alpha(O_r)$. Dans le cas présent, nous avons défini α comme

$$\alpha(O_r) = (1 - \varepsilon) \cdot \exp(-\gamma \cdot O_r^2) + \varepsilon \quad (8.1)$$

(a) Carte d'élongation de l'image MRS pour le jeu de données **ELSAU**.(b) Carte d'élongation de l'image HRS pour le jeu de données **ELSAU**.FIGURE 8.3 – Cartes d'élongation obtenues pour le jeu de données **ELSAU**.

où la fonction γ peut être définie comme

$$\gamma = \frac{1}{\delta^2} \cdot \ln \left(\frac{2 - 2\varepsilon}{1 - 2\varepsilon} \right) \quad (8.2)$$

avec $\delta \in [0, 1]$ et $\varepsilon \in [0, \frac{1}{2}[$. Le paramètre δ permet de contrôler la balance entre la fonction α (*i.e.*, la confiance accordée à l'homogénéité radiométrique O_r) et la fonction $1 - \alpha$ (*i.e.*, la confiance accordée à l'homogénéité géométrique O_g). Le paramètre ε permet quant à lui de contrôler le comportement asymptotique de la fonction α , tout en conservant un seuil minimal de confiance pour l'homogénéité radiométrique O_r . La figure 8.4 illustre l'utilisation des paramètres δ et ε . Ces paramètres peuvent être facilement déterminés : une première application de l'approche de segmentation hiérarchique avec $\varepsilon = 0$ et $\delta = 0$ peut fournir des indices relatifs au comportement global de α . Les valeurs de δ et de ε peuvent ensuite être adaptées en fonction des besoins de l'utilisateur. Par exemple, si l'image traitée n'est pas totalement structurée par des éléments linéaires, les valeurs des paramètres ε et δ peuvent être augmentées. Dans le cadre de ces expérimentations, ces valeurs ont été fixées respectivement à $\varepsilon = 0,2$ et $\delta = 0,3$ (Figure 8.4).

Apprentissage de la coupe exemple Suite à la construction des p_i BPT, l'utilisateur peut définir un (ou plusieurs) exemple(s) de coupes pertinent(s) à travers les p_i BPT disponibles. Durant ces expérimentations, nous avons étudié l'influence du nombre d'exemples de coupes fournis sur la qualité des résultats de segmentation/classification obtenus. Pour ce faire, nous avons observé l'impact du ratio $\eta_{ex} \in [0, 1]$ de surface couverte

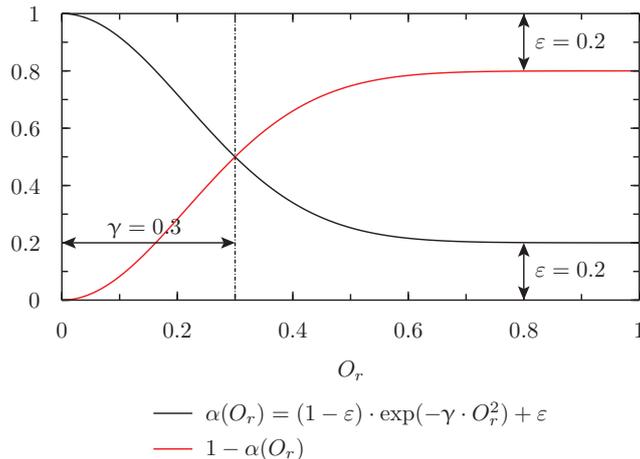


FIGURE 8.4 – Représentation des fonctions α et $(1 - \alpha)$ avec $\varepsilon = 0,2$ et $\delta = 0,3$ (Équations (8.1) et (8.2)).

par les exemples de coupes (en termes du nombre de pixels) sur la qualité des résultats de segmentation. Dans le cas où les p_i segments ont une taille similaire, ce ratio peut être défini comme $\eta_{ex} = \frac{\# \text{ exemples de coupes}}{p_i}$. Nous avons employé l’approche de segmentation hiérarchique pour traiter des images à différentes résolutions spatiales, en faisant varier ce ratio ($\eta_{ex} = 3\%, 5\%, 7\%, 10\%, 15\%, 30\%, 50\%$ et 70%). Le résultat de chacune de ces expériences a ensuite été évalué par le biais des mesures de qualité présentées ci-après. Le résultat de cette étude d’impact est détaillé dans la section 8.3.2.

Dans le but de reproduire au mieux les exemples de coupes fournis par les utilisateurs dans les BPT restants, il convient d’apprendre ces exemples. Le principe de la stratégie proposée est alors d’utiliser l’algorithme de classification K -MEANS, pour extraire u centroïdes qui caractérisent les exemples de coupes fournis par l’utilisateur. Nous avons étudié l’impact du nombre de centroïdes u sur la qualité des résultats finaux de segmentation/classification. Pour ce faire, nous avons appliqué l’approche de segmentation hiérarchique basée sur des exemples comme suit. Dans un premier temps, un BPT a été construit pour l’image \mathcal{I} . À partir de celui-ci, l’utilisateur a interactivement sélectionné une coupe pertinente C . Cette coupe C a ensuite été traitée par l’intermédiaire du processus d’apprentissage en utilisant différents nombres de centroïdes ($u \in [1, 20]$). Finalement, l’exemple appris a été utilisé pour re-couper le BPT précédent, associé à l’image \mathcal{I} . Chaque expérience a ainsi conduit à l’obtention d’une coupe particulière. Chaque coupe a alors été comparée qualitativement à la coupe C définie initialement par l’utilisateur.

Pour trouver les valeurs de u les plus adaptées (pour chaque résolution), ces expériences ont été réalisées en utilisant toutes les résolutions disponibles. Le résultat de cette étude d’impact est présenté dans la section 8.3.2.

Reproduction automatique de la coupe exemple Pour reproduire les coupes exemples traitées par l’étape d’apprentissage, nous avons défini précédemment un algorithme d’escalade. Ce dernier cherche à minimiser une mesure d’inertie calculée entre des paires d’histogrammes ordinaux. La distance standard, couramment utilisée pour la comparaison d’histogrammes ordinaux, est la distance euclidienne. Cependant, cette distance engendre des problèmes liés à la translation des valeurs des barres des histogrammes [Cha et Srihari, 2002], qui ne sont pas souhaitables lors de la comparaison d’histogrammes.

Pour pallier cet inconvénient, une solution consiste à utiliser la version contrainte (*i.e.*, temporellement) de la mesure de similarité temporelle Dynamic Time Warping (DTW) [Sakoe et Chiba, 1978]. Cette mesure autorise de faibles translations au sein des valeurs des barres des histogrammes (*i.e.*, axe radiométrique). Son calcul requiert cependant plus d’opérations que pour le calcul de la distance euclidienne. Nous avons étudié expérimentalement cette différence de complexité. Les résultats ont montré que le calcul de DTW nécessite environ dix fois plus d’opérations que celui de la distance euclidienne. La complexité requise pour comparer deux histogrammes $H(\mathcal{I}_1), H(\mathcal{I}_2) : V \rightarrow \mathbb{N}$ en utilisant la distance euclidienne est $\Theta(\text{card}(V))$ tandis que la complexité est $\Theta(\text{card}(V) \times c)$ (où c représente la tolérance aux distorsions des valeurs des barres) en utilisant la mesure de similarité DTW contrainte. Dans le pire cas, lors de l’utilisation de l’algorithme d’escalade, le

nombre d’histogrammes à comparer est de l’ordre de $\mathcal{N} \times u$, ce qui reste acceptable dans notre cas d’utilisation. Associée à cette mesure, une méthode de moyenne a été présentée dans [Petitjean et al., 2011]. Elle peut être utilisée pour définir les histogrammes moyens caractérisant les centroïdes.

B. Classification multirésolution

À chaque étape $t \in \llbracket 1, 3 \rrbracket$, les segments résultants (obtenus en regroupant les k segmentations produites) ont été catégorisés par l’approche de classification non-supervisée multirésolution (Chapitre 6), puis projetés dans la résolution suivante τ_{t+1} (excepté pour la dernière étape (Étape 3), pour laquelle une méthode de classification non-supervisée standard a été utilisée).

Les résultats de ces expérimentations ont montré que l’approche de classification ne permet pas de former directement tous les clusters appropriés, correspondant directement aux différentes classes thématiques recherchées. Pour pallier ce problème, une solution couramment utilisée consiste à extraire un plus grand nombre de clusters k_t que le nombre de classes thématiques extractibles des données considérées : l’expert peut ensuite choisir de fusionner certains de ces clusters. Par exemple, six classes thématiques (au niveau des quartiers urbains) peuvent être extraites à partir de l’image MRS (voir la table 8.3). En conséquence, il a été choisi en accord avec l’expert, d’extraire $k_1 = 10$ clusters à partir de l’image MRS. En suivant la même démarche, nous avons fixé $k_2 = 15$ pour traiter l’image HRS, et $k_3 = 13$ pour traiter l’image THRS.

8.3.2 Précision

Comme les techniques d’évaluation non-supervisée ne sont pas appropriées pour estimer la qualité des résultats fournis par les approches de segmentation interactives [McGuinness et O’Connor, 2010], nous nous sommes focalisés dans cette étude sur les approches d’évaluation supervisée. Ces dernières nécessitent la création d’ensembles de cartes de vérité terrain pour l’évaluation des résultats.

A. Cartes de vérité terrain

Les résultats de classification obtenus par le processus PCSC ont été évalués qualitativement par des comparaisons avec des cartes de vérité terrain certifiées, extraites de différentes bases de données relatives à la couverture des sols. Pour chaque jeu de données, un ensemble de trois cartes de vérité terrain est disponible (*i.e.*, une par niveau considéré). Pour mesurer la précision des résultats, ces cartes doivent être aussi précises que possible. En effet, d’éventuelles erreurs affecteraient directement la précision des tests de performance.

La figure 8.5 présente les cartes de vérité terrain associées au jeu de données **ELSAU**. La première carte contient six classes thématiques correspondant au niveau sémantique des quartiers (Figure 8.5(a)). Elle a été produite par un expert *via* une interprétation visuelle de l’image MRS. La seconde carte contient neuf classes thématiques correspondant au niveau sémantique des blocs urbains (Figure 8.5(b)). Cette dernière a été extraite d’une base de données régionale (© BDOCS – CIGAL2002²). La dernière carte est utilisée pour évaluer l’extraction des objets urbains simples. Cette carte contient une unique classe thématique correspondant à l’emprise au sol des bâtiments (Figure 8.5(c)). Elle a été extraite d’une base de données produite par l’institut national de l’information géographique et forestière (© IGN, 2002). Toutes les classes thématiques associées aux nomenclatures définies précédemment et aux différentes cartes de vérité terrain sont présentées dans la table 8.3. Les cartes de vérité terrain associées aux jeux de données **ILLKIRCH** et **HAUTEPIERRE** sont reportées en Annexe B (Figure B.3 et Figure B.4).

B. Évaluation des résultats

Pour comparer les résultats de segmentation/classification obtenus, aux cartes de vérité terrain, nous avons calculé des indices de qualité à un niveau local et à un niveau global (Table 8.4).

2. Ces données sont accessibles à l’adresse suivante : http://cigalsace.net/recette/produits_cigal.htm.



(a) Niveau des quartiers urbains (6 classes sémantiques).



(b) Niveau des blocs urbains (9 classes sémantiques).



(c) Niveau des objets urbains (1 classe sémantique).

FIGURE 8.5 – Cartes de vérité terrain associées aux images du jeu de données **ELSAU**. Voir la table 8.3 pour obtenir des informations relatives aux couleurs et aux concepts sémantiques associés aux classes thématiques représentées.

Niveau	Indice	Classe	Couleur
1. Quartiers	(1)	Quartiers résidentiels	
	(2)	Quartiers spécifiques	
	(3)	Surfaces d'eau	
	(4)	Routes	
	(5)	Zones agricoles	
	(6)	Zones de végétation	
2. Blocs	(1)	Blocs d'habitations collectives	
	(2)	Blocs d'habitations individuelles	
	(3)	Blocs urbains à fortes densités	
	(4)	Blocs urbains industriels	
	(5)	Surfaces d'eau	
	(6)	Routes	
	(7)	Zones agricoles	
	(8)	Zones de végétation urbaine	
	(9)	Forêts	
3. Objets	(1)	Ombres	
	(2)	Surfaces d'eau	
	(3)	Routes	
	(4)	Voies ferrées	
	(5)	Sols nus	
	(6)	Végétations chlorophylliennes	
	(7)	Forêts	
	(8)	Toits ardoises	
	(9)	Toits tuiles rouges	
	(10)	Toits tuiles grises	
	(11)	Toits métalliques	

} (12) Bâtiments

TABLE 8.3 – Classes thématiques (et leurs couleurs, voir la figure 8.5) associées aux cartes de vérité terrain.

Indices locaux Les indices de qualité locaux permettent d'évaluer indépendamment l'extraction de chaque classe thématique. Pour ce faire, pour chaque classe thématique à extraire, les clusters les plus représentatifs (en termes de partitions) ont été sélectionnés. Nous avons ensuite calculé :

- le pourcentage de faux positifs ($f^{(p)}$);
- le pourcentage de faux négatifs ($f^{(n)}$);
- le pourcentage de vrais positifs ($t^{(p)}$).

Ces indices sont utilisés pour estimer la précision \mathcal{P} et le rappel \mathcal{R} des résultats obtenus *via* l'approche proposée

$$\mathcal{P} = \frac{t^{(p)}}{t^{(p)} + f^{(p)}} \quad \text{et} \quad \mathcal{R} = \frac{t^{(p)}}{t^{(p)} + f^{(n)}} \quad (8.3)$$

Nous avons aussi calculé l'indice de qualité F-mesure \mathcal{F} , qui correspond à la moyenne harmonique de la précision et du rappel

$$\mathcal{F} = 2 \cdot \frac{\mathcal{P} \cdot \mathcal{R}}{\mathcal{P} + \mathcal{R}} \quad (8.4)$$

Symbole	Indice de qualité	Type
\mathcal{P}	Précision	} Précision locale
\mathcal{R}	Rappel	
\mathcal{F}	F-mesure	
\mathcal{K}	Kappa	} Précision globale
$\overline{\mathcal{F}}$	Moyenne harmonique des F-mesures	

TABLE 8.4 – Indices de qualité.

Indices globaux Pour évaluer la pertinence des résultats, nous avons par ailleurs calculé des indices de qualité globaux. Pour chaque expérience, nous avons évalué la moyenne harmonique pondérée $\overline{\mathcal{F}}$ des F-mesures (pondérée par les cardinaux des classes thématiques), et l'indice Kappa [Congalton, 1991] \mathcal{K} , qui est une mesure globale de la qualité d'une classification, définie comme

$$\mathcal{K} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (8.5)$$

avec $\Pr(a)$ et $\Pr(e)$ correspondant respectivement à l'accord entre les partitions et à la probabilité d'un accord aléatoire. L'indice Kappa prend ses valeurs dans l'intervalle $[0, 1]$. Il décroît proportionnellement au désaccord entre le résultat de classification et la carte de vérité terrain. Un moyen pour calculer cet indice consiste à considérer chaque couple de points $(\mathbf{x}_1, \mathbf{x}_2)$ et à analyser la configuration de ces points dans chacune des deux partitions (*i.e.*, le résultat de classification et la carte de vérité terrain). Il y a alors quatre configurations possibles ; pour chacune d'elles, un compteur est incrémenté chaque fois qu'une configuration apparaît :

1. les points \mathbf{x}_1 et \mathbf{x}_2 sont labellisés *via* la même étiquette dans le résultat de classification et dans la carte de vérité terrain (compteur s_s) ;
2. les points \mathbf{x}_1 et \mathbf{x}_2 sont labellisés *via* la même étiquette dans le résultat de classification mais pas dans la carte de vérité terrain (compteur s_d) ;
3. les points \mathbf{x}_1 et \mathbf{x}_2 sont labellisés *via* la même étiquette dans la carte de vérité terrain mais pas dans le résultat de classification (compteur d_s) ;
4. les points \mathbf{x}_1 et \mathbf{x}_2 ne sont labellisés *via* la même étiquette ni dans le résultat de classification, ni dans la carte de vérité terrain (compteur d_d).

Finalement, l'indice Kappa peut être calculé comme

$$\Pr(a) = \frac{s_s + d_d}{s_s + s_d + d_s + d_d} \quad \text{et} \quad \Pr(e) = \frac{(s_s + s_d) \cdot (s_s + d_s) + (s_d + d_d) \cdot (d_s + d_d)}{(s_s + s_d + d_s + d_d)^2} \quad (8.6)$$

C. Analyse des résultats

Ces différents indices ont ensuite été utilisés pour évaluer les résultats de segmentation/classification obtenus *via* PCSC. Dans un premier temps, nous détaillons une étude de l'impact de la proportion du nombre d'exemples de coupes η_{ex} fournis par l'utilisateur. Dans un second temps, nous détaillons l'impact du nombre de centroïdes u sur la qualité des résultats. Finalement, nous présentons une analyse globale de la pertinence des résultats et nous comparons ces derniers aux résultats obtenus par des méthodes issues de la littérature (sur d'autres jeux de données).

Impact de la proportion du nombre d'exemples de coupes Nous discutons ci-après de l'impact de la proportion du nombre d'exemples de coupes fournis par l'utilisateur sur la qualité des résultats produits par l'approche de segmentation hiérarchique basée sur des exemples. Les résultats de cette étude d'impact sont présentés dans les figures 8.6(a-c).

En analysant ces graphes, on constate qu'entre 0% et 15% d'exemples de coupes fournis par l'utilisateur, la qualité et la pertinence des u centroïdes (extraits durant l'étape d'apprentissage) augmente. L'étape de

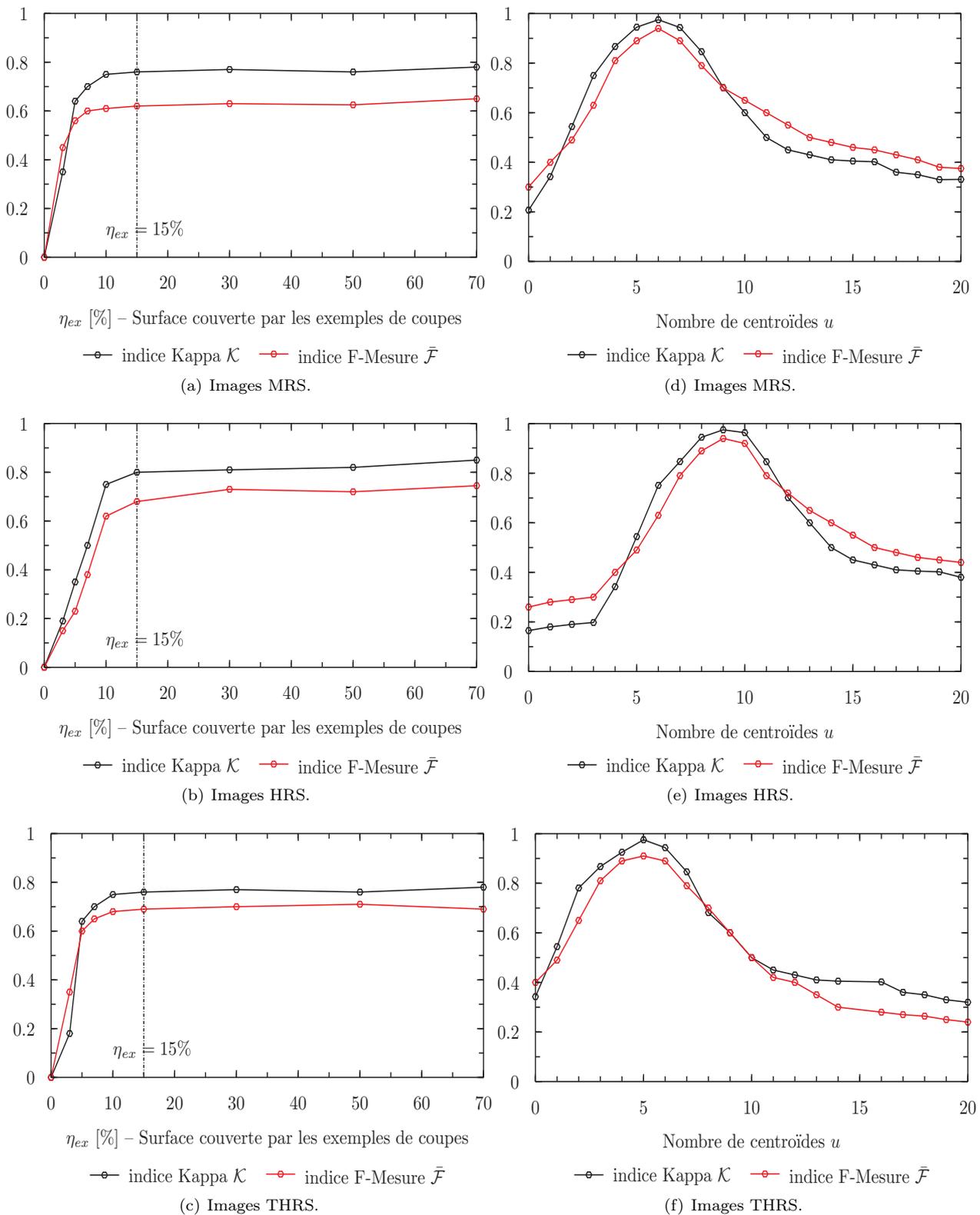


FIGURE 8.6 – Résultats des différentes études d'impact. (a-c) Impact de la proportion du nombre d'exemples de coupes η_{ex} fournis par l'utilisateur sur la qualité des résultats finaux de classification. (d-f) Impact du nombre de centroïdes u sur la qualité des résultats finaux de classification. L'évolution de l'indice de qualité \mathcal{K} (respectivement l'indice $\bar{\mathcal{F}}$) est représentée en noir (respectivement en rouge).

Niveau	Classe	ELSAU					ILLKIRCH					HAUTEPIERRE				
		\mathcal{P}	\mathcal{R}	\mathcal{F}	$\mathcal{K} \pm \sigma$	$\overline{\mathcal{F}} \pm \sigma$	\mathcal{P}	\mathcal{R}	\mathcal{F}	$\mathcal{K} \pm \sigma$	$\overline{\mathcal{F}} \pm \sigma$	\mathcal{P}	\mathcal{R}	\mathcal{F}	$\mathcal{K} \pm \sigma$	$\overline{\mathcal{F}} \pm \sigma$
1. Quartiers	(1)	0,72	0,77	0,74	$0,38 \pm 0,02$	$0,66 \pm 0,01$	0,59	0,80	0,68	$0,41 \pm 0,02$	$0,61 \pm 0,02$	0,61	0,86	0,70	$0,32 \pm 0,01$	$0,63 \pm 0,02$
	(2)	0,56	0,47	0,51			0,43	0,54	0,48			0,42	0,59	0,49		
	(3)	0,69	0,91	0,78			0,39	0,91	0,54			•	•	•		
	(4)	0,05	0,29	0,08			0,10	0,42	0,16			0,52	0,60	0,56		
	(5)	0,67	0,73	0,70			0,70	0,45	0,55			0,89	0,83	0,86		
	(6)	0,26	0,72	0,37			0,64	0,83	0,72			0,37	0,62	0,46		
2. Blocs	(1)	0,66	0,61	0,62	$0,36 \pm 0,02$	$0,69 \pm 0,03$	0,60	0,67	0,64	$0,41 \pm 0,02$	$0,62 \pm 0,03$	0,55	0,62	0,58	$0,34 \pm 0,02$	$0,64 \pm 0,02$
	(2)	0,72	0,75	0,74			0,68	0,42	0,52			0,57	0,64	0,60		
	(3)	0,64	0,63	0,64			•	•	•			•	•	•		
	(4)	0,55	0,52	0,53			0,65	0,72	0,68			0,59	0,42	0,49		
	(5)	0,75	0,84	0,79			0,78	0,87	0,82			•	•	•		
	(6)	0,31	0,28	0,29			0,33	0,31	0,32			0,28	0,54	0,37		
	(7)	0,71	0,76	0,72			0,68	0,75	0,71			0,55	0,49	0,52		
	(8)	0,21	0,63	0,31			0,35	0,77	0,48			0,12	0,27	0,17		
	(9)	0,25	0,49	0,33			0,46	0,53	0,48			0,35	0,62	0,45		
3. Objets	(12)	0,83	0,64	0,72	$0,38 \pm 0,01$	$0,71 \pm 0,02$	0,82	0,72	0,77	$0,39 \pm 0,02$	$0,77 \pm 0,02$	0,74	0,63	0,68	$0,39 \pm 0,02$	$0,68 \pm 0,02$

TABLE 8.5 – Résultats d'évaluation obtenus pour les jeux de données **ELSAU**, **ILLKIRCH**, **HAUTEPIERRE**. Le symbole (•) signifie que la classe thématique considérée n'est pas disponible dans le jeu de données courant.

reproduction automatique (et en particulier, l'algorithme d'escalade) est directement affectée par la qualité de ces u centroïdes qui modélisent les exemples de segmentations. À partir de 15% d'exemples fournis, la qualité des centroïdes n'augmente plus d'une façon notable et tend à devenir constante (relativement à \mathcal{K} et $\overline{\mathcal{F}}$).

Cependant, la définition de ces exemples de coupes par l'utilisateur est une étape consommatrice de temps. Les expériences ont montré que $\eta_{ex} = 5\%$ est un compromis satisfaisant entre (1) un temps raisonnable passé à définir les exemples et (2) une qualité de segmentation pertinente pour les images MRS (Figure 8.6(a)). En conséquence, fournir 5% d'exemples de coupes parmi les k régions à segmenter (soit deux exemples en moyenne par cluster) permet d'obtenir une étape de reproduction automatique des coupes exemples qui n'est pas excessivement coûteuse en temps pour l'utilisateur (*i.e.*, 30 secondes par exemple). Lors du traitement d'images HRS, $\eta_{ex} = 10\%$ d'exemples semble une valeur suffisante pour obtenir des résultats pertinents (Figure 8.6(b)) tandis que lors du traitement d'images THRS, la valeur de η_{ex} peut être fixée à $\eta_{ex} = 7\%$ (Figure 8.6(c)). Finalement, l'utilisateur est libre de modifier η_{ex} si la qualité des résultats obtenus n'est pas suffisante pour l'application courante.

Impact de la proportion du nombre de centroïdes Les expérimentations ont montré que le choix du nombre de centroïdes u a aussi une influence considérable sur la qualité des résultats obtenus.

Les figures 8.6(d-f) présentent l'évolution des indices \mathcal{K} et $\overline{\mathcal{F}}$ relativement à la variation des valeurs de u pendant le traitement des images aux différentes résolutions spatiales. Les meilleurs résultats ont été obtenus quand le nombre de centroïdes est fixé à $u = 7$ pour les images MRS (Figure 8.6(d)), $u = 10$ pour les images HRS (Figure 8.6(e)) et $u = 6$ pour les images THRS (Figure 8.6(f)). Les expérimentations ont montré que lors du traitement d'images MRS et THRS, il est nécessaire d'utiliser un nombre plus faible de centroïdes u que lors du traitement d'images HRS. Ce comportement peut être justifié par le fait que le nombre de matériaux analysables à partir d'images HRS est plus important qu'à partir d'images MRS (du fait de leurs faibles résolutions spatiales) et THRS (du fait de leurs faibles résolutions spectrales). Au demeurant, le nombre de centroïdes peut être adapté par l'utilisateur qui peut estimer, en visualisant les images, combien de matériaux différents peuvent composer les segments à raffiner.

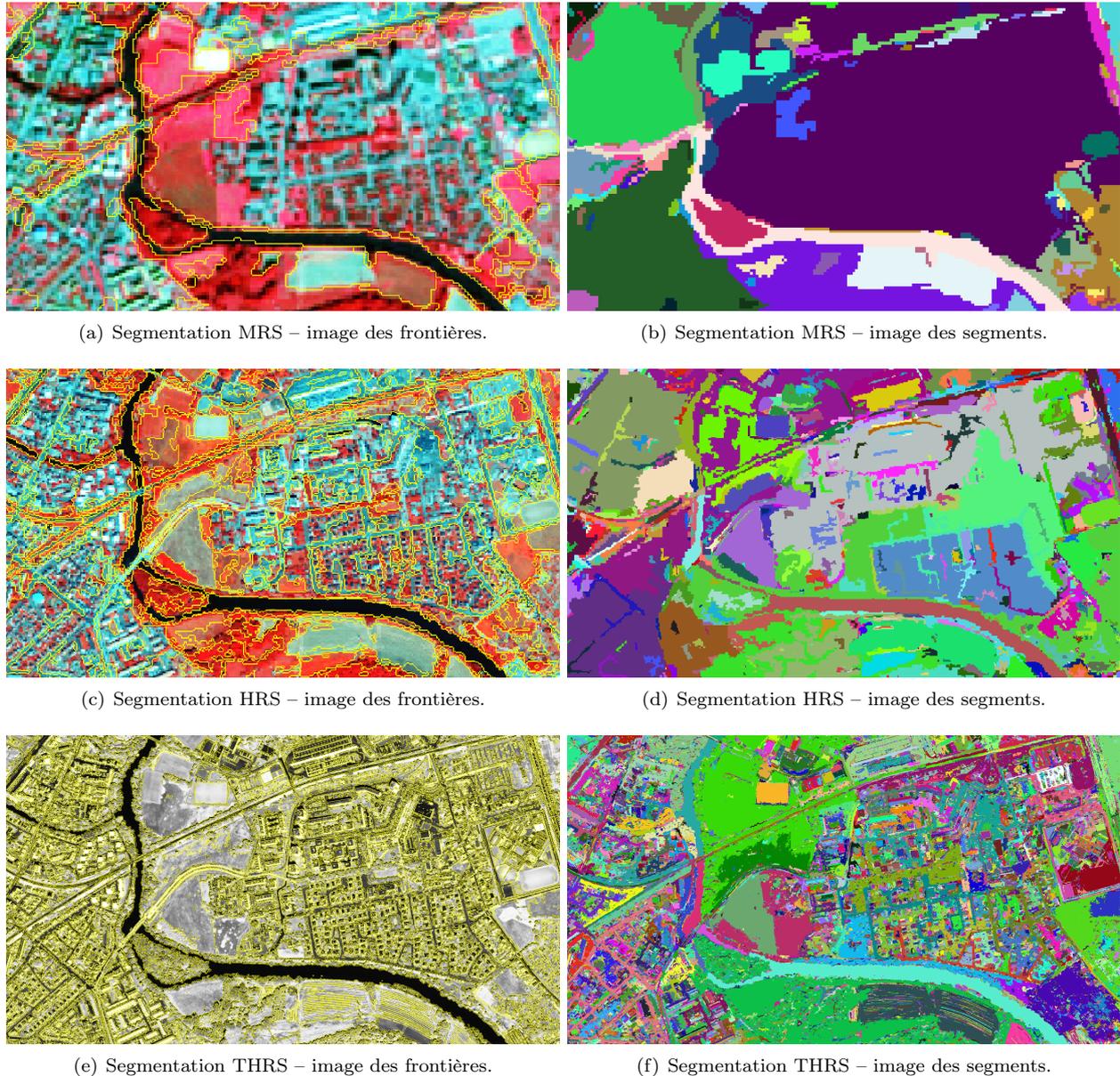


FIGURE 8.7 – Résultats de segmentation pour le jeu de données **ELSAU** ($6\,576\text{ m} \times 2\,793\text{ m}$). (a) Niveau des quartiers. (b) Niveau des blocs. (c) Niveau des objets urbains.

Évaluation globale des résultats Nous avons par la suite fixé les différents paramètres de PCSC aux valeurs discutées précédemment et appliqué ce processus sur les trois jeux de données considérés. Nous présentons ci-après les meilleurs résultats de segmentation/classification permettant d'évaluer globalement la pertinence de la méthode. La table 8.5 détaille les différents scores d'évaluation obtenus. La figure 8.7 illustre une partie des résultats de segmentation obtenus sur le jeu de données **ELSAU**. Les résultats fournis par l'algorithme *K-MEANS* étant sensibles à l'initialisation des centroïdes, chaque expérience a été répétée dix fois en faisant varier les graines de l'algorithme. Nous avons ensuite calculé la valeur de variance σ obtenue pour les différents indices considérés.

L'étape 1 a été appliquée sur les images MRS afin de séparer les grandes structures de la scène (*e.g.*, quartiers urbains, zones de forêt, surfaces d'eau). Les résultats de segmentation/classification obtenus sont présentés par les figures 8.8(a), 8.9(a), 8.10(a). Après l'étape de classification, les comparaisons entre les régions classifiées résultantes et les cartes de vérité terrain permettent d'obtenir des valeurs de Kappa et de



(a) Classification MRS – niveau des quartiers.



(b) Classification HRS – niveau des blocs.

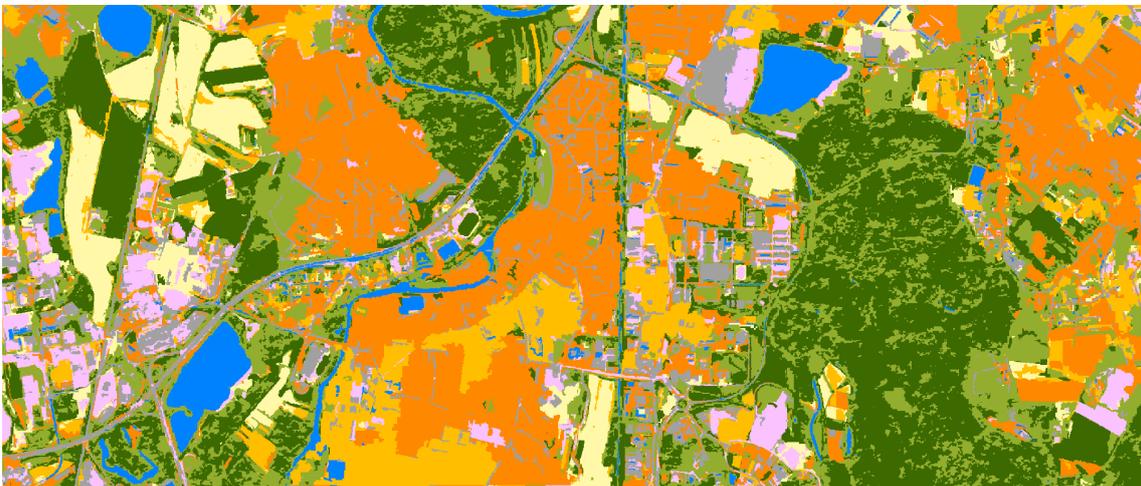


(c) Classification THRS – niveau des objets urbains.

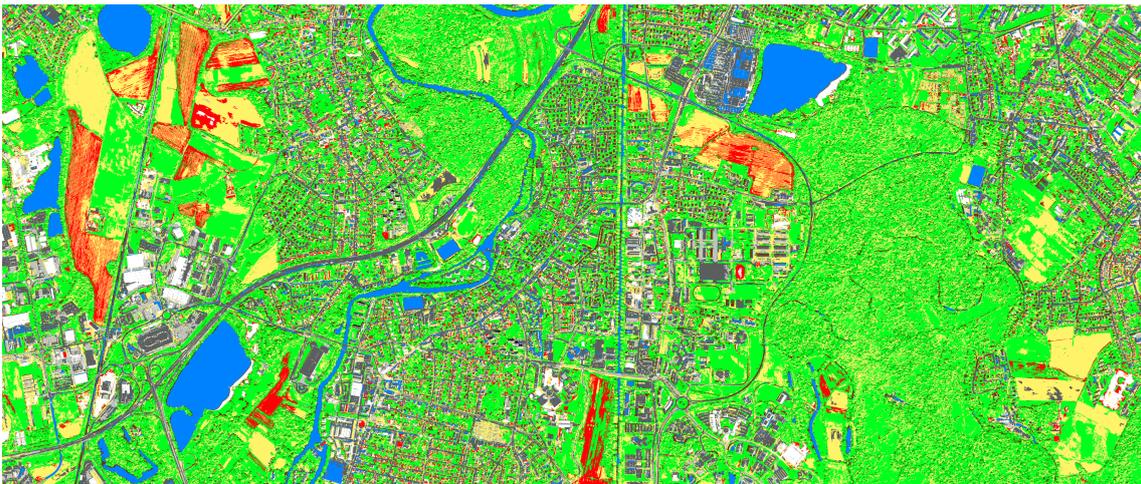
FIGURE 8.8 – Résultats de classification pour le jeu de données **ELSAU** (6 576 m × 2 793 m). Les couleurs des clusters ont été choisies pour correspondre aux couleurs des classes dans les cartes de vérité terrain.



(a) Classification MRS – niveau des quartiers.

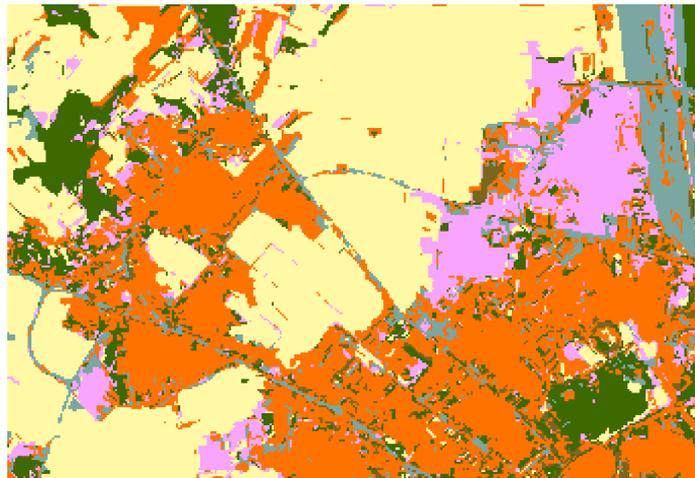


(b) Classification HRS – niveau des blocs.

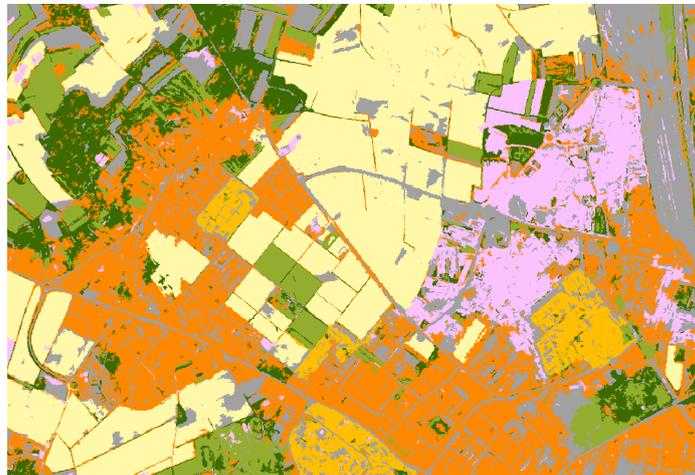


(c) Classification THRS – niveau des objets urbains.

FIGURE 8.9 – Résultats de classification pour le jeu de données **ILLKIRCH** (6 576 m × 2 803 m). Les couleurs des clusters ont été choisies pour correspondre aux couleurs des classes dans les cartes de vérité terrain.



(a) Classification MRS – niveau des quartiers.



(b) Classification HRS – niveau des blocs.



(c) Classification THRS – niveau des objets urbains.

FIGURE 8.10 – Résultats de classification pour le jeu de données **HAUTEPIERRE** (3 293 m × 2 246 m). Les couleurs des clusters ont été choisies pour correspondre aux couleurs des classes dans les cartes de vérité terrain.

F-mesure égales à ($\mathcal{K} = 0,38$ et $\overline{\mathcal{F}} = 0,66$) pour le jeu de données **ELSAU**, ($\mathcal{K} = 0,41$ et $\overline{\mathcal{F}} = 0,61$) pour le jeu de données **ILLKIRCH**, et ($\mathcal{K} = 0,32$ et $\overline{\mathcal{F}} = 0,63$) pour le jeu de données **HAUTEPIERRE**. Les classes thématiques les mieux extraites sont celles correspondant aux quartiers résidentiels (1), aux quartiers spécifiques (2) et aux zones agricoles (5). Cependant, la méthodologie ne permet pas d'extraire directement (d'une manière pertinente), à cette échelle, la classe des routes (4). On constate que les valeurs de F-mesure associées à cette classe sont faibles dans deux des trois jeux de données ($\mathcal{F} = 0,08$ pour le jeu de données **ELSAU** et $\mathcal{F} = 0,16$ pour le jeu de données **ILLKIRCH**).

L'étape 2 a été appliquée sur les images HRS pour partitionner les quartiers urbains (extraits au cours de l'étape précédente) en différents segments pouvant correspondre à des quartiers urbains mixtes, des sous-quartiers commerciaux ou industriels, des blocs urbains d'habitations, *etc.* Les résultats de segmentation/classification obtenus sont présentés par les figures 8.8(b), 8.9(b), 8.10(b). Les comparaisons entre les régions classifiées résultantes et les cartes de vérité terrain permettent d'obtenir des valeurs de Kappa et de F-mesure égales à ($\mathcal{K} = 0,38$ et $\overline{\mathcal{F}} = 0,69$) pour le jeu de données **ELSAU**, ($\mathcal{K} = 0,41$ et $\overline{\mathcal{F}} = 0,62$) pour le jeu de données **ILLKIRCH**, et ($\mathcal{K} = 0,34$ et $\overline{\mathcal{F}} = 0,64$) pour le jeu de données **HAUTEPIERRE**. Les classes relatives aux blocs urbains industriels et d'habitations (1-4) sont particulièrement bien extraites. Les valeurs de F-mesure associées à ces classes sont proches de 0,60 pour tous les jeux de données. On peut cependant remarquer que bien que certains résultats de partition soient composés de différentes régions correspondant correctement aux objets d'intérêt urbains recherchés, ces partitions sont aussi composées de nombreux petits segments représentant les structures linéaires de la scène et les zones de végétation. Ces problèmes de sur-segmentation sont probablement liés au critère géométrique utilisé par l'algorithme de construction de BPT (*i.e.*, le critère d'élongation), qui n'est pas adapté pour l'extraction de zones de végétation.

L'étape 3 a été appliquée sur les images THRS pour extraire les objets urbains simples (*e.g.*, bâtiments de l'habitation collective/individuelle, végétation, rues/parkings, ombres portées) à partir des blocs urbains extraits précédemment. Les résultats de segmentation et de classification obtenus sont présentés par les figures 8.8(c), 8.9(c), 8.10(c). Du fait de la non disponibilité de toutes les informations de classe pour les cartes de vérité terrain correspondant aux images THRS, nous avons uniquement évalué les précisions relatives aux classes de bâtiments. Les comparaisons entre les régions classifiées résultantes et les cartes de vérité terrain permettent d'obtenir des valeurs de Kappa et de F-mesure égales à ($\mathcal{K} = 0,38$ et $\mathcal{F} = 0,71$) pour le jeu de données **ELSAU**, ($\mathcal{K} = 0,39$ et $\mathcal{F} = 0,77$) pour le jeu de données **ILLKIRCH**, et ($\mathcal{K} = 0,39$ et $\mathcal{F} = 0,68$) pour le jeu de données **HAUTEPIERRE**. De plus, les comparaisons réalisées visuellement montrent que les autres catégories d'objets d'intérêt urbains simples sont extraits de façon pertinente, bien que les bords des segments ne soient pas parfaits.

Pour chaque résultat de segmentation/classification, on peut aussi remarquer le raffinement progressif des partitions initiales. Ce phénomène est obtenu par le biais des informations additionnelles apportées par les résolutions les plus fines. Une illustration de ce phénomène est présentée par la figure 8.11.

Comparaison à d'autres méthodes La plupart des approches proposées dans la littérature sont basées sur des critères d'homogénéité radiométrique [Inglada, 2007; Özdemir et al., 2010] (certaines de ces approches utilisent notamment des stratégies de segmentation hiérarchique [Akçay et Aksoy, 2008]). De telles méthodes sont efficaces pour extraire/classifier des objets urbains simples, mais semblent moins pertinentes pour détecter des structures plus larges, ce qui est le principal avantage du processus PCSC. Nous avons ainsi choisi d'évaluer séparément l'efficacité relative à l'extraction des objets d'intérêt simples et l'efficacité relative à l'extraction des objets d'intérêt les plus complexes.

Les méthodes auxquelles PCSC peut être comparé, et en particulier celles considérées dans le cadre de ces travaux ([Inglada, 2007; Akçay et Aksoy, 2008; Özdemir et al., 2010] pour l'extraction des objets d'intérêt simples, et [Gaetano et al., 2009] pour l'extraction des objets d'intérêt complexes) ne peuvent pas être directement évaluées sur les jeux de données décrits dans cette étude expérimentale (en particulier du fait de la non disponibilité de ces méthodes, et de la difficulté de les reprogrammer correctement en nous basant uniquement sur les détails fournis dans les publications). En conséquence, nous avons choisi de comparer ces méthodes à PCSC en se basant sur les scores de qualité fournis dans ces publications.

L'évaluation de la qualité des résultats obtenus sur les trois jeux de données montre que les pourcentages d'objets d'intérêt simples correctement reconnus/extraits sont comparables à ceux obtenus dans

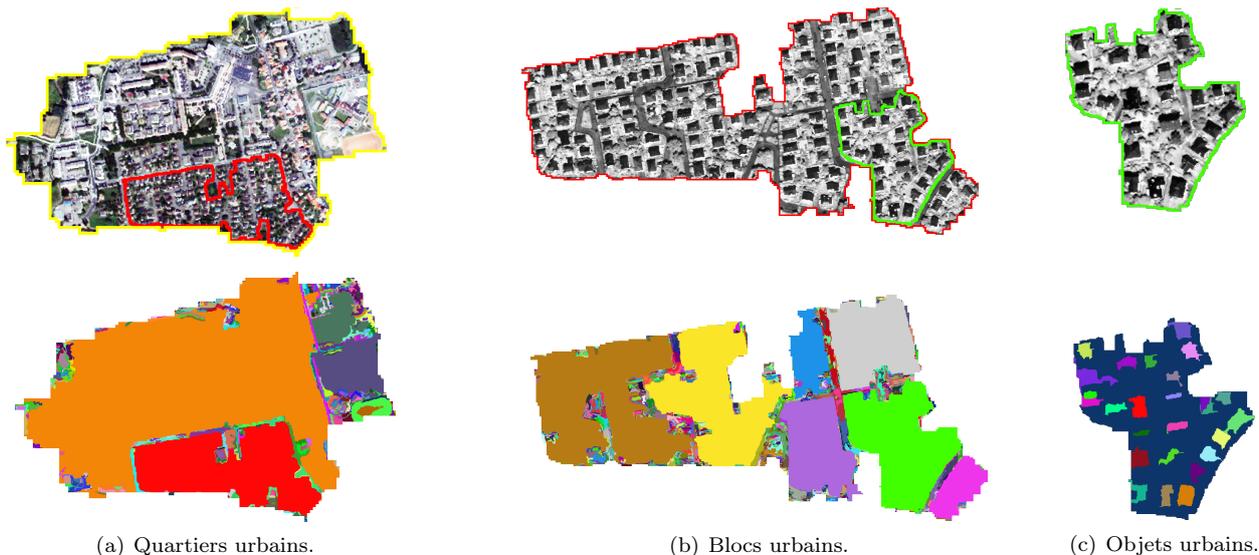


FIGURE 8.11 – Illustration d’une partie des résultats de segmentation du jeu de données **ELSAU**. (a) Segmentation HRS. (b–c) Segmentation THRS.

[Özdemir et al., 2010], où six différentes approches (dédiées à l’extraction d’objets d’intérêt simples) sont présentées. Le processus multirésolution proposé dans le cadre de ces travaux permet d’extraire, en moyenne, un pourcentage satisfaisant des objets d’intérêt simples ($\overline{\mathcal{K}} = 0,38$ et $\overline{\mathcal{F}} = 0,71$) tandis que les pourcentages de bâtiments individuellement correctement extraits atteignent 78% dans [Özdemir et al., 2010], où des indices de qualité similaires sont utilisés. La différence de précision peut être expliquée par le fait que les six approches présentées dans cet article ont été élaborées de manière *ad hoc* pour des besoins spécifiques, et requièrent ainsi une importante supervision (*e.g.*, de nombreux seuils doivent être fixés et/ou des marqueurs doivent être positionnés sur les bâtiments). Nous avons aussi comparé la précision obtenue *via* PCSC à la précision obtenue par l’intermédiaire d’une approche hiérarchique [Akçay et Aksoy, 2008]. Cette dernière a été appliquée pour extraire différentes catégories d’objets d’intérêt simples (*e.g.*, routes, bâtiments, végétation urbaine). Les scores obtenus pour la classe de bâtiments atteignent ($\overline{\mathcal{P}} = 0,69$ et $\overline{\mathcal{R}} = 0,77$) ce qui est aussi comparable à ceux obtenus par l’intermédiaire de PCSC ($\overline{\mathcal{P}} = 0,79$ et $\overline{\mathcal{R}} = 0,66$).

En ce qui concerne l’extraction des objets d’intérêt complexes (comme les blocs ou les quartiers urbains), nous avons comparé les résultats obtenus par l’intermédiaire de PCSC à ceux présentés dans [Gaetano et al., 2009], où une approche de segmentation ascendante (traitant des images multirésolutions) est proposée. Cette dernière a été utilisée pour extraire des objets urbains complexes (*e.g.*, grands blocs urbains, parkings, zones résidentielles) à partir de couples d’images HRS/THRS captées par le satellite IKONOS. Les scores des résultats obtenus en utilisant cette méthodologie atteignent $\overline{\mathcal{K}} = 0,35$, ce qui est similaire aux scores obtenus avec PCSC ($\overline{\mathcal{K}} = 0,38$ et $\overline{\mathcal{F}} = 0,69$ pour les blocs urbains, et $\overline{\mathcal{K}} = 0,38$ et $\overline{\mathcal{F}} = 0,66$ pour les quartiers urbains).

En conclusion de cette étude comparative, on peut observer que le processus PCSC fournit des résultats de segmentation/classification comparables aux résultats des approches similaires que l’on trouve dans la littérature. De plus, PCSC est conçu pour extraire simultanément différents niveaux d’objets d’intérêt, ce qui n’est pas toujours le cas en utilisant d’autres méthodes, ne permettant l’extraction que d’un seul niveau de sémantique.

8.3.3 Robustesse

A. Déterminisme

Protocole suivi Pour évaluer le degré de déterminisme du processus PCSC et en particulier celui de l’approche de segmentation hiérarchique, nous avons appliqué le protocole suivant.

L’approche de segmentation basée sur des exemples a été appliquée p fois en faisant varier les exemples

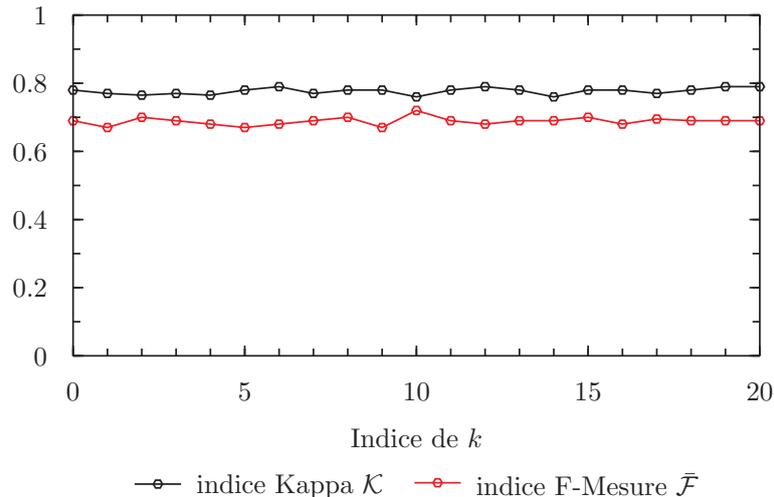


FIGURE 8.12 – Impact du choix de l'exemple de coupe parmi $p = 20$ possibilités. L'évolution de l'indice de qualité \mathcal{K} (respectivement, l'indice $\bar{\mathcal{F}}$) est représentée en noir (respectivement en rouge).

de coupes fournis par l'utilisateur. Pour ce faire, pour un ensemble homogène particulier de p imagettes à segmenter ($p = 20$ dans l'expérience considérée), nous avons exhaustivement choisi un BPT parmi les p BPT à couper. Ce BPT a ensuite été coupé interactivement par l'utilisateur. Pour chaque expérience, la coupe obtenue a été utilisée pendant l'étape d'apprentissage afin de définir le modèle de segmentation. Après l'étape de reproduction automatique, le résultat a été évalué, et le processus a été répété en utilisant un autre BPT pour définir l'exemple de coupe (parmi les p arbres possibles).

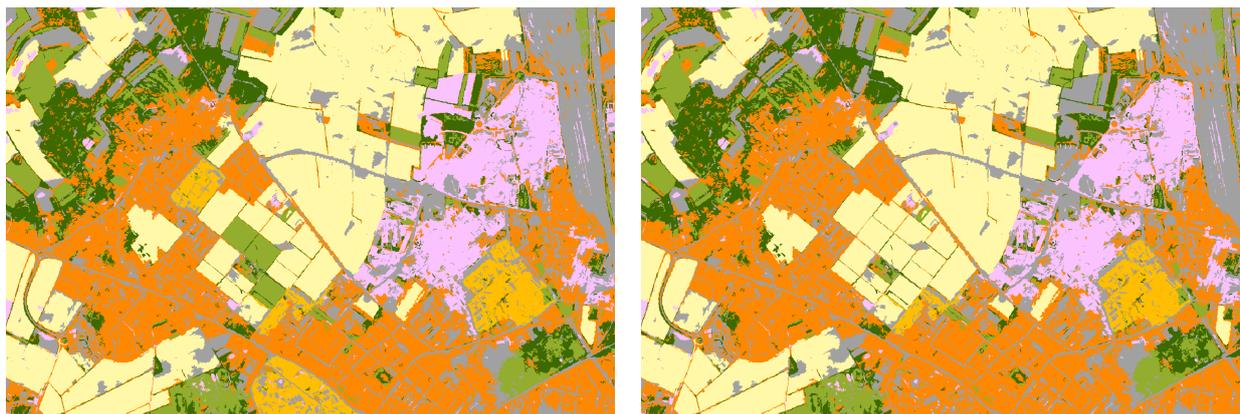
Analyse des résultats Les résultats obtenus sont présentés par la figure 8.12. À partir de ce graphe, on observe que le choix de la région qui est utilisée pour fournir l'exemple de coupe n'a pas d'impact notable sur la qualité des résultats. En effet, les valeurs de Kappa et de F-mesure obtenues demeurent similaires suivant les régions utilisées pour définir les exemples de coupes ($\mathcal{K} = 0,79 \pm 0,02$, $\bar{\mathcal{F}} = 0,70 \pm 0,04$). Ce comportement peut être expliqué par le fait que les p imagettes à segmenter représentent des contenus de l'image similaires car ces dernières ont été précédemment regroupées dans le même cluster par l'approche de classification multirésolution. Ainsi, le choix d'une imagette particulière (parmi l'ensemble des imagettes) pour effectuer l'exemple de coupe n'affecte pas la qualité des résultats obtenus par l'approche proposée.

B. Adaptation au domaine

Maximiser le potentiel des exemples de segmentations Partant du constat que la définition des exemples de segmentations (sur un jeu de données spécifique) est une tâche consommatrice de temps, il semble pertinent de maximiser leur potentiel en les réutilisant pour traiter d'autres jeux de données. Pour ce faire, de nombreuses approches ont été proposées pour effectuer le transfert de modèles d'apprentissage [Daumé III et Marcu, 2006]. En particulier, les approches d'adaptation au domaine permettent de profiter des avantages des connaissances apprises sur un jeu de données source, dans le but de déduire un modèle approprié pour le traitement d'un jeu de données similaire. Ces approches ont déjà été appliquées avec succès pour la classification d'images satellites [Bruzzone et al., 1999; Bahirat et al., 2012].

Pour évaluer la robustesse de PCSC, nous proposons d'étendre le paradigme d'adaptation au domaine pour permettre le transfert des exemples de segmentations requis par ce dernier pour le traitement d'un jeu de données source, dans le but de traiter automatiquement un jeu de données cible.

Transfert des éléments appris Une fois qu'un jeu de données source a été traité (*i.e.*, segmenté et classifié), il est possible de transférer les exemples de segmentations et les clusters appris pour traiter un



(a) Résultat de segmentation/classification.

(b) Résultat de l'expérience d'adaptation au domaine.

FIGURE 8.13 – Illustration des résultats de l'expérience d'adaptation au domaine pour le jeu de données **HAUTEPIERRE**. Le jeu de données **ELSAU** a été utilisé comme jeu de données source.

Jeu de données	Niveau	Indices d'évaluation	
		$\overline{\mathcal{F}} \pm \sigma$	$\mathcal{K} \pm \sigma$
ELSAU	1. Quartiers	0,56 ± 0,02	0,38 ± 0,02
	2. Blocs	0,69 ± 0,01	0,39 ± 0,01
	3. Objets	0,72 ± 0,03	0,36 ± 0,04
HAUTEPIERRE	1. Quartiers	0,67 ± 0,03	0,37 ± 0,01
	2. Blocs	0,64 ± 0,04	0,38 ± 0,05
	3. Objets	0,68 ± 0,01	0,39 ± 0,01
HAUTEPIERRE	1. Quartiers	0,65 ± 0,01	0,36 ± 0,01
Adaptation au domaine	2. Blocs	0,61 ± 0,02	0,37 ± 0,03
	3. Objets	0,64 ± 0,02	0,36 ± 0,01

TABLE 8.6 – Évaluation globale des résultats de l'expérience d'adaptation au domaine.

ensemble d'images multirésolution cible. Pour ce faire, on peut transférer du jeu de données source au jeu de données cible les éléments suivants :

1. les Q exemples de coupes fournis par l'utilisateur (pour chaque classe thématique de chaque résolution) durant l'approche de segmentation hiérarchique basée sur des exemples ($Q = q_1 + \dots + q_n$ ensembles d'histogrammes). Ces Q exemples de coupes peuvent ensuite être directement utilisés en entrée de l'approche de segmentation hiérarchique basée sur des exemples pour couper automatiquement les BPT associés aux n images du jeu de données cible ;
2. les K centroïdes (de chaque image) des clusters appris durant l'approche de classification multirésolution $K = k_1 + \dots + k_n$ ensembles de centroïdes. Ces K centroïdes peuvent ensuite être directement utilisés en entrée de l'approche de classification multirésolution pour classifier les segments extraits précédemment.

Du fait que les exemples de segmentations et les clusters appris sont liés à des zones thématiques spécifiques du jeu de données source, on peut faire l'hypothèse que le jeu de données cible représente les mêmes types de zones thématiques. Dans le cas où certaines de ces zones diffèrent du jeu de données source au jeu de données cible, la stratégie d'adaptation au domaine proposée peut être modifiée en ajoutant ou en supprimant des connaissances apprises lors de l'étape de transfert.

Expérimentations, résultats et analyses Dans un premier temps, le jeu de données **ELSAU** a été traité par l'intermédiaire du processus PCSC. L'approche de segmentation hiérarchique et l'approche de classification multirésolution ont été appliquées itérativement ($n = 3$) pour extraire les trois niveaux de sémantique

considérés. Pour ce faire, $Q = 16$ exemples de coupes ont été fournis par l'expert ($q_1 = 1, q_2 = 8, q_3 = 7$) et $K = 26$ clusters ont été extraits ($k_1 = 8, k_2 = 7, k_3 = 11$). La figure 8.10(c) présente le résultat de classification obtenu au niveau des blocs urbains sur le jeu de données **ELSAU**.

Pour valider la stratégie d'adaptation au domaine proposée, nous avons réalisé des expériences de transfert de connaissances. Le jeu de données **ELSAU** a été utilisé comme jeu de données source tandis que le jeu de données **HAUTEPIERRE** a été utilisé comme jeu de données cible. D'une manière plus pratique, les $Q = 16$ exemples de coupes et les $K = 26$ centroïdes des clusters résultant du traitement du jeu de données **ELSAU**, ont été transférés pour traiter le jeu de données **HAUTEPIERRE**. Cependant, comme les classes thématiques correspondant aux surfaces d'eau (classe bleu) et aux zones d'habitation denses (classe rouge) ne sont pas présentes dans le jeu de données **HAUTEPIERRE**, les exemples de segmentations et les clusters appris sur le jeu de données **ELSAU** et correspondant à ces classes n'ont pas été conservés durant l'étape de transfert.

Les résultats obtenus sont présentés dans la table 8.6 et dans la figure 8.13. À partir de ces résultats, on observe que les scores obtenus sur le jeu de données **HAUTEPIERRE** sont légèrement plus faibles quand les exemples de segmentations et les clusters appris sont transférés du jeu de données **ELSAU** (expérience d'adaptation au domaine (b)) que quand ces derniers sont appris directement sur le jeu de données (expérience (a)). Cependant, la qualité de ces résultats reste suffisante pour aider un utilisateur dans une tâche de cartographie. De plus, ces expériences ont montré qu'il est possible de maximiser le potentiel des exemples de segmentations définis par l'utilisateur pour traiter de grands jeux de données.

8.3.4 Complexité de calcul

Cette section propose une évaluation expérimentale de la complexité et des temps de calcul engendrés par l'utilisation du processus PCSC. Pour ce faire, nous comparons (en termes de temps de calcul et de mémoire) l'utilisation d'un BPT classique à l'utilisation de PCSC pour segmenter les images du jeu de données **ELSAU**. Finalement, nous comparons les temps de calcul de PCSC à ceux liés à l'application d'une approche de segmentation multirésolution descendante [Goffe et al., 2011].

Analyse expérimentale de la complexité La table 8.7 présente les temps de calcul et les consommations mémoire pour les segmentations et les classifications des images contenues dans les trois jeux de données. Les expérimentations ont été réalisées sur une machine équipée d'un processeur Intel® Core™2 Quad cadencé à 2,4 GHz avec 8 Go de RAM. Comme le montre la troisième colonne de la table 8.7, les traitements ont une complexité linéaire avec la taille des images. Ainsi, le traitement d'une image HRS qui contient 16 fois plus de pixels qu'une image MRS, requiert 16 fois plus d'opérations et de temps que le traitement d'une image MRS. Du fait que l'approche de classification multirésolution a une complexité linéaire, on peut faire l'hypothèse que l'approche de segmentation hiérarchique est aussi linéaire relativement aux données. Cependant, on remarque que la consommation mémoire reste non-négligeable dans le cas des images THRS (Table 8.7, colonne 4).

Comparaison des temps de calcul La table 8.8 permet de mettre en évidence les avantages de l'utilisation d'une stratégie descendante à travers la résolution : tandis que les BPT classiques ne permettent pas le traitement d'images THRS (*i.e.*, limitations mémoire), l'approche descendante proposée peut traiter des images de taille importante avec une consommation mémoire raisonnable. De plus, les temps de traitement sont comparables à ceux obtenus dans [Goffe et al., 2011], où un schéma de construction de pyramides irrégulières est présenté pour segmenter des images multirésolutions, en utilisant une stratégie descendante. Par exemple, cette approche requiert 44 min à 1 h 48 min pour traiter des grands jeux de données multirésolutions composés d'images à différentes résolutions spatiales (*e.g.*, 625×625 , $2\,500 \times 2\,500$, et $10\,000 \times 10\,000$ pixels).

8.4 Bilan scientifique

8.4.1 Contributions

Ce chapitre a présenté les expérimentations effectuées dans le cadre de la validation du processus PCSC, appliqué à l'extraction de hiérarchies d'objets d'intérêt à partir de trois ensembles d'images multirésolutions.

Jeu de données	Image (taille en pixels)	Temps de traitement	Mémoire (RAM)
ELSAU	MRS (685 × 291)	46 s	67 Mo
	HRS (2 740 × 1 164)	5 min 27 s	569 Mo
	THRS (10 960 × 4 656)	1 h 24 min	3,51 Go
ILLKIRCH	MRS (685 × 292)	56 s	69 Mo
	HRS (2 740 × 1 168)	5 min 48 s	583 Mo
	THRS (10 960 × 4 672)	1 h 36 min	3,59 Go
HAUTEPIERRE	MRS (343 × 234)	33 s	54 Mo
	HRS (1 372 × 936)	4 min 12 s	431 Mo
	THRS (5 488 × 3 744)	53 min	2,81 Go

TABLE 8.7 – Temps de calcul et consommation mémoire pour la segmentation et la classification des images contenues dans les jeux de données considérés.

Jeu de données	Image (taille)	BPT classique			PCSC		
		Temps de traitement	Taille BPT (nb. de nœuds)	Mémoire (RAM)	Temps de traitement	Taille BPT (nb. de nœuds)	Mémoire (RAM)
ELSAU	MRS (685 × 291)	46 s	$3,98 \times 10^4$	67 Mo	46 s	$3,98 \times 10^4$	67 Mo
	HRS (2 740 × 1 164)	3 min 49 s	$9,12 \times 10^6$	1,42 Go	5 min 27 s	$4,49 \times 10^5$	569 Mo
	THRS (10 960 × 4 656)	n/a	n/a	n/a	1 h 24 min	$2,94 \times 10^6$	3,51 Go

TABLE 8.8 – Comparaison des temps de calcul et des consommations mémoire entre le calcul d’un BPT classique et le processus PCSC. Les images sont celles du jeu de données **ELSAU**.

Les résultats de ces expérimentations ont montré que les cartes de segmentation et de classification obtenues par le biais de ce processus descendant peuvent être employées par un utilisateur pour accomplir des tâches de cartographie. D’une manière plus générale ces résultats mettent en évidence l’intérêt de l’utilisation d’une approche collaborative segmentation–classification, appliquée d’une manière descendante à travers la résolution, pour l’extraction d’objets complexes à partir d’images à (très) hautes résolution spatiales.

8.4.2 Valorisation scientifique

Ces différentes expérimentations ont donné lieu à une contribution [Kurtz et al., 2012b] dans la conférence internationale de télédétection *IEEE IGARSS 2012, International Geoscience & Remote Sensing Symposium*, publiée par *IEEE Geoscience & Remote Sensing Society*.

Troisième partie

Vers un processus itératif guidé par des connaissances

★ Résumé ★

Dans cette partie, nous étudions des mécanismes permettant de remettre en cause et corriger les résultats d'analyse obtenus aux différentes résolutions. Ces mécanismes sont principalement basés sur l'intégration de connaissances du domaine de haut niveau (apportées par un utilisateur) et de connaissances liées au contenu des images traitées (générées par l'intermédiaire du processus d'analyse descendant) dans un post-processus, appliqué à la suite du processus collaboratif descendant. Dans ce contexte, nous présentons une distance originale basée sur la sémantique issue des connaissances du domaine, permettant de comparer des données caractérisées par le biais d'attributs multirésolutions décrits sous la forme d'histogrammes. Nous évoquons ensuite la perspective d'un processus itératif permettant une re-segmentation et une re-classification « corrective » des résultats liés aux différents niveaux d'analyse, dans le but d'en améliorer la précision. Nous présentons une validation expérimentale de ce post-processus et de cette distance, dans le cadre de l'analyse d'environnements urbains à partir d'ensembles d'images satellites multirésolutions.

Connaissances et remise en cause des résultats

Sommaire

9.1 Limites de PCSC	111
9.1.1 Problèmes inhérents au processus	111
9.1.2 Intégration de connaissances : vers un processus correctif	113
9.2 Différents types de connaissances	114
9.2.1 Connaissances liées aux images	114
9.2.2 Connaissances du domaine	115
9.3 Remise en cause des résultats	117
9.3.1 Comment intégrer les connaissances ?	117
9.3.2 Quand intégrer les connaissances ?	118
9.3.3 Vers un post-processus de remise en cause des résultats	119

Nous décrivons dans ce chapitre l'intérêt de l'ajout de connaissances de haut niveau dans le processus global d'analyse d'images présenté précédemment. Pour ce faire, nous commençons par énoncer les limites et les problèmes de PCSC (Section 9.1). Nous introduisons ensuite l'ajout de connaissances de haut niveau, comme une solution possible pour améliorer ce processus (Section 9.2). Finalement, nous proposons un post-processus combinant plusieurs approches basées sur l'intégration de connaissances, permettant de remettre en cause et de corriger les résultats d'analyse initialement obtenus (Section 9.3).

9.1 Limites de PCSC

9.1.1 Problèmes inhérents au processus

Le processus PCSC, proposé dans la partie précédente, permet d'extraire et de classifier des hiérarchies de régions pouvant correspondre à des objets d'intérêt complexes, apparaissant comme hétérogènes au sein d'images (T)HRS. L'originalité de ce processus est d'extraire ces hiérarchies d'une manière descendante à partir d'images multirésolutions par l'intermédiaire d'une collaboration entre une approche de segmentation hiérarchique basée sur des exemples et une approche de classification non-supervisée, successivement appliquées à travers la résolution. De par sa nature non-supervisée, ce processus ne nécessite que très peu d'informations *a priori*, relatives au contenu de la scène observée et à la nature des objets d'intérêt à extraire.

Cependant, les approches non-supervisées employées peuvent présenter certaines limites, inhérentes à la forte hétérogénéité des images. Des erreurs peuvent par exemple avoir été commises lors de l'étape de segmentation et/ou lors de l'étape de classification, pour chacun des niveaux de résolution traités. La propriété descendante de ce processus va alors impliquer une propagation de ces erreurs dans les niveaux suivants.

Ce problème est principalement dû au fait que le processus actuel ne permet pas une correction *a posteriori* des erreurs de segmentation et de classification. Nous détaillons par la suite les sources potentielles de ces erreurs.



FIGURE 9.1 – Extrait d’une segmentation (dont les frontières des segments sont délimitées en rouge) réalisée sur une image MRS (9,6 m), projetée sur une image HRS (2,4 m). On peut observer que sur cette image HRS, les frontières des segments apparaissent comme crênelés.

A. Problèmes liés à l’approche de segmentation hiérarchique basée sur des exemples

La première limite de l’approche de segmentation hiérarchique basée sur des exemples (Chapitre 5) est directement liée au processus de reproduction des exemples de coupes définis par l’utilisateur. Dans le cas où ce processus commet une erreur dans la reproduction d’une coupe, la zone de l’image ne sera pas segmentée conformément aux attentes de l’utilisateur. Par exemple, une coupe exemple donnée par ce dernier est caractérisée par un niveau d’échelle particulier que l’utilisateur a choisi. Si l’algorithme d’escalade (défini dans la section 5.3) commet une erreur durant la reproduction de cette coupe exemple aux zones de l’image à segmenter, alors ces dernières ne seront pas partitionnées avec un niveau d’échelle similaire.

Le deuxième limite est inhérente au caractère descendant multirésolution du processus PCSC, et en particulier à la différence de résolution entre les images traitées. Les segments extraits à une résolution τ_t (pouvant correspondre à des objets d’intérêt complexes) par l’approche de segmentation sont ensuite projetés dans la résolution τ_{t+1} pour y être affinés et décomposés en segments pouvant correspondre à des objets plus élémentaires. Le problème est que l’approche de segmentation ne ré-adapte pas les bords des segments extraits à une résolution τ_t quand ceux-ci sont projetés dans une résolution τ_{t+1} . Comme la différence de résolution peut être importante entre les images considérées, les bords de ces segments peuvent apparaître comme crênelés et ne plus correspondre exactement aux frontières de ces objets d’intérêt dans l’image à la résolution τ_{t+1} . Ce phénomène est illustré par la figure 9.1, où une segmentation réalisée sur une image MRS est projetée sur une image HRS. Dans cet exemple, les quartiers urbains sont délimités en rouge. On peut noter que les frontières de ces derniers sont globalement correctes mais néanmoins affectées d’un problème de crênelage dû à l’importante différence de résolution entre l’image MRS et l’image HRS.

B. Problèmes liés à l’approche de classification multirésolution

Une première limite de l’approche de classification multirésolution (Chapitre 6) est liée à l’étape de caractérisation des segments produits par l’étape de segmentation. Durant cette étape, les segments extraits d’une image à une résolution τ_t sont caractérisés en fonction de leur composition à une résolution τ_{t+1} , en prenant en compte la distribution des labels de classes/clusters dans cette résolution plus fine. Or, ces classes/clusters sont obtenu(e)s à la résolution τ_{t+1} par le biais d’une approche de classification non-supervisée basée pixels, qui peut comporter des erreurs. En conséquence, les histogrammes de composition résultant de cette étape de caractérisation, et utilisés pour caractériser les segments, peuvent être erronés.

Une seconde limite est inhérente à l'utilisation d'une approche de classification non-supervisée pour labelliser ces régions. Un premier problème réside notamment dans le choix du nombre de clusters utilisés lors de cette classification (voir la section 1.2.3). Si le nombre de clusters choisi n'est pas adapté pour extraire l'ensemble des classes des images, alors les résultats de classification obtenus pourront ne pas être pertinents. Dans le cadre de cette approche, l'algorithme K -MEANS est utilisé. En conséquence, un deuxième problème est lié au choix de la position des centroïdes initiaux. Ce choix est effectué de manière aléatoire et peut ainsi conduire à des clusters ne reflétant pas correctement la structure des données, menant à des résultats de classification/labellisation comportant des erreurs. Finalement, un troisième problème est lié à la distance utilisée pour comparer les données (*i.e.*, les histogrammes de composition) lors de l'étape de classification. En effet, la plupart des algorithmes de classification non-supervisée requièrent une mesure de distance (et une moyenne) pour comparer les couples d'objets à classer. Dans le cas où ces données sont représentées par des histogrammes, la plupart des distances existantes ne permettent pas d'obtenir des résultats pertinents. En conséquence, cette étape de classification non-supervisée peut conduire à des erreurs de labellisation.

C. Propagation des erreurs durant la phase descendante

Finalement, la limite principale de PCSC est liée au principe descendant de ce processus : les erreurs de segmentation et de classification, potentiellement commises à un niveau d'analyse t , sont par construction propagées dans les niveaux suivants.

Considérons par exemple un segment correspondant à une zone de végétation, extrait à une résolution τ_t , et classifié avec le même label que les segments correspondant à des zones industrielles (*i.e.*, erreur de classification). Ce dernier sera alors projeté dans la résolution τ_{t+1} pour y être décomposé en segments plus élémentaires. Cependant, ce segment sera décomposé (par l'étape de segmentation hiérarchique basée sur des exemples) de la même façon qu'un segment correspondant à une zone industrielle (*i.e.*, avec un même niveau d'échelle) entraînant ainsi d'éventuelles erreurs de segmentation à la résolution τ_{t+1} . En conséquence, une erreur de classification commise à une résolution τ_t affectera potentiellement l'étape de segmentation réalisée à une résolution τ_{t+1} .

Considérons maintenant l'exemple d'une zone de l'image correspondant à un bloc urbain industriel analysable à une résolution τ_t . Faisons l'hypothèse que la segmentation réalisée au niveau t comporte des erreurs et qu'aucun segment ne correspond à ce bloc industriel (*i.e.*, erreur de segmentation). Ce segment erroné sera ensuite caractérisé par un histogramme de composition lui aussi potentiellement erroné, conduisant alors à une erreur de classification. Ainsi, une erreur de segmentation commise à une résolution τ_t affectera l'étape de classification réalisée à cette même résolution.

En conséquence, les erreurs d'analyse résultant du traitement d'une image à une résolution τ_t peuvent perturber les approches de segmentation et de classification réalisées au cours du traitement des images aux résolutions suivantes.

9.1.2 Intégration de connaissances : vers un processus correctif

Les limites du processus PCSC sont, d'une manière plus générale, inhérentes aux faiblesses des approches non-supervisées. Avec l'augmentation de la complexité des données, ces dernières peuvent montrer leurs limites et produire des résultats qui parfois ne sont pas suffisamment pertinents pour les besoins applicatifs de l'utilisateur. L'augmentation du nombre de données, du nombre d'attributs, ou encore du nombre de clusters, sont des exemples des problèmes les plus courants.

Il est à noter que bien que des informations relatives au contenu des images soient générées durant l'analyse des données aux résolutions les plus fines, elles ne sont pas prises en compte ailleurs pour guider les résultats obtenus aux résolutions les plus grossières. Ainsi, l'information multirésolution prise en compte pendant la caractérisation des segments extraits à une résolution τ_t , relative aux clusters obtenus à la résolution τ_{t+1} , n'est pas prise en compte dans le sens inverse. Considérons l'exemple d'un bloc urbain d'habitations individuelles, extrait par l'étape de segmentation à une résolution τ_t . Durant l'étape de classification multirésolution, ce segment est caractérisé par un histogramme modélisant sa composition en termes de clusters dans la résolution τ_{t+1} . Ainsi, des informations relatives au contenu de l'image à la résolution τ_{t+1} sont utilisées lors de l'analyse de l'image à la résolution τ_t . Cependant, aucune connaissance liée au contenu de la scène à la résolution

τ_{t-1} n'est utilisée lors du traitement de la résolution τ_t . Or, il pourrait être pertinent d'utiliser ce type de connaissances durant la totalité du processus d'analyse. En effet, savoir par exemple que la région que l'on est en train de classifier à la résolution τ_t avec un label de bloc urbain industriel se trouve, à la résolution τ_{t-1} , dans un cluster correspondant à une zone de végétation, peut apporter une information quant à la pertinence du label affecté à ce segment.

De façon plus générale en fouille de données complexes, pour tenter d'apporter une solution à ces problèmes, une approche communément employée par la communauté scientifique consiste à utiliser des connaissances lors du processus d'analyse. Ces connaissances, sont intégrées de manière croissante dans les processus de fouille de données, et permettent ainsi de guider ces derniers vers de meilleurs résultats. De nombreux travaux [Yoon et al., 1999; Kopanas et al., 2002; Forestier, 2010] ont montré le rôle important joué par ces connaissances du domaine (ainsi que par l'utilisateur) dans le processus de fouille de données. D'après les conclusions de ces travaux, le processus d'extraction d'informations à partir de données ne peut souvent pas être (totalement) automatique, et il est nécessaire d'étudier des mécanismes permettant d'établir des interactions entre, d'une part, les traitements automatisables et, d'autre part, la supervision de l'utilisateur qui guide ces traitements.

Pour résoudre les différents problèmes du processus PCSC, nous proposons d'introduire des mécanismes de prise en compte des connaissances au sein de ce dernier. L'objectif est d'utiliser ces connaissances pour remettre en cause et corriger les différents niveaux de segmentation et de classification obtenus, dans le but d'aider l'utilisateur à mieux comprendre la scène représentée. Nous décrivons dans la section suivante les différents types de connaissances pouvant être considérés dans ce contexte.

9.2 Différents types de connaissances

Deux types de connaissances peuvent être considérés pour guider les algorithmes de segmentation et/ ou de classification [Baltsavias, 2004] :

1. les connaissances liées aux images (générées à partir de ces dernières par l'analyse préalable des différents niveaux de résolution *via* PCSC) ;
2. les connaissances spécifiques au domaine, apportées par l'utilisateur.

Pour chaque type de connaissances, deux sous-catégories peuvent être considérées : les connaissances dites monorésolutions et celles dites multirésolutions. Nous détaillons ci-après ces différents types de connaissances.

9.2.1 Connaissances liées aux images

A. Connaissances monorésolutions

Le traitement *via* PCSC des images considérées permet d'extraire de ces données des ensembles de segments/régions offrant des connaissances *a posteriori* relatives aux frontières potentielles des objets d'intérêt présents dans les images, ainsi qu'aux caractéristiques permettant de décrire ces objets dans ces dernières. De plus, on dispose de connaissances sur les regroupements de ces régions en clusters offrant des informations pertinentes quant aux zones de l'image partageant des caractéristiques communes. Il est également possible d'obtenir des informations relatives à l'organisation spatiale de ces clusters au sein d'une image.

Considérons l'exemple de trois clusters de régions (K_1 , K_2 et K_3) extraits d'une image à une résolution τ_t . En analysant les relations spatiales des régions issues de ces trois clusters, on peut trouver en général, des configurations atypiques et/ou fréquentes (*e.g.*, les régions de K_1 sont toujours adjacentes aux régions de K_2 , les régions de ce cluster étant caractérisées par de fortes valeurs d'élongation) entre ces dernières, probablement dues à l'organisation structurelle des objets d'intérêt représentés. On peut ainsi inférer des connaissances relatives à l'organisation spatiale de ces clusters. Ce type de connaissances est illustré par la figure 9.2. Dans cet exemple, les relations spatiales entre les clusters extraits de l'image sont modélisées par le biais de chaînes de Markov reflétant les probabilités de passage d'un cluster à un autre (*i.e.*, en considérant différentes directions géographiques).

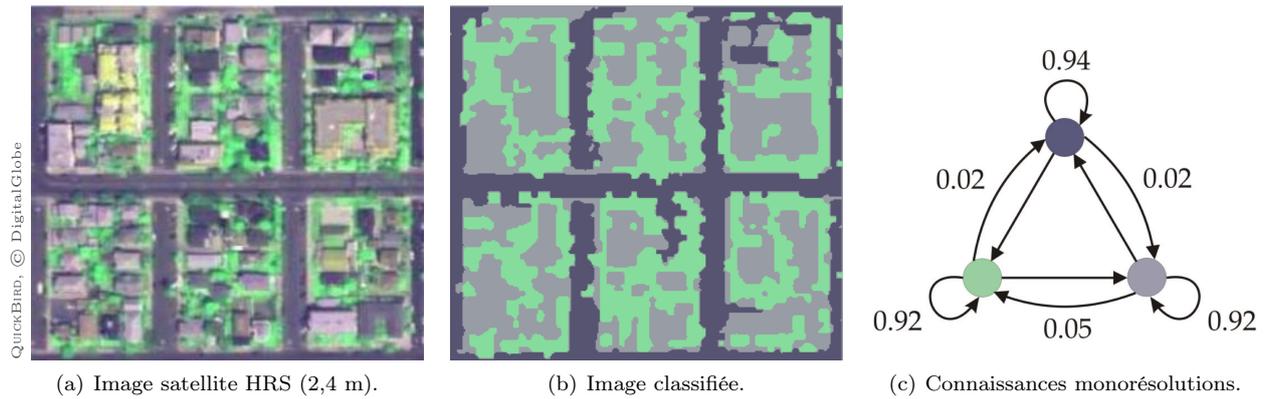


FIGURE 9.2 – Connaissances monorésolutions extraites d’un résultat de classification [Gaetano et al., 2009]. (c) Les relations spatiales entre les clusters sont modélisées par l’intermédiaire d’une chaîne de Markov.

B. Connaissances multirésolutions

L’organisation hiérarchique, à travers la résolution, des régions/clusters permet d’obtenir des connaissances multirésolutions relatives à l’organisation spatiale et structurelle, au sein des images, des objets d’intérêt.

Considérons l’exemple d’un cluster de régions (nommé K), extrait d’une image à une résolution τ_t . Le cluster K regroupe les régions de l’image correspondant à un certain type de couverture des sols. En construisant un histogramme de composition moyen pour le cluster K (relatif à sa composition en termes de clusters dans l’image à la résolution τ_{t+1}), on a l’information, qu’en moyenne dans les images considérées, les régions appartenant à ce cluster sont composées de ($x\%$ de régions catégorisées dans le cluster 0, \dots , $z\%$ de régions catégorisées dans le cluster 4) dans l’image à la résolution τ_{t+1} . La figure 9.3 illustre ce type de connaissances multirésolutions. Ces informations apportent ainsi des connaissances sur l’organisation structurelle des objets d’intérêt observables à partir d’une image à une résolution τ_t , en fonction d’objets d’intérêt observables à partir d’une image à une résolution τ_{t+1} .

Sur le même principe, l’organisation hiérarchique des régions/clusters apporte aussi des informations structurelles inverses (*i.e.*, relatives aux objets d’intérêt analysables à partir d’une image à une résolution τ_{t+1} en fonction d’objets d’intérêt observables à partir d’une image à une résolution τ_t). Par exemple, on peut savoir qu’en moyenne dans les images considérées, les régions du cluster 1 (extrait à la résolution τ_{t+1}) participent à la composition des clusters 4 et 5 (extraits à la résolution τ_t) mais ne participent jamais (ou très peu) à la composition du cluster 8 (extrait à la résolution τ_t).

9.2.2 Connaissances du domaine

Dans notre contexte, les connaissances du domaine sont des informations géographiques que l’utilisateur peut apporter au processus d’extraction d’informations. Ces connaissances sont des informations générales, relatives à un certain type d’environnement urbain, qui ne sont pas spécifiques au contenu de la scène représentée (ainsi qu’aux jeux de données considérés). Le but de cette section n’est pas de faire une revue exhaustive des connaissances du domaine disponibles (et des travaux relatifs à leur modélisation), mais d’en introduire certaines nous paraissant pertinentes pour remettre en cause les résultats obtenus précédemment.

A. Connaissances monorésolutions

Les connaissances du domaine monorésolutions sont des informations thématiques, relatives à l’occupation des sols observés à partir d’une image à une résolution spécifique, que l’utilisateur connaît *a priori*. Ces connaissances peuvent être liés aux types d’objets d’intérêt (et de classes) que l’on peut extraire de ces territoires urbains. Par exemple, l’utilisateur peut apporter des informations sur les classes thématiques (*e.g.*, nombres, labels) que l’on peut extraire d’un type précis d’environnements.

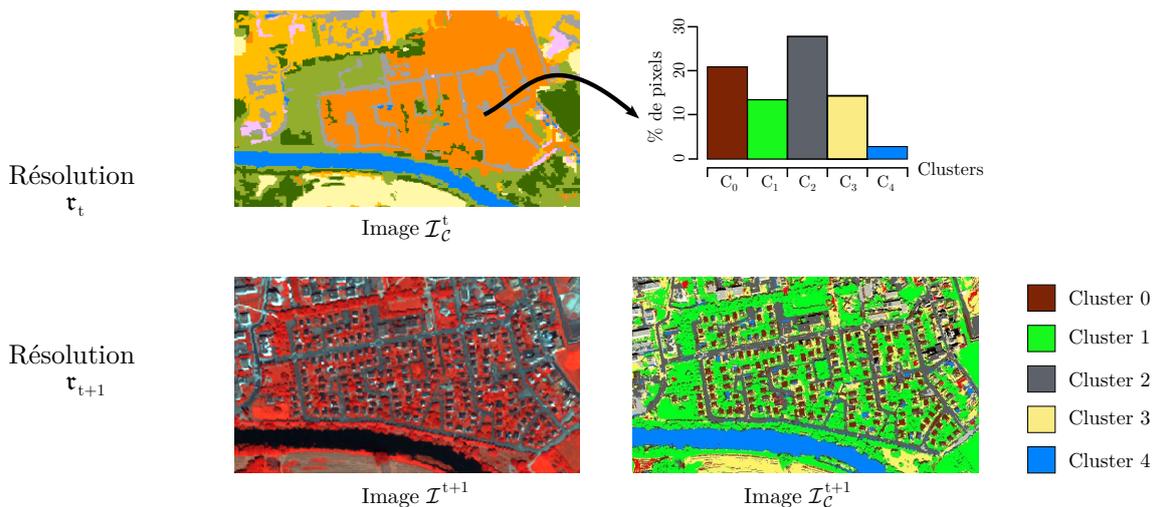


FIGURE 9.3 – Connaissances multirésolutions liées aux images : le cluster K , représenté en orange sur l'image \mathcal{I}_C^t , peut être caractérisé par un histogramme de composition modélisant la répartition de ses pixels en termes de labels de clusters dans l'image classifiée \mathcal{I}_C^{t+1} .

Pour ces classes thématiques, l'utilisateur peut ensuite fournir des connaissances de bas niveau comme des exemples d'objets étiquetés (similaires à ceux requis dans le cadre de la classification supervisée, voir la section 1.2.2), ou encore des connaissances exprimées sous la forme de relations *must-link* et *cannot-link* pour contraindre le processus de segmentation et/ou de classification. Un lien *must-link* indique que deux objets devraient être dans le même cluster à la fin du processus alors qu'un lien *cannot-link* signifie le contraire. Ce type de connaissances est parfois plus simple à obtenir qu'un ensemble classique d'objets étiquetés. En effet, il n'est pas nécessaire de connaître le nombre de classes, leurs étiquettes, leurs caractéristiques, *etc.* (seule l'information entre couples d'objets est modélisée).

Les connaissances du domaine monorésolutions peuvent aussi être relatives à l'organisation spatiale des objets géographiques dans l'image : les surfaces de végétation sont adjacentes aux surfaces d'eau, un toit de bâtiment est constitué de deux parties adjacentes (une face éclairée et une face à l'ombre), *etc.* Ce type de connaissances est définissable dans le cadre de l'organisation spatiale d'objets d'intérêt simples (*i.e.*, bâtiments, routes). Cependant, de telles connaissances ne sont pas définissables (et modélisables) dans le cadre de l'organisation spatiale d'objets d'intérêt complexes (*i.e.*, blocs urbains, quartiers urbains), dû à la forte diversité de ce type de structures urbaines. Ainsi, nous préférons ici utiliser des connaissances du domaine plus génériques.

Un autre type de connaissances est lié à la sémantique des classes recherchées : l'utilisateur peut apporter des connaissances relatives à la proximité sémantique de ces différentes classes. Considérons l'exemple de la classe correspondant aux blocs urbains d'habitations individuelles. Cette classe est plus proche sémantiquement (et thématiquement) de la classe correspondant aux blocs urbains d'habitations collectives qu'à celle correspondant aux blocs urbains industriels. Nous verrons par la suite que ces connaissances sémantiques peuvent être très utiles pour remettre en cause les labels de classes/clusters attribués aux régions des différentes résolutions.

B. Connaissances multirésolutions

Une connaissance du domaine qualifiée de multirésolution peut apporter des informations relatives à l'occupation des sols observés, à travers la résolution. De telles connaissances peuvent être relatives à l'organisation spatiale ou structurelle des objets d'intérêt (et des classes) que l'on peut extraire de ces territoires urbains aux différents niveaux d'échelle.

Par exemple, l'utilisateur peut savoir qu'en général, un objet d'intérêt complexe (*e.g.*, un bloc urbain) est composé d'une certaine proportion d'objets urbains simples, comme des toits de bâtiments, des parkings, des

routes ou des parcelles de végétation ou au contraire que l'on ne trouve jamais de bâtiments industriels au sein d'une zone agricole. De telles connaissances peuvent alors être utilisées pour remettre en cause les résultats de classification obtenus.

Les connaissances du domaine multirésolutions sont des informations difficiles à exprimer par l'utilisateur, et il n'existe à ce jour que très peu de bases de données (et/ou d'ontologies) répertoriant de telles informations.

9.3 Remise en cause des résultats

Dans la section précédente, nous avons vu que de nombreuses connaissances peuvent être extraites *a posteriori* des images *via* l'utilisation de PCSC. De plus, l'utilisateur peut apporter des connaissances du domaine *a priori*, relatives à la couverture des sols étudiés. Parallèlement, nous avons montré que le processus PCSC présente certaines limites et qu'une solution potentielle pour les corriger consiste à ajouter des connaissances au sein du processus de fouille de données. Nous analysons dans cette section comment et quand intégrer ces connaissances dans les différentes étapes de ce processus.

9.3.1 Comment intégrer les connaissances ?

Le processus PCSC est basé sur deux approches principales : une approche de segmentation hiérarchique et une approche de classification multirésolution. Chacune de ces deux approches peut générer des erreurs. Nous décrivons par la suite comment y intégrer des connaissances, dans le but de remettre en cause et de corriger les résultats.

A. Ajout de connaissances dans l'approche de segmentation hiérarchique

Nous avons vu précédemment que la première limite de l'approche de segmentation hiérarchique est directement liée au processus de reproduction des coupes exemples définies par l'utilisateur. Dans le cas où ce processus commet une erreur dans la reproduction d'une coupe, la zone de l'image segmentée ne correspondra pas forcément aux attentes de l'utilisateur (induisant des zones de l'image pouvant ne pas être correctement segmentées).

Il est possible d'utiliser des connaissances extraites des images pour corriger ce type de problèmes. L'idée sous-jacente est que, si un algorithme de segmentation commet une erreur lors du partitionnement d'une image à une résolution τ_t , cette erreur pourra être corrigée par les connaissances apportées par l'analyse d'une image à une résolution différente. De même, la classification des régions de l'image à la résolution τ_t peut aussi être utilisée pour remettre en cause cette segmentation. Par exemple, si le résultat de l'algorithme de classification comporte des régions « outliers », c'est-à-dire n'ayant pas été classifiées correctement (*e.g.*, éloignées de tous les centroïdes), on peut par exemple postuler que le segment correspondant à cette région n'a pas été délimité (et/ou caractérisé) de façon pertinente. Il est à noter que la remise en cause de la segmentation à une résolution τ_t implique une remise en cause de la classification à cette même résolution, car celle-ci en est directement dépendante.

La deuxième limite est liée au caractère descendant multirésolution du processus PCSC, et en particulier à la différence importante de résolution entre les images traitées. De fait, les bords des segments peuvent apparaître comme crénelés et ne plus correspondre exactement aux frontières des objets d'intérêt géographiques recherchés. Pour remettre en cause et corriger les frontières de ces segments, il nous semble possible d'utiliser des connaissances extraites des images aux résolutions les plus fines. Le principe est ici d'utiliser les informations géométriques des segments extraits à une résolution τ_t pour corriger les frontières des segments extraits à des résolutions plus grossières τ_{t-1} , τ_{t-2} , *etc.* Ce type d'approches a déjà été employé [Gaetano et al., 2009] et a montré des résultats prometteurs.

B. Ajout de connaissances dans l'approche de classification multirésolution

Nous avons vu précédemment que la première limite de l'approche de classification multirésolution est liée au fait que les clusters utilisés pour caractériser les segments peuvent comporter un certain nombre d'erreurs. Cependant, il est possible d'utiliser des connaissances extraites des images pour corriger ce problème. En effet,

à la fin du traitement de l'image à la résolution τ_{t+1} par PCSC, un résultat de classification basée régions de cette image est généré. Ce dernier est supposé plus pertinent que le résultat de classification basée pixels, utilisé lors de l'étape de caractérisation multirésolution des segments de l'image à la résolution τ_t . Il est alors possible de caractériser les segments de τ_t *via* les clusters de régions formés dans ce résultat de classification basée régions. Ces connaissances extraites des images peuvent permettre de remettre en cause et de corriger les résultats de classification des régions des images considérées.

La deuxième limite est inhérente à la distance utilisée pour comparer ces données décrites par des histogrammes de composition. La plupart des distances proposées dans la littérature ne sont pas pertinentes quand les valeurs des barres de ces histogrammes sont corrélées. Cependant, en considérant que l'utilisateur est capable (1) d'affecter aux étiquettes des clusters issus du résultat de classification obtenu à la résolution τ_{t+1} , une sémantique correspondant aux classes d'intérêt recherchées (*e.g.*, Cluster 1 \rightarrow Toits d'habitations rouges, ..., Cluster 7 \rightarrow Parcelles de végétation) et (2) d'évaluer des valeurs de proximité sémantique entre ces classes (*e.g.*, la classe Toits d'habitations rouges est plus proche sémantiquement de la classe Toits d'habitations en ardoise que de la classe Parcelles de végétation), il est possible d'utiliser ces connaissances pour améliorer la distance employée afin de comparer les données à classifier (et ainsi remettre en cause les résultats de classification à la résolution τ_t). Ces connaissances peuvent être utilisées pour pallier au problème de corrélation sémantique entre les barres des histogrammes.

La troisième limite est inhérente à l'utilisation d'une approche de classification non-supervisée pour labelliser ces régions. Pour apporter une solution au problème du choix du nombre de clusters utilisés, il est possible d'utiliser des connaissances extraites des différentes images. En effet, les résultats de classification obtenus à chaque résolution, peuvent apporter des informations quant au choix du nombre de clusters pour les algorithmes de classification. On peut par exemple faire collaborer ces résultats afin de trouver des nombres de clusters permettant d'unifier chacun des résultats de classification. Ainsi, les résultats de classification obtenus aux résolutions τ_{t-1} et τ_{t+1} peuvent aider à choisir un nombre de clusters pertinent pour la classification de la résolution τ_t .

Pour résoudre le problème lié au choix de la position des centroïdes initiaux, il est possible d'utiliser des connaissances du domaine multirésolutions, pour définir ces centroïdes d'une manière plus réaliste. Par exemple, des connaissances relatives à la composition moyenne des blocs urbains d'habitations individuelles en termes de pourcentages de routes, de toits d'habitations rouges, de toits d'habitations en ardoise, de parcelles de végétation peuvent être utilisées pour initialiser l'histogramme du centroïde.

9.3.2 Quand intégrer les connaissances ?

Nous avons montré dans la section précédente qu'il est possible d'intégrer différents types de connaissances, à plusieurs niveaux, dans les approches de segmentation et de classification associées au processus PCSC. Néanmoins, il est nécessaire d'étudier à quel moment considérer ces connaissances au sein de ce dernier. Deux solutions sont alors envisageables : (1) ajouter ces connaissances directement pendant l'exécution de PCSC ou (2) ajouter ces connaissances dans un post-processus permettant la remise en cause et la correction des résultats obtenus par PCSC.

Les connaissances générées par l'analyse des images (monorésolutions et multirésolutions) ne sont pas toutes disponibles au cours du processus descendant à travers la résolution. En effet, lors du traitement de l'image à la résolution τ_t , des connaissances relatives aux contenus des images τ_{t-1} , τ_{t-2} , *etc.* traitées dans les étapes précédentes sont disponibles. En revanche, aucune connaissance relative au contenu des images aux résolutions τ_{t+1} , τ_{t+2} , *etc.* n'est disponible pendant l'exécution du processus descendant.

Les connaissances du domaine sont des informations géographiques générales que l'utilisateur possède *a priori* sur les environnements étudiés. Il est alors envisageable de prendre en compte ces connaissances directement pendant le déroulement de PCSC à travers la résolution. Certaines connaissances du domaine sont d'ailleurs déjà intrinsèquement prise en compte dans l'approche de segmentation hiérarchique (*i.e.*, les exemples de coupes correspondent à des niveaux d'échelle) et dans l'approche de classification multirésolution (*i.e.*, le nombre de clusters).

Cependant, l'un des intérêts majeurs de PCSC est d'être basé sur des approches de segmentation et de classification non-supervisées. En conséquence, ce processus ne nécessite que très peu de connaissances prove-

nant des utilisateurs. Bien que les résultats de ce dernier puissent comporter des erreurs, la plupart des zones de la scène sont correctement classifiées. Il semble alors plus pertinent de maximiser l'aspect non-supervisé du processus et de n'appliquer qu'un post-processus correcteur, intégrant des connaissances du domaine et des connaissances extraites préalablement des images afin de corriger les éventuelles erreurs produites par ce processus descendant.

9.3.3 Vers un post-processus de remise en cause des résultats

A. Proposition

De nombreuses perspectives de recherche sont ouvertes par l'intégration de connaissances dans un post-processus appliqué à la suite de PCSC : certaines de ces perspectives sont réalisables à court terme tandis que d'autres le sont à plus long terme. Parmi ces perspectives, nous avons privilégié celles à court terme, et fait le choix d'orienter nos recherches vers :

1. **l'utilisation de connaissances extraites des images pour remettre en cause les résultats de segmentation** obtenus à partir des différentes résolutions, et améliorer ainsi l'approche de segmentation hiérarchique basée sur des exemples ;
2. **l'utilisation de connaissances du domaine au sein d'une distance entre histogrammes** pour améliorer l'approche de classification multirésolution ;

D'une manière plus générale, nos recherches se sont penchées sur **une réflexion relative à un post-processus de correction des résultats**, combinant ces deux approches, dédié à la correction des erreurs présentes dans les résultats d'analyse obtenus aux différentes résolutions. Ce post-processus rejoint ici les possibilités offertes par l'utilisation d'approches collaboratives permettant la remise en cause des résultats d'analyse obtenus aux différentes sources.

B. Principe de la méthodologie proposée

Le principe du post-processus correcteur proposé consiste à parcourir la hiérarchie de régions initialement obtenue, et à ré-appliquer l'approche de classification multirésolution en tenant compte de ces connaissances. Notre idée est de proposer une nouvelle distance pour la comparaison d'histogrammes, utilisant des connaissances du domaine monorésolutions couplées à des connaissances extraites des images (générées pendant la descente). Cette distance permet d'améliorer la comparaison d'histogrammes de composition et ainsi les résultats produits par l'approche de classification multirésolution. Nous verrons par la suite que la remise en cause des résultats de classification permet de remettre en cause les résultats de segmentation.

Les résultats de segmentation et de classification, obtenus à une résolution r_t , étant dépendants des résultats obtenus à une résolution r_{t-1} , nous proposons dans ces travaux que ce post-processus correcteur soit basé sur une approche ascendante. D'une manière plus générale, partant du fait que ces segments/clusters ont été produits par un processus descendant, sans connaissance *a priori*, nous proposons un schéma méthodologique permettant de remonter dans la hiérarchie (et dans la résolution) afin de corriger les segments/clusters en utilisant des connaissances acquises pendant la descente et d'autres données par l'utilisateur. Cette stratégie a pour avantages (1) de maintenir l'aspect non-supervisé de PCSC et (2) de permettre l'utilisation des connaissances générées à partir des images par l'application préalable de PCSC, et des connaissances du domaine fournies par l'utilisateur.

Dans le chapitre suivant (Chapitre 10), nous décrivons en détail une distance originale pour la comparaison d'histogrammes, utilisant des connaissances du domaine. Nous introduisons ensuite (Chapitre 11) comment cette distance peut être intégrée dans l'approche de classification multirésolution proposée précédemment et comment cette dernière peut être ré-appliquée dans le cadre d'un post-processus ascendant correcteur, désigné par la suite *via* l'acronyme C-PCSC.

HSBD, une distance entre histogrammes basée sur les connaissances

Sommaire

10.1 Mesurer la distance entre des attributs multirésolutions	121
10.1.1 Motivations	121
10.1.2 Problèmes liés à la comparaison d’histogrammes	123
10.2 Distances entre histogrammes	124
10.2.1 Distance entre les niveaux de la mesure	124
10.2.2 Comparaison d’histogrammes	125
10.2.3 Histogrammes et hiérarchies	128
10.3 Schéma méthodologique de HSBD	129
10.3.1 Principe général	129
10.3.2 Construction de la hiérarchie de fusion sémantique	129
10.3.3 Calcul de HSBD	131
10.4 Bilan scientifique	135
10.4.1 Contributions	135
10.4.2 Valorisation scientifique	135

Ce chapitre présente une nouvelle distance pour la comparaison de données caractérisées par des attributs multirésolutions, modélisés sous la forme d’histogrammes. Cette distance permet d’intégrer des connaissances du domaine apportées par l’utilisateur, relatives aux proximités sémantiques des barres des histogrammes, dans le calcul de cette dernière. Pour ce faire nous commençons par présenter les problèmes liés à la comparaison de données décrites par le biais d’histogrammes (Section 10.1). Nous détaillons ensuite différentes distances, présentes dans la littérature et permettant de comparer de telles structures, ainsi que leurs avantages et inconvénients respectifs (Section 10.2). Finalement, nous présentons une nouvelle distance basée sur un mécanisme hiérarchique et permettant de prendre en compte des connaissances du domaine pour la comparaison d’histogrammes (Section 10.3). Dans une dernière section (Section 10.4), nous énonçons le bilan scientifique de ces travaux.

10.1 Mesurer la distance entre des attributs multirésolutions

10.1.1 Motivations

Dans le cadre de nos travaux, nous proposons de caractériser les segments extraits des images par des attributs multirésolutions modélisés sous la forme d’histogrammes. Dans de nombreux domaines, comme ceux de la fouille de données et de l’analyse d’images, il est nécessaire de classifier d’importants jeux de données dans lesquels chaque élément à catégoriser est caractérisé par un ou plusieurs histogrammes. Par exemple, il est souvent nécessaire de classifier des populations en fonction de la distribution d’une mesure particulière (*e.g.*, la distribution de la taille des individus contenus dans ces populations). Les histogrammes sont donc des structures utiles pour modéliser de nombreux types de données et permettent de prendre en considération leurs propriétés statistiques.

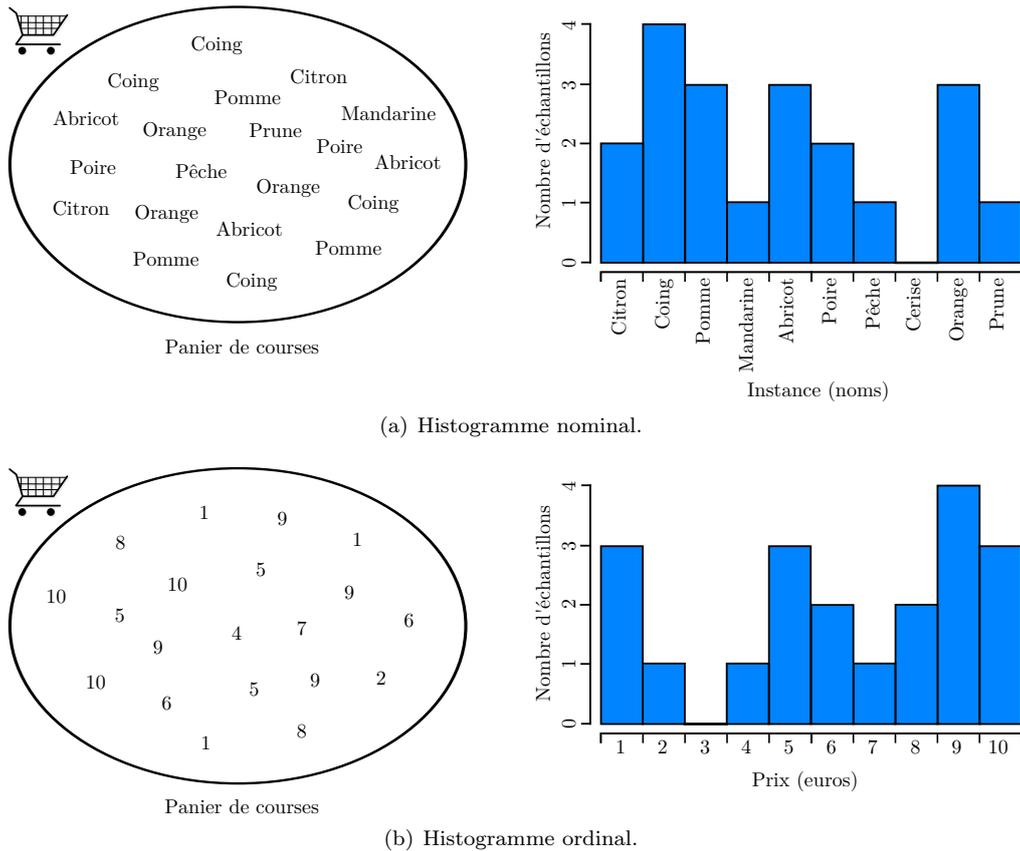


FIGURE 10.1 – Deux types d’histogrammes modélisant la composition d’un panier de courses. (a) Histogramme nominal modélisant la distribution des instances du concept sémantique FRUIT. (b) Histogramme ordinal modélisant la distribution des prix des articles.

Il existe différentes catégories d’histogrammes, relatives à des types de mesures spécifiques : nominales, ordinales (et modulo, qui représentent un cas particulier des mesures ordinales) [Cha et Srihari, 2002]. Dans une mesure nominale, chaque valeur est nommée et/ou peut représenter une instance d’un concept sémantique particulier (*e.g.*, le concept FRUIT peut prendre différentes valeurs/instances comme Citron, Prune, Pomme, Mandarine, Abricot, *etc.*). Ainsi, un histogramme de type nominal peut modéliser la composition d’un panier de courses en fonction du nombre et des types de fruits qu’il contient (Figure 10.1(a)). Dans un tel histogramme, les niveaux de la mesure peuvent être permutés, car il n’existe pas d’ordre (total) entre ces niveaux (propriété d’invariance à la permutation). Au contraire, dans une mesure ordinale, les valeurs sont totalement ordonnées (*e.g.*, le prix des légumes peut être quantifié en 10 valeurs discrètes entre 1 et 10 euros). Ainsi, un histogramme de type ordinal peut modéliser la composition d’un panier de courses en fonction du nombre et des prix des fruits qu’il contient (Figure 10.1(b)).

Mesurer la similarité entre des histogrammes est une opération cruciale au cours d’une étape de classification non-supervisée [Jain et al., 1999; Ng et al., 2005], mais également dans de nombreuses applications comme la reconnaissance de formes [Bressan et al., 2003], la recherche d’images basée sur le contenu [Brunelli et Mich, 2001; Nascimento et al., 2002; Al-Omari et Al-Jarrah, 2005; Liu et al., 2007], la classification de textes [Capdevila et Florez, 2009; Fabrizio, 2002] ou encore l’analyse de séries temporelles [Strelkov, 2008]. En effet, mesurer la distance entre un couple d’histogrammes permet d’évaluer la similarité entre les propriétés statistiques modélisées par ces structures. Au cours des dernières décennies, de nombreuses mesures de similarité entre histogrammes ont ainsi été proposées. On peut distinguer deux catégories principales de distances entre histogrammes : les distances *barre-à-barre* et les distances *barres-croisées*. Les distances *barre-à-barre* considèrent un histogramme comme un vecteur de dimension fixe et comparent uniquement le contenu des barres correspondantes des histogrammes, tandis que les distances *barres-croisées* considèrent un

	Citron	Coing	Pomme	Mandarine	Abricot	Poire	Pêche	Cerise	Orange	Prune
H_1	9	0	0	0	0	0	0	0	1	0
H_2	1	0	0	0	0	0	0	0	1	8
H_3	1	0	0	0	0	0	0	0	8	1

TABLE 10.1 – Histogrammes modélisant les compositions des trois paniers de courses C_1, C_2, C_3 , chaque panier étant composé de 10 fruits.

histogramme comme l'estimation d'une *fonction de densité de probabilité* et comparent aussi bien les barres correspondantes que les non-correspondantes.

10.1.2 Problèmes liés à la comparaison d'histogrammes

Dans ces travaux, nous nous intéressons à la comparaison de très grands jeux de données dans lesquels chaque donnée est caractérisée par un ou plusieurs histogrammes nominaux, pour lesquels des connaissances relatives à la proximité sémantique entre les barres peuvent être considérées. Pour illustrer ces histogrammes particuliers, on peut revenir à l'exemple présenté par la figure 10.1(a). Chaque barre de l'histogramme représente la proportion d'un type de fruits, qui est une instance du concept sémantique FRUIT. Étant donné que les citrons et les oranges sont des agrumes, l'instance Citron peut être considérée comme plus proche de l'instance Orange que de l'instance Prune. Il est ainsi possible d'évaluer les similarités sémantiques entre les barres composant cet histogramme nominal. Dans ce contexte, il est important de prendre en compte de telles similarités sémantiques pour améliorer la comparaison de ces histogrammes, et ainsi les approches de fouille de données dont les résultats dépendent de la pertinence de ces comparaisons.

D'une part, les distances barre-à-barre sont bien adaptées au traitement de grands jeux de données, en particulier grâce à leur faible coût de calcul. Cependant, par définition, de telles distances ne peuvent pas considérer les proximités sémantiques entre les différentes barres. (De plus ces distances souffrent des problèmes liés à la translation des barres et au choix de la taille des barres.) Considérons, par exemple, trois paniers de courses C_1, C_2, C_3 , représentés par la table 10.1. Les différents fruits peuvent être de type {Citron, Coing, ..., Prune}. Ainsi, la composition d'un panier de courses C_i peut être modélisée par un histogramme $H_i(\#Citron, \#Coing, \dots, \#Prune)$ où $\#x$ représente le nombre d'occurrences de l'instance x dans le panier C_i . En utilisant une distance barre-à-barre standard, par exemple la distance de Manhattan d_{L_1} , on obtient une distance similaire entre H_1 et H_2 et entre H_1 et H_3 ($d_{L_1}(H_1, H_2) = d_{L_1}(H_1, H_3) = 16$). Cependant, C_1 est sémantiquement plus proche de C_3 qu'il ne l'est de C_2 , car C_1 et C_3 sont tous les deux des paniers composés principalement d'agrumes.

D'autre part, les distances barres-croisées, qui comparent d'une manière plus exhaustive à la fois les barres correspondantes et non-correspondantes, permettent de considérer les proximités sémantiques entre les différentes barres. Ceci peut être réalisé en assignant à chaque paire d'instances un poids (*i.e.*, une valeur numérique) modélisant le degré de proximité sémantique entre les instances comparées. Considérons l'exemple du panier de courses. Le poids associé au couple d'instances (Citron, Orange) doit être plus faible que celui associé au couple (Citron, Prune) ou au couple (Orange, Prune). La contrepartie de cette stratégie est d'induire un temps de calcul quadratique pour la comparaison des barres des histogrammes (*i.e.*, chaque barre d'un histogramme H_1 doit être comparée à l'ensemble des autres barres d'un histogramme H_2), qui n'est pas compatible avec notre contexte méthodologique et nos besoins applicatifs.

En se basant sur ces considérations, on remarque que pour la plupart des méthodes dédiées à la comparaison d'histogrammes nominaux, la prise en compte de la proximité sémantique entre les instances modélisées par les barres des histogrammes est incompatible avec un faible temps de calcul. Dans ces travaux, nous proposons une nouvelle distance apportant une solution à ce problème. Son calcul est basé sur une stratégie hiérarchique, fusionnant progressivement les instances considérées (et les barres associées) en fonction de leurs proximités

sémantiques. Pour chaque niveau de cette hiérarchie, une distance barre-à-barre classique est calculée entre les paires d'histogrammes correspondantes. Dans le but d'obtenir la distance proposée, ces distances barre-à-barre sont ensuite fusionnées en prenant en compte la cohérence sémantique associée à chaque niveau de la hiérarchie. Cette distance préserve les avantages des distances barres-croisées (*i.e.*, robustesse à la translation des valeurs des histogrammes et au choix de la taille des barres), mais hérite néanmoins du faible temps de calcul des distances barre-à-barre.

10.2 Distances entre histogrammes

10.2.1 Distance entre les niveaux de la mesure

Un histogramme $H(A)$ représente la distribution des valeurs quantifiées d'une mesure x parmi les échantillons d'un ensemble A . Relativement à deux types de mesures (*i.e.*, nominales et ordinales), nous définissons deux fonctions d_{nom} et d_{ord} qui mesurent la différence entre deux valeurs de mesure $x_i, x_j \in X$. Dans la littérature, la différence entre deux valeurs (ou niveaux) de mesure est appelée la « distance au sol » (*ground distance*).

A. Mesure ordinale

Dans une mesure ordinale, les valeurs x_i sont totalement ordonnées, et il est possible de déterminer une distance atomique $\Delta_{(x_i, x_{i+1})} \in \mathbb{R}_+$ entre chaque paire de niveaux successifs x_i et x_{i+1} . Ainsi, nous définissons la distance au sol entre deux valeurs d'une mesure ordinale x_i et x_j comme la somme des distances atomiques entre chaque niveau successif de i à j

$$d_{ord}(x_i, x_j) = \sum_{k=i}^{j-1} \Delta_{(x_k, x_{k+1})} \quad (10.1)$$

Quand les valeurs de la mesure sont numériques (*i.e.*, chaque $x_i \in \mathbb{R}$), la distance au sol entre deux niveaux d'une mesure ordinale est la différence absolue entre ces niveaux

$$d_{ord}(x_i, x_j) = \sum_{k=i}^{j-1} |x_k - x_{k+1}| = |x_i - x_j| \quad (10.2)$$

B. Mesure nominale

Dans une mesure nominale, deux cas sont envisageables :

1. il n'est pas possible de définir des relations de proximité entre les valeurs x_i . Ainsi, nous définissons la distance au sol entre ces valeurs comme valeurs « correspondantes » ou « non-correspondantes »

$$d_{nom}(x_i, x_j) = \begin{cases} 0 & \text{si } x_i = x_j \\ 1 & \text{sinon} \end{cases} \quad (10.3)$$

2. il est possible de définir des relations de proximité sémantique entre les valeurs x_i . Comme la définition précédente de d_{nom} ne permet pas de considérer les proximités sémantiques entre les barres d'un histogramme nominal, nous étendons la définition de la distance au sol entre deux niveaux de la mesure nominale comme

$$d_{nom}(x_i, x_j) = d_{nom}(x_j, x_i) = \begin{cases} 0 & \text{si } x_i = x_j \\ \alpha_{(x_i, x_j)} & \text{sinon} \end{cases} \quad (10.4)$$

où $\alpha_{(x_i, x_j)} \in]0, 1]$ reflète une valeur de dissimilarité sémantique entre x_i et x_j reposant sur les connaissances du domaine de l'utilisateur. Ainsi, il est possible de définir une matrice de dissimilarité \mathcal{M}^{dis} de taille $v \times v$ qui modélise les relations entre chaque instance $x \in X = \{x_0, x_1, \dots, x_{v-1}\}$ du concept modélisé par l'histogramme

$$\mathcal{M}^{dis} = \begin{bmatrix} \alpha_{(x_0, x_0)} & \cdots & \alpha_{(x_0, x_{v-1})} \\ \vdots & \ddots & \vdots \\ \alpha_{(x_{v-1}, x_0)} & \cdots & \alpha_{(x_{v-1}, x_{v-1})} \end{bmatrix} \quad (10.5)$$

x_i	Citron	Coing	Pomme	Mandarine	Abricot	Poire	Pêche	Cerise	Orange	Prune
Citron	0,00	0,80	0,90	0,20	0,70	0,80	0,90	0,80	0,10	0,85
Coing	-	0,00	0,20	0,75	0,40	0,20	0,45	0,50	0,78	0,48
Pomme	-	-	0,00	0,90	0,40	0,02	0,50	0,45	0,95	0,45
Mandarine	-	-	-	0,00	0,92	0,85	0,75	0,90	0,15	0,95
Abricot	-	-	-	-	0,00	0,40	0,15	0,07	0,90	0,10
Poire	-	-	-	-	-	0,00	0,40	0,40	0,90	0,40
Pêche	-	-	-	-	-	-	0,00	0,10	0,90	0,05
Cerise	-	-	-	-	-	-	-	0,00	0,90	0,10
Orange	-	-	-	-	-	-	-	-	0,00	0,90
Prune	-	-	-	-	-	-	-	-	-	0,00

TABLE 10.2 – Matrice de dissimilarité \mathcal{M}^{dis} associée aux instances du concept FRUIT. Comme d_{nom} est symétrique, \mathcal{M}^{dis} est une matrice symétrique (nous ne représentons ici que le triangle supérieur droit).

Nous présentons dans le tableau 10.2 un exemple de matrice de dissimilarité pour le concept FRUIT¹ introduit dans la section précédente.

C. Propriété d'une métrique

Les mesures d_{ord} et d_{nom} définies dans les équations (10.2–10.3) satisfont les propriétés suivantes d'une métrique [Cha et Srihari, 2002] et peuvent donc être considérées comme des distances

1. non-négativité : $d(x_i, x_j) \geq 0$;
2. symétrie : $d(x_i, x_j) = d(x_j, x_i)$;
3. identité : $d(x_i, x_j) = 0 \Leftrightarrow x_i = x_j$;
4. inégalité triangulaire : $d(x_i, x_k) \leq d(x_i, x_j) + d(x_j, x_k)$.

Par construction, la mesure définie par l'équation (10.4) satisfait, au moins, les propriétés suivantes

1. non-négativité : $d_{nom}(x_i, x_j) \geq 0$;
2. symétrie : $d_{nom}(x_i, x_j) = d_{nom}(x_j, x_i)$;
3. identité : $d_{nom}(x_i, x_j) = 0 \Leftrightarrow x_i = x_j$.

Cependant, comme cette mesure est construite sur des notions relatives à la sémantique et à la perception humaine, la propriété d'inégalité triangulaire n'est généralement pas satisfaite. En conséquence, cette version étendue de d_{nom} , n'est pas une distance.

10.2.2 Comparaison d'histogrammes

Au cours des dernières décennies, de nombreuses mesures de similarité ont été proposées. Nous présentons ci-après les distances les plus couramment employées dans la littérature.

1. Les valeurs de dissimilarité contenues ont été définies dans un but d'exemple et ne correspondent pas à une réalité sémantique.

A. Distances barre-à-barre

Comme nous l'avons évoqué précédemment, les distances barre-à-barre considèrent un histogramme comme un vecteur de dimension fixe, et comparent uniquement le contenu des barres correspondantes des histogrammes. Pour ce faire, il est possible d'utiliser différentes métriques/fonctions de distance. Les plus couramment utilisées sont : la distance de Manhattan (L_1), la distance Euclidienne (L_2), l'intersection entre les histogrammes, la distance χ^2 . Par la suite, ces distances seront notées \mathcal{D}_{L_1} , \mathcal{D}_{L_2} , \mathcal{D}_\cap , et \mathcal{D}_{χ^2} . Pour deux histogrammes $H(A)$ et $H(B)$, composés de v barres, ces distances peuvent être définies comme

$$\mathcal{D}_{L_1}(H(A), H(B)) = \sum_{i=0}^{v-1} |H_i(A) - H_i(B)| \quad (10.6)$$

$$\mathcal{D}_{L_2}(H(A), H(B)) = \sqrt{\sum_{i=0}^{v-1} |H_i(A) - H_i(B)|^2} \quad (10.7)$$

$$\mathcal{D}_\cap(H(A), H(B)) = \sum_{i=0}^{v-1} \min(H_i(A), H_i(B)) \quad (10.8)$$

$$\mathcal{D}_{\chi^2}(H(A), H(B)) = \sum_{i=0}^{v-1} \frac{|H_i(A) - H_i(B)|^2}{2(H_i(A) + H_i(B))} \quad (10.9)$$

Ces dernières présentent des propriétés spécifiques discutées dans [Cha, 2008].

Comme ces mesures de similarité ne comparent que les barres correspondantes, et qu'elles ignorent les corrélations entre les barres adjacentes, les distances barre-à-barre sont rapides à calculer et peuvent être utilisées pour mesurer des similarités dans de grands jeux de données. De plus, comme elles ne requièrent pas un ordre parmi les barres, elles peuvent être utilisées pour comparer des histogrammes nominaux ou ordinaux. Cependant, elle souffrent du problème de translation : une faible translation des valeurs de l'histogramme peut significativement affecter la distance entre histogrammes. De plus, les distances barre-à-barre sont fortement liées au choix de la « taille » des barres des histogrammes : des barres trop étendues n'offriront pas une capacité de discrimination suffisante tandis que des barres trop fines sépareront des caractéristiques similaires/corrélées en différentes barres qui ne seront jamais comparées. Les distances barres-croisées permettent de prendre en compte ces phénomènes.

B. Distances barres-croisées

De nombreuses distances barres-croisées ont déjà été proposées pour la comparaison d'histogrammes. Elles permettent d'obtenir des résultats de similarité plus pertinents que les distances barre-à-barre. Ces distances peuvent être divisées en deux famille principales : celles nécessitant une matrice de (dis)similarité pour modéliser les relations de proximités entre les barres des histogrammes, et celles nécessitant un ordre total parmi les barres.

Distances basées sur une matrice Parmi les distances basées sur une matrice, celles de formes quadratiques [Niblack et al., 1993; Hafner et al., 1995] utilisent une matrice de similarité \mathcal{M}^{sim} (qui est « l'opposée » de \mathcal{M}^{dis}) pour modéliser les relations de proximité entre les barres. Le calcul de la distance entre histogrammes est généralement basé sur un produit matriciel

$$\mathcal{D}_{quad}(H(A), H(B)) = [(H(A) - H(B))^T \mathcal{M}^{sim} (H(A) - H(B))]^{\frac{1}{2}} \quad (10.10)$$

Ces approches fournissent une première solution pour considérer les relations de proximité entre les barres de ces histogrammes. Cependant, les algorithmes liés au calcul de ces dernières nécessitent des temps de calcul importants. De plus, il a été montré dans [Stricker et Orengo, 1995] que l'utilisation de distances quadratiques pour la recherche d'images basée sur leurs contenus conduisait à des résultats peu pertinents, dans lesquels la similarité mutuelle de la distribution des couleurs est surestimée.

Les distances par correspondances (*matching distances*) représentent une autre forme de distances barres-croisées. Le principe de ces approches est d'estimer le coût de la transformation permettant de passer d'un

histogramme à un autre. Parmi celles-ci, la distance EMD (Earth Mover's Distance)² est reconnue comme l'une des plus efficaces. Elle a été largement étudiée au cours des dix dernières années [Rubner et al., 2000]. Deux notions sont à considérer pour le calcul de cette dernière : la « terre » et les « trous ». Le problème de la mesure de la distance est alors transformé en un problème de déplacement de terre, où le coût minimum du déplacement de toute la terre dans tous les trous est calculé. Comme pour les distances quadratiques, la distance EMD requiert une matrice de similarité \mathcal{M}^{sim} pour modéliser les relations de similarité entre les barres des histogrammes. Cette distance permet de prendre en compte les corrélations entre les barres et de réduire la sensibilité de la distance au choix de la taille de ces dernières. Par ailleurs, il a été montré dans [Rubner et al., 2001] que l'utilisation de la distance EMD dans des tâches de reconnaissance d'images fournissait des résultats dont la qualité surpasse celle obtenue avec la plupart des autres distances barres-croisées. Cependant, le principal inconvénient de EMD repose sur son coût de calcul élevé. En effet, la résolution du problème de programmation linéaire associé au calcul de la distance engendre une complexité algorithmique considérable. Bien que différentes méthodes aient été proposées pour réduire cette complexité [Ling et Okada, 2007], l'utilisation de EMD pour la comparaison de grands jeux de données reste difficilement exploitable.

Distances basées sur un ordre Quand un ordre total est disponible sur les instances du concept considéré, la distance EMD peut être simplifiée [Cha et Srihari, 2002; Serratos et Sanfeliu, 2006] comme suit

$$\mathcal{D}_{EMD}(H(A), H(B)) = \sum_{i=0}^{v-1} \left| \sum_{j=0}^i (H_j(A) - H_j(B)) \right| \quad (10.11)$$

Cette définition est alors équivalente à celle proposée dans [Shen et Wong, 1983] où la distance par correspondance entre deux histogrammes est définie comme la distance L_1 entre les histogrammes cumulatifs correspondants

$$\mathcal{D}_{match}(H(A), H(B)) = \sum_{i=0}^{v-1} |\widehat{H}_i(A) - \widehat{H}_i(B)| \quad (10.12)$$

où $\widehat{H}_i(\cdot) = \sum_{j \leq i} H_j(\cdot)$. Cette distance est efficace pour les histogrammes monodimensionnels mais il reste cependant difficile de l'utiliser pour la comparaison d'histogrammes de plus grandes dimensions. En effet la mise en correspondance dans des espaces de grandes dimensions nécessite la résolution de problèmes complexes de correspondance entre graphes [Werman et al., 1985].

Les mesures de similarité entre séries temporelles peuvent aussi être vues comme des cas particuliers des distances barres-croisées basées sur des ordres. Par exemple, dans [Strelkov, 2008], la similarité est relative aux distances entre les positions spatiales et les formes des pics des histogrammes comparés. Le principal inconvénient de ces mesures est de considérer principalement les pics des histogrammes, ce qui n'est généralement pas suffisant pour comparer des structures de données dans lesquelles les barres vides portent aussi une information. Une autre solution consiste à utiliser la mesure de similarité Dynamic Time Warping (DTW) [Rath et Manmatha, 2003; Petitjean et al., 2011]. Cette mesure de similarité temporelle autorise de faibles distorsions pendant l'étape de mise en correspondance entre les histogrammes considérés. Cependant, son utilisation nécessite de fixer un paramètre de distorsion. Si celui-ci est fixé à des valeurs faibles, cette mesure est équivalente à la distance barre-à-barre de Manhattan. Dans le cas contraire, cette mesure peut supporter de grandes distorsions entre les histogrammes (induisant un coût de calcul élevé). Cependant, dans le cadre de ces travaux, les histogrammes considérés sont des histogrammes nominaux pour lesquels aucun ordre « naturel » n'est disponible. Bien qu'il soit possible de déterminer aléatoirement un ordre pour ordonner les barres de ces structures, l'emploi de distances basées sur un ordre ne conduirait pas à des résultats pertinents.

En conclusion, les distances barre-à-barre permettent de comparer des histogrammes avec un faible coût de calcul mais sans prendre en compte les corrélations sémantiques possibles entre les barres. Par ailleurs, les distances barres-croisées, et en particulier les plus générales, basées sur les matrices de similarité, permettent la prise en compte de ces corrélations mais nécessitent un coût de calcul beaucoup plus important.

	<i>Agrume</i>	<i>Coing</i>	<i>Pomme</i>	<i>Abricot</i>	<i>Poire</i>	<i>Pêche</i>	<i>Cerise</i>	<i>Prune</i>
H'_1	10	0	0	0	0	0	0	0
H'_2	2	0	0	0	0	0	0	8
H'_3	9	0	0	0	0	0	0	1

TABLE 10.3 – Histogrammes modélisant les compositions des trois paniers de courses C_1, C_2, C_3 (voir le tableau 10.1) après la création de l’instance Agrume.

10.2.3 Histogrammes et hiérarchies

De nombreux travaux ont montré qu’organiser des informations sémantiques portées par des données, sous la forme de hiérarchies, pouvait faciliter certaines tâches d’extraction de connaissances [Tousch et al., 2012]. Il existe en effet une grande variété d’approches d’analyse d’images utilisant des histogrammes et qui se basent sur ce paradigme. Dans ces approches, la hiérarchie d’histogrammes est généralement composée d’histogrammes construits sur une hiérarchie de données, comparés avec des distances standards. Cependant, les approches hiérarchiques utilisées spécifiquement pour le calcul d’une distance entre histogrammes semblent fournir une solution pertinente pour traiter le problème évoqué dans la section 10.1.2. Pour illustrer cette assertion, revenons à l’exemple du panier de courses introduit précédemment.

On considère à nouveau trois paniers de courses C_1, C_2, C_3 (chacun composé de 10 fruits, voir la table 10.1). On rappelle que la composition d’un panier de courses C_i peut être modélisée par un histogramme $H_i(\#\text{Citron}, \#\text{Coing}, \dots, \#\text{Prune})$. Supposons maintenant que l’on fusionne les instances Citron, Orange et Mandarine pour créer une nouvelle instance nommée Agrume. La composition d’un panier de courses C_i est dorénavant modélisée par un histogramme $H'_i(\#\text{Agrume}, \#\text{Coing}, \dots, \#\text{Prune})$ où $\#\text{Agrume} = \#\text{Citron} + \#\text{Mandarine} + \#\text{Orange}$. Les compositions résultantes des trois paniers de courses C_1, C_2, C_3 sont présentées dans la table 10.3. La distance barre-à-barre de Manhattan d_{L_1} est maintenant plus grande entre H'_1 et H'_2 ($d_{L_1}(H'_1, H'_2) = 16$) qu’entre H'_1 et H'_3 ($d_{L_1}(H'_1, H'_3) = 2$). Cette valeur de distance reflète de manière plus pertinente les similarités sémantiques entre les paniers C_1 et C_3 qui sont tous les deux des paniers de courses principalement composés d’agrumes.

Comme illustré dans cet exemple, les distances basées sur des relations hiérarchiques permettent naturellement de considérer les corrélations sémantiques multiniveaux entre les distributions modélisées par ces histogrammes. À notre connaissance, la seule distance entre histogrammes impliquant une telle stratégie hiérarchique a été proposée dans [Ma et al., 2010]. Son calcul est basé sur une fusion itérative des barres des histogrammes les plus proches pour créer des histogrammes plus grossiers. Étant donné que la valeur de cette mesure de distance est obtenue en calculant itérativement une distance barre-à-barre choisie, son coût algorithmique est plus bas que celui induit par le calcul des distances barres-croisées. Cette distance, qui a été proposée pour des tâches de recherche d’images basée sur le contenu, a fourni des résultats encourageants. Cependant, comme son processus de fusion requiert un ordre total sur les barres, elle peut uniquement être utilisée pour la comparaison d’histogrammes ordinaux. Dans le cadre de ces travaux, où nous cherchons à comparer des histogrammes nominaux associés à une matrice de dissimilarité, cette distance n’est donc pas applicable.

Dans la prochaine section, nous proposons une solution à ce problème en définissant une distance hiérarchique pouvant être employée pour la comparaison de tels histogrammes. En particulier, cette nouvelle distance, nommée HSBD (*Hierarchical Semantic-Based Distance*), combine l’efficacité des distances barre-à-barre (*i.e.*, faible coût algorithmique) et les avantages offerts par les distances barres-croisées (*i.e.*, robustesse aux problèmes liés aux translations des valeurs des histogrammes et au choix de la taille des barres).

2. La distance EMD est aussi appelée la distance de Mallows [Mallows, 1972] dans le domaine des statistiques [Levina et Bickel, 2001].

10.3 Schéma méthodologique de HSBD

10.3.1 Principe général

Le calcul de la distance HSBD, entre deux histogrammes $H(A)$ et $H(B)$ composés de v barres, requiert deux paramètres :

1. une matrice de dissimilarité \mathcal{M}^{dis} modélisant les valeurs de proximité sémantique entre les v instances composant $H(A)$ et $H(B)$;
2. une distance barre-à-barre entre histogrammes \mathcal{D}_{bin} .

Afin de pouvoir calculer la distance HSBD entre $H(A)$ et $H(B)$, la stratégie adoptée (basée sur un modèle de calcul fin-à-grossier) requiert de définir un moyen de fusionner hiérarchiquement les différentes instances représentées par les histogrammes en « clusters » d'instances (*i.e.*, des instances de niveaux sémantiques plus élevés). Cette étape de pré-traitement, décrite dans la section 10.3.2, repose principalement sur la construction d'un dendrogramme \mathcal{D} induit par \mathcal{M}^{dis} , et modélisant la hiérarchie de fusion des instances. Il est à noter que cette étape de pré-traitement ne doit être effectuée qu'une seule fois pour une matrice de dissimilarité donnée.

Une fois que le dendrogramme \mathcal{D} a été construit, la distance HSBD peut être calculée entre $H(A)$ et $H(B)$. Son calcul se décompose en deux étapes (détaillées en section 10.3.3) :

1. **Étape 1. Calcul des sous-distances barre-à-barre hiérarchiques** Durant un processus itératif de fusion scannant chaque étage du dendrogramme (de ses feuilles jusqu'à sa racine), les histogrammes $H^k(A)$ et $H^k(B)$ associés à $H(A)$ et $H(B)$, sont construits. Ces différents histogrammes sont induits par la fusion des instances composant chaque cluster de l'étage courant S_k . Après chaque itération, la distance barre-à-barre D_{bin} est calculée entre chaque couple d'histogrammes grossiers $H^k(A)$ et $H^k(B)$.
2. **Étape 2. Fusion des sous-distances barre-à-barre** Les distances barre-à-barre calculées pour tous les étages du dendrogramme, et l'énergie sémantique nécessaire pour aller d'un étage à l'autre, sont ensuite fusionnées en une fonction qui est finalement intégrée pour fournir la valeur de la distance HSBD.

Ces deux étapes sont détaillées dans la section 10.3.3. Le lecteur pourra aussi se référer à la figure 10.3 pour une illustration du calcul de HSBD.

10.3.2 Construction de la hiérarchie de fusion sémantique

Le principe de l'approche proposée est de calculer plusieurs fois une distance barre-à-barre entre des paires d'histogrammes, en fusionnant progressivement les barres/instances les plus proches pour créer des histogrammes plus grossiers (*i.e.*, de plus hauts niveaux sémantiques). Pour ce faire, il est nécessaire de définir une hiérarchie de fusion des instances qui permet de déterminer l'ordre des fusions entre les instances d'un concept sémantique.

En partant des valeurs de proximité sémantique contenues dans \mathcal{M}^{dis} , il est possible de calculer une hiérarchie de fusion des instances en utilisant par exemple un algorithme de classification hiérarchique ascendante (AHC) [Ward, 1963]. Les quatre étapes de cet algorithme sont énoncées ci-dessous :

1. **Étape 1.** Chaque groupe est initialement construit avec uniquement une instance basique (*i.e.*, v groupes où v est le nombre d'instances considérées).
2. **Étape 2.** Les valeurs de dissimilarité entre chaque paire de groupes sont ensuite calculées, et la matrice \mathcal{M}^{dis} est mise à jour.
3. **Étape 3.** Les deux groupes les plus proches (*i.e.*, les groupes qui ont la plus faible valeur de dissimilarité relativement à \mathcal{M}^{dis}) sont ensuite fusionnés, et la matrice \mathcal{M}^{dis} est modifiée pour prendre en compte la création de ce nouveau groupe (en fusionnant les lignes/colonnes associées aux deux groupes fusionnés).
4. **Étape 4.** Si le nombre de groupes restants est toujours supérieur à 1, on retourne à l'étape 2.

Cet algorithme construit hiérarchiquement des clusters d'instances minimisant une mesure d'inertie intra-cluster³. Pour calculer les valeurs de dissimilarité entre chaque couple de groupes, il est nécessaire de choisir

³ Des travaux récents [Soille et Najman, 2012] se sont intéressés aux relations existantes entre les approches de classification hiérarchique et les approches morphologiques de segmentation reposant sur la construction de hiérarchies de segments.

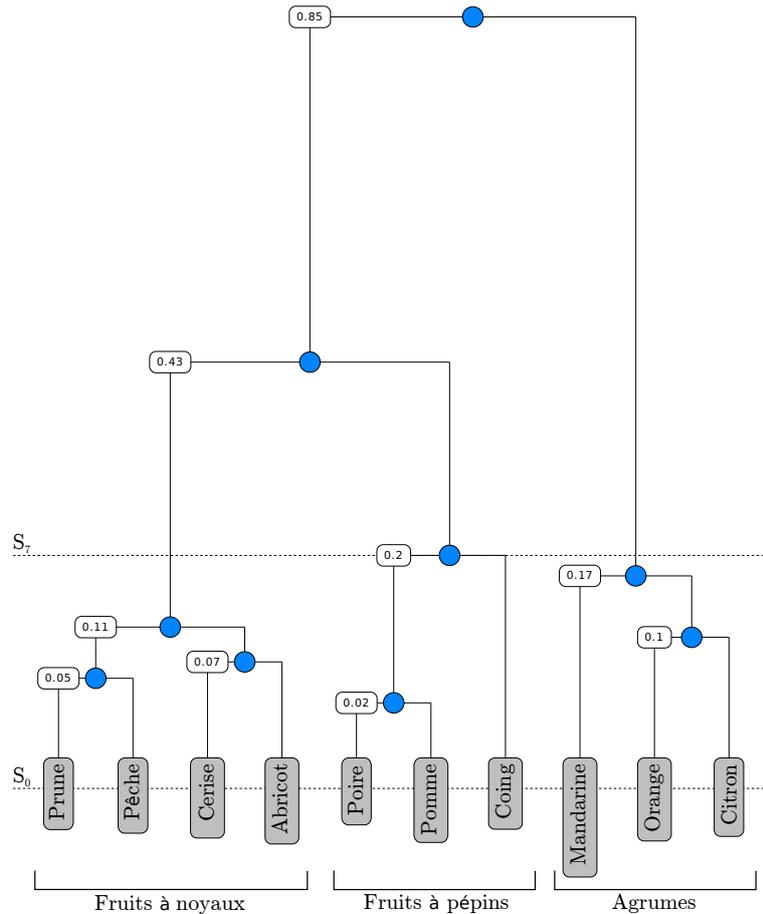


FIGURE 10.2 – Dendrogramme associé à la matrice de dissimilarité présentée dans la table 10.2. Les instances initiales sont représentées par des rectangles gris tandis que les instances de plus hauts niveaux sémantiques sont représentées par des disques bleus.

un critère de fusion. Nous détaillons, dans la partie de validations, des expériences relatives au choix de ce critère.

Cette hiérarchie de fusion est modélisée par un dendrogramme \mathcal{D} de s étages⁴, dont la racine est le cluster qui regroupe toutes les instances. Chaque étage de \mathcal{D} correspond à un niveau de sémantique particulier. La valeur minimale de s ($s_{min} = 2$) est atteinte quand \mathcal{M}^{dis} est une matrice où $\alpha(x_i, x_j) = 1$ si $x_i \neq x_j$ et $\alpha(x_i, x_j) = 0$ sinon (*i.e.*, aucune connaissance du domaine). Dans ce cas particulier, le dendrogramme présente un étage pour les feuilles qui sont les instances de base et un étage pour la racine. À l'inverse, la valeur maximale de s ($s_{max} = v$) est atteinte quand \mathcal{D} est un dendrogramme totalement déséquilibré.

Nous définissons, à partir de ce dendrogramme :

- une fonction $f_{\mathcal{D}}$, qui prend en entrée l'indice k de l'étage S_k ($k \in \llbracket 0, s-1 \rrbracket$) et fournit en sortie la liste $\mathcal{L}_m^k = \langle \mathcal{L}_{v_0}, \dots, \mathcal{L}_{v_{m-1}} \rangle$ composée des m listes de fusion d'instances \mathcal{L}_{v_i} ($i \in \llbracket 0, m-1 \rrbracket$) induites par \mathcal{D} à cet étage (*i.e.*, m clusters) ;
- une fonction $h_{\mathcal{D}}$, qui prend en entrée l'indice k de l'étage S_k et fournit en sortie sa hauteur $h_{\mathcal{D}}(k)$ dans le dendrogramme \mathcal{D} .

La hauteur $h_{\mathcal{D}}(k)$ correspond à « l'énergie » nécessaire pour construire les clusters d'instances induits par l'étage S_k (*i.e.*, l'inertie inter-cluster calculée par AHC pendant la création de \mathcal{D}).

La figure 10.2 illustre le dendrogramme associé à la matrice de dissimilarité définie dans la table 10.2. Dans cet exemple, la hauteur de l'étage S_7 est donnée par $h_{\mathcal{D}}(7) = 0.2$. La liste \mathcal{L}_3^7 composée de trois listes de fusion

4. Ces étages sont généralement indexés par leur profondeur dans le dendrogramme (*i.e.*, de la racine aux feuilles). Dans notre cas, et pour des raisons de lisibilité, nous indexons ces étages par leur niveau dans la hiérarchie (*i.e.*, des feuilles à la racine).

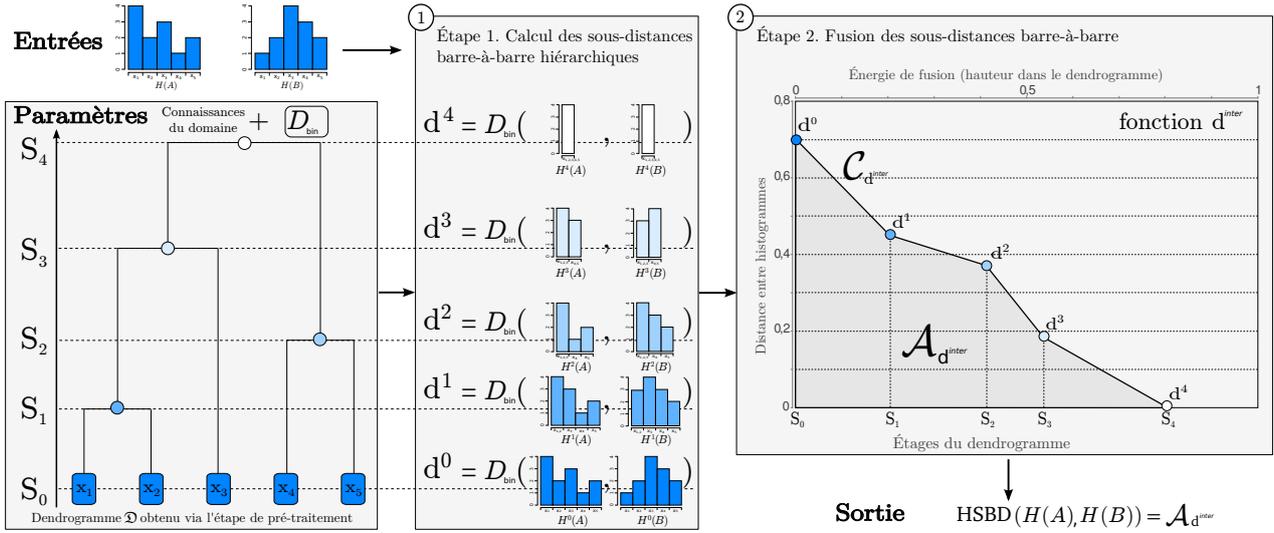


FIGURE 10.3 – Schéma méthodologique du calcul de la distance HSBD.

d'instances à l'étage S_7 est donnée par $f_{\mathcal{D}}(7) = \langle \langle \text{Prune, Pêche, Cerise, Abricot} \rangle, \langle \text{Poire, Pomme, Coing} \rangle, \langle \text{Mandarine, Orange, Citron} \rangle \rangle$. Il est à noter que ces trois listes de fusion d'instances correspondent à trois classes d'un niveau sémantique plus élevé : les fruits à pépins, les fruits à noyaux et les fruits de la famille des agrumes.

10.3.3 Calcul de HSBD

Une fois l'étape de pré-traitement réalisée, nous disposons d'un dendrogramme \mathcal{D} qui modélise la hiérarchie de fusion des instances. Il est maintenant possible de calculer la distance HSBD entre les deux histogrammes $H(A)$ et $H(B)$ composés de v barres. Pour une illustration du calcul de HSBD, le lecteur pourra se référer à la figure 10.3.

A. Étape 1. Calcul des sous-distances barre-à-barre hiérarchiques

Pour calculer les distances barre-à-barre hiérarchiques durant le processus itératif de fusion, nous définissons la fonction \mathbf{d}^k . Cette fonction prend en entrée deux histogrammes $H(A)$ et $H(B)$ et l'indice k de l'étage S_k du dendrogramme, et fournit en sortie la distance barre-à-barre D_{bin} calculée entre les histogrammes grossiers $H^k(A)$ et $H^k(B)$ associés à $H(A)$ et $H(B)$ à l'étage S_k de \mathcal{D} . Cette fonction \mathbf{d}^k peut être définie comme

$$\mathbf{d}^k(H(A), H(B)) = \mathcal{D}_{bin}(H^k(A), H^k(B)) \quad (10.13)$$

où $H^k(A)$ et $H^k(B)$ sont des histogrammes grossiers (fournissant un plus haut niveau de sémantique) induits par le regroupement des instances considérées à l'étage S_k . De tels histogrammes peuvent être construits en utilisant la fonction $f_{\mathcal{D}}(k)$ qui fournit une liste $\mathcal{L}_m^k = \langle \mathcal{L}_{v_0}, \dots, \mathcal{L}_{v_{m-1}} \rangle$ composée des m listes de fusion d'instances induits par l'étage S_k . Un histogramme $H^k(Y)$ pourra être défini comme

$$H^k(Y) = \langle H_0^k(Y), H_1^k(Y), \dots, H_{m-1}^k(Y) \rangle \quad (10.14)$$

où chaque barre $H_i^k(Y)$ peut être calculée par

$$H_i^k(Y) = \sum_{j \in \mathcal{L}_{v_i}} H_j(Y) \quad (10.15)$$

Pour être plus concis, la fonction $\mathbf{d}^k(H(A), H(B))$ pourra être notée \mathbf{d}^k . Par la suite, les valeurs induites par cette fonction seront appelées des valeurs de *sous-distance*.

Le processus de fusion itératif fin-à-grossier fonctionne de la façon suivante : dans un premier temps, la sous-distance barre-à-barre \mathbf{d}^0 est calculée pour $H(A)$ et $H(B)$ en considérant toutes les barres des histogrammes (*i.e.*, $f_{\mathcal{D}}(0) = \langle \langle x_0 \rangle, \langle x_1 \rangle, \dots, \langle x_{v-1} \rangle \rangle$ et $h_{\mathcal{D}}(0) = 0$). Ensuite, en considérant l'étage suivant S_i du dendrogramme, les barres les plus proches (sémantiquement) des histogrammes (données par $f_{\mathcal{D}}(i)$) sont fusionnées, et une nouvelle valeur de sous-distance \mathbf{d}^i est calculée entre les deux histogrammes résultants $H^i(A)$ et $H^i(B)$. Cette nouvelle valeur de sous-distance permet d'évaluer la similarité à un niveau particulier de « taille de barres » (et ainsi de sémantique). Cette étape est répétée pour chaque étage $S_k, k \in \llbracket 1, s-1 \rrbracket$ jusqu'à ce que le nombre de barres soit égal à 1 (*i.e.*, le processus se termine quand la racine de l'arbre est atteinte). Une série de sous-distances hiérarchiques $\mathbf{d}^0, \mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^{s-1}$ est ainsi obtenue. Il est à noter que \mathbf{d}^{s-1} est toujours égale à 0 car les histogrammes $H^{s-1}(A)$ et $H^{s-1}(B)$ sont toujours composés d'une seule barre représentant l'instance du plus haut niveau sémantique du dendrogramme (*i.e.*, la racine).

B. Étape 2. Fusion des sous-distances barre-à-barre

Une fois que les sous-distances hiérarchiques $\mathbf{d}^0, \mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^{s-1}$ ont été calculées, il est possible de les fusionner pour obtenir une mesure globale de distance entre les deux histogrammes $H(A)$ et $H(B)$.

Pour justifier la stratégie employée, considérons la représentation graphique de la fonction \mathbf{d}^k illustrée en figure 10.4. Ces deux courbes représentent la fonction \mathbf{d}^k calculée entre des couples d'histogrammes exemples. On peut observer que plus les histogrammes comparés sont sémantiquement/thématiquement similaires, plus les valeurs de \mathbf{d}^k décroissent rapidement (Figure 10.4 (a)), et réciproquement (Figure 10.4 (b)).

Ce comportement est lié aux propriétés analytiques de la fonction induite par les valeurs \mathbf{d}^k . Plus précisément, considérons la fonction affine par morceaux $\mathbf{d}^{inter} : [h_{\mathcal{D}}(0), h_{\mathcal{D}}(s-1)] \rightarrow \mathbb{R}_+$ définie par $\mathbf{d}^{inter}(h_{\mathcal{D}}(k)) = \mathbf{d}^k$ pour tout $k \in \llbracket 0, s-1 \rrbracket$ (voir la figure 10.4). Le taux de décroissance de \mathbf{d}^{inter} , qui caractérise la similarité entre les histogrammes, est lié à la valeur de l'intégrale $\mathcal{A}_{\mathbf{d}^{inter}}$ de cette fonction (Figure 10.3-②). Cette assertion justifie la définition de HSBD comme

$$\text{HSBD}(H(A), H(B)) = \int_{h_{\mathcal{D}}(0)}^{h_{\mathcal{D}}(s-1)} \mathbf{d}^{inter}(t) \cdot dt \quad (10.16)$$

Pour calculer HSBD nous utilisons la méthode des trapèzes, qui est une approche classique pour calculer une intégrale finie. Ainsi, HSBD peut être calculée de la façon suivante

$$\text{HSBD}(H(A), H(B)) = \frac{1}{2} \sum_{k=0}^{s-2} [(\mathbf{d}^{k+1} + \mathbf{d}^k) (h_{\mathcal{D}}(k+1) - h_{\mathcal{D}}(k))] \quad (10.17)$$

Remarque 1. *Il est possible d'utiliser HSBD avec une matrice de dissimilarité \mathcal{M}^{dis} « partielle ». En effet, dans le cas où une proximité ne peut pas être évaluée entre certaines des instances considérées, les valeurs de dissimilarité associées sont fixées à 1 dans \mathcal{M}^{dis} . Pour ces parties de la matrice, le comportement de HSBD sera équivalent à celui de la distance barre-à-barre \mathcal{D}_{bin} utilisée. En particulier, si aucune connaissance du domaine n'est disponible, \mathcal{M}^{dis} est une matrice où $\alpha_{(x_i, x_j)} = 1$ si $x_i \neq x_j$ et $\alpha_{(x_i, x_j)} = 0$ sinon. Dans de telles conditions, HSBD est équivalente à la sous-distance barre-à-barre utilisée*

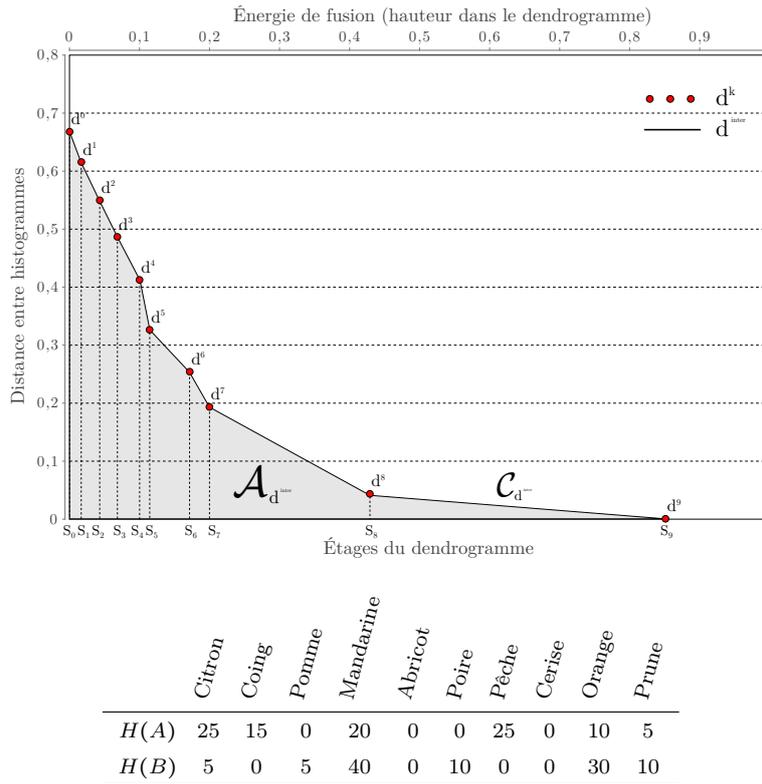
$$\text{HSBD}(H(A), H(B)) = \frac{1}{2} \mathcal{D}_{bin}(H(A), H(B)) \quad (10.18)$$

C. Propriétés

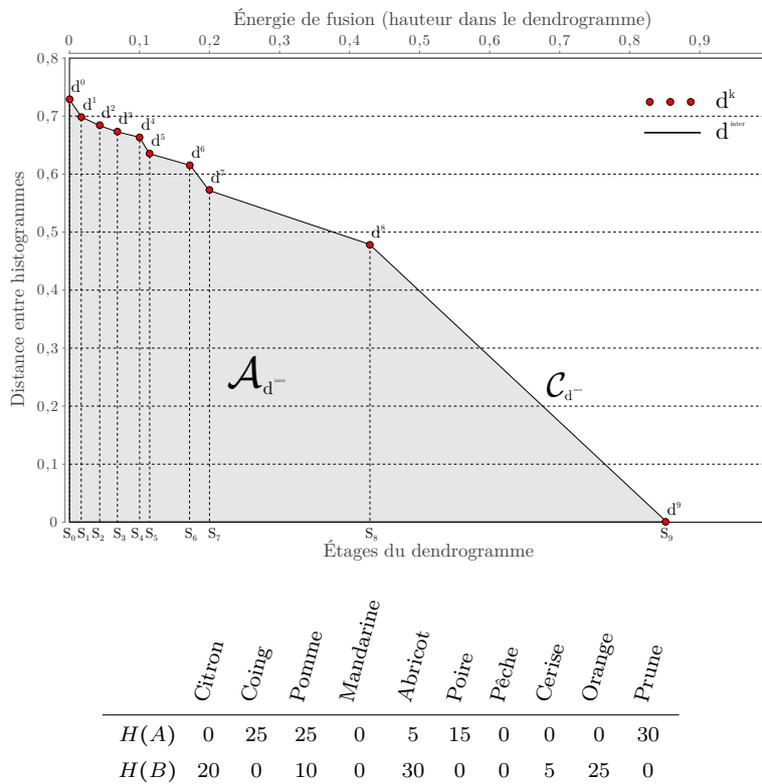
Pour être une distance, une mesure de similarité doit avoir les propriétés suivantes : non-négativité, symétrie, identité et inégalité triangulaire. La mesure HSBD peut être définie comme une somme pondérée de sous-distances \mathbf{d}^k (Équation (10.17)) ce qui est équivalent à une somme de sous-distances barre-à-barre \mathcal{D}_{bin} . Nous montrons ci-après que HSBD hérite des propriétés de \mathcal{D}_{bin} et peut donc être considérée comme une distance.

Propriété 1 (Non-négativité). $\text{HSBD}(H(A), H(B)) \geq 0$.

Démonstration. Le résultat dérive de la propriété de non-négativité de la fonction \mathbf{d}^k et de la décroissance de la fonction $h_{\mathcal{D}}$. \square



(a) Comparaison d'un couple d'histogrammes sémantiquement similaires.



(b) Comparaison d'un couple d'histogrammes sémantiquement différents.

FIGURE 10.4 – Représentation graphique des fonctions \mathbf{d}^k et \mathbf{d}_{inter}^k calculées entre des couples d'histogrammes exemples (modélisant la composition d'un panier de 100 fruits). Suivant le contenu des histogrammes $H(A)$ et $H(B)$, le comportement des fonctions \mathbf{d}^k et \mathbf{d}_{inter}^k diffère.

Propriété 2 (Symétrie). $HSBD(H(A), H(B)) = HSBD(H(B), H(A))$.

Démonstration. Le résultat dérive de la propriété de symétrie de \mathcal{D}_{bin} . □

Propriété 3 (Identité). $HSBD(H(A), H(B)) = 0 \Leftrightarrow H(A) = H(B)$.

Démonstration. La partie « \Leftarrow » dérive de la propriété d'identité de \mathcal{D}_{bin} . La partie « \Rightarrow » est obtenue de la manière suivante

$$\begin{aligned} HSBD(H(A), H(B)) = 0 &\Rightarrow \frac{1}{2} \sum_{k=0}^{s-2} \left[\underbrace{(\mathbf{d}^{k+1} + \mathbf{d}^k)}_{\geq 0} \underbrace{(h_{\mathfrak{D}}(k+1) - h_{\mathfrak{D}}(k))}_{> 0} \right] = 0 \\ &\Rightarrow \sum_{k=0}^{s-2} [\mathcal{D}_{bin}(H^{k+1}(A), H^{k+1}(B)) + \mathcal{D}_{bin}(H^k(A), H^k(B))] = 0 \\ &\Rightarrow \forall k \in [0, s-2], \mathcal{D}_{bin}(H^k(A), H^k(B)) = 0 \\ &\Rightarrow H(A) = H(B) \end{aligned}$$

□

Propriété 4 (Inégalité triangulaire). $HSBD(H(A), H(B)) + HSBD(H(B), H(C)) \geq HSBD(H(A), H(C))$.

Démonstration. Le résultat dérive de la propriété d'inégalité triangulaire de \mathcal{D}_{bin} . □

D. Complexité algorithmique

Nous détaillons ci-après la complexité algorithmique de la distance proposée. Pour calculer HSBD, il est dans un premier temps nécessaire de réaliser une étape de pré-traitement (voir Section 10.3.2), qui consiste à construire l'ordre des fusions (*i.e.*, construire le dendrogramme \mathfrak{D}). Pour ce faire, on utilise l'algorithme AHC [Ward, 1963]. La complexité de l'algorithme AHC naïf est $\Theta(v^3)$ où v représente le nombre d'instances basiques. Il est en effet nécessaire de parcourir exhaustivement la matrice \mathcal{M}^{dis} de taille $v \times v$ pour trouver la valeur de dissimilarité la plus faible à chacune des $v - 1$ itérations (voir Section 10.3.2, étape 4). Cependant, cette complexité peut être réduite à $\Theta(v^2 \log v)$ en utilisant un algorithme à base de files d'attente.

Une fois que le dendrogramme a été construit, il devient possible de calculer HSBD. Ce calcul nécessite le pré-calcul de la série de sous-distances $\mathbf{d}^0, \dots, \mathbf{d}^{s-1}$ (voir l'équation (10.17)) où s représente le nombre d'étages de \mathfrak{D} . La complexité du calcul de chaque \mathbf{d}^k est directement liée à celle de \mathcal{D}_{bin} , qui équivaut en général à $\Theta(v)$ où v représente le nombre de barres dans les histogrammes (voir les équations (10.6–10.9)). Ainsi, la complexité de HSBD dépend à la fois du nombre de barres v dans les histogrammes et du nombre d'étages s dans le dendrogramme \mathfrak{D} .

Suivant la valeur de s (qui est corrélée à la forme de \mathfrak{D}), différents cas peuvent être considérés :

- si \mathfrak{D} est un dendrogramme « plat » (*i.e.*, un dendrogramme à deux étages), alors $s = 2$ et la complexité de HSBD est égale à

$$\Theta(v + 1) = \Theta(v) \tag{10.19}$$

- si \mathfrak{D} est un dendrogramme totalement équilibré, alors $s = \log_2(v)$ et la complexité de HSBD est égale à

$$\Theta\left(\underbrace{v + \frac{v}{2} + \frac{v}{4} + \dots + 1}_{\log_2(v) \text{ termes}}\right) = \Theta(2v) = \Theta(v) \tag{10.20}$$

- si \mathfrak{D} est un dendrogramme quelconque, alors $s = v$ et la complexité de HSBD est égale à

$$\Theta\left(\underbrace{v + (v-1) + \dots + 1}_v\right) = \Theta(v^2) \tag{10.21}$$

Ainsi, le coût algorithmique de HSBD est borné par $\Theta(v)$ et $\Theta(v^2)$. À cette complexité, il est nécessaire d'ajouter celle nécessaire pour la construction du dendrogramme \mathcal{D} , qui requiert $\Theta(v^2 \log v)$ opérations. Cependant, comme cette opération n'est réalisée qu'une seule fois, sa complexité peut être considérée comme négligeable lors de la comparaison de grands jeux de données.

Pour conclure sur cette étude, le calcul de HSBD nécessite plus d'opérations que celui requis pour une distance barre-à-barre classique, mais largement moins d'opérations que pour les distances barres-croisées, pouvant nécessiter des temps super-cubiques. Ainsi, HSBD peut être utilisée efficacement pour comparer des grands jeux de données caractérisées par des histogrammes nominaux.

10.4 Bilan scientifique

10.4.1 Contributions

Nous avons présenté dans ce chapitre une distance dédiée à la comparaison d'histogrammes nominaux. La principale originalité de cette mesure est de prendre en compte les relations de proximité/corrélation sémantique entre les barres des histogrammes considérés. Pour ce faire, nous avons proposé une stratégie reposant sur les connaissances du domaine de l'utilisateur. Cette distance est basée sur un modèle de calcul fin-à-grossier apportant une solution aux problèmes liés à la translation des valeurs de l'histogramme et au choix de la taille des barres et garantissant des faibles temps de calcul rendant possible le traitement de grands jeux de données.

Cette distance entre histogrammes constitue la principale contribution de ce chapitre.

10.4.2 Valorisation scientifique

Ces travaux ont donné lieu à une contribution [Kurtz, 2012] dans les actes de la conférence francophone *EGC 2012, Extraction et Gestion des Connaissances* publiés aux éditions *Hermann*. De plus, un article est actuellement en cours de relecture dans une revue internationale.

C-PCSC, un processus de post-classification guidé par des connaissances

Sommaire

11.1 Remise en cause des résultats de classification	137
11.1.1 Ajout de connaissances dans l'approche de classification multirésolution	137
11.1.2 Post-classification des régions	138
11.2 Schéma méthodologique de C-PCSC	141
11.2.1 Entrées et sorties	141
11.2.2 Schéma méthodologique	141
11.2.3 Correction des résultats de classification	141
11.3 Vers un enrichissement itératif des résultats	143
11.3.1 Enrichissement de la segmentation par la classification	143
11.3.2 Vers une approche ascendante itérative	143
11.3.3 Vers des ontologies multiniveaux et multirésolutions	144
11.4 Bilan scientifique	145

Nous introduisons dans ce chapitre une première proposition d'un post-processus ascendant, permettant une remise en cause des résultats de segmentation et de classification obtenus par le biais de PCSC. L'intérêt majeur de ce processus est de prendre en compte d'une part, des connaissances relatives au contenu des images traitées (générées pendant le traitement initial des données par le biais de PCSC), et d'autre part de pouvoir intégrer des connaissances de haut niveau (fournies par l'utilisateur) dans le but de corriger les différents résultats d'analyse. Pour ce faire, nous commençons par présenter un mécanisme permettant la remise en cause d'un résultat de classification à une résolution particulière (Section 11.1). Nous décrivons ensuite le schéma méthodologique du post-processus ascendant proposé, nommé C-PCSC, basé sur le mécanisme de remise en cause présenté précédemment, permettant de corriger les résultats de classification obtenus par le biais de PCSC (Section 11.2). Finalement, nous introduisons des perspectives relatives à la définition d'un post-processus itératif permettant de corriger les résultats de segmentation et de classification à travers la résolution (Section 11.3). Dans une dernière section (Section 11.4), nous énonçons le bilan scientifique de ces travaux.

11.1 Remise en cause des résultats de classification

11.1.1 Ajout de connaissances dans l'approche de classification multirésolution

Le processus PCSC, permet d'extraire et de classifier des hiérarchies de régions de manière descendante à partir d'images multirésolutions. Cependant, nous avons vu précédemment (Chapitre 9) que les différentes étapes de ce processus engendrent certains problèmes. En particulier, l'approche de classification multirésolution est limitée par sa nature non-supervisée. L'une des principales limites est la non prise en compte des connaissances extraites du contenu des images, ainsi que des connaissances du domaine, lors de la comparaison des histogrammes de composition.

Pour illustrer ce phénomène, considérons deux régions S_i et S_j , extraites par PCSC à une résolution \mathbf{r}_t , correspondant dans l'image à des blocs urbains d'habitations collectives. Ces régions ont été caractérisées (au cours de la descente) par deux histogrammes modélisant leur composition en termes des labels des clusters extraits d'une image à une résolution \mathbf{r}_{t+1} (par une classification basée pixels). On fait l'hypothèse que $H(S_i) = (80\% - \text{Cluster 1}, 10\% - \text{Cluster 2}, 10\% - \text{Cluster 3})$ et que $H(S_j) = (10\% - \text{Cluster 1}, 80\% - \text{Cluster 2}, 10\% - \text{Cluster 3})$, et que l'algorithme de classification a classé ces deux régions dans deux clusters différents. Cependant, ces deux régions représentent le même type d'objets géographiques : la classification a donc commis une erreur à l'étape t . À la fin de l'étape $t + 1$, une classification basée régions de l'image est générée par PCSC. On fait aussi l'hypothèse que l'on peut affecter une sémantique aux labels des nouveaux clusters issus de cette classification (Cluster 1' \rightarrow toits d'habitations individuelles en tuiles, Cluster 2' \rightarrow toits d'habitations individuelles en ardoise, Cluster 3' \rightarrow routes). Cette dernière apporte alors une information de plus haut niveau sur le contenu de l'image à la résolution \mathbf{r}_{t+1} . Par exemple, on peut obtenir comme information que la région S_i est composée, à la résolution \mathbf{r}_{t+1} , de (80% de toits d'habitations individuelles en tuiles, 10% de toits d'habitations individuelles en ardoise, 10% de routes) et que la région S_j est composée à la résolution \mathbf{r}_{t+1} de (10% de toits d'habitations individuelles en tuiles, 80% de toits d'habitations individuelles en ardoise, 10% de routes). Ces deux régions sont donc toutes les deux composées, en majorité, de toits d'habitations individuelles et de routes. En conséquence, il semble pertinent de prendre en compte de telles connaissances lors de la classification de ces régions à l'étape t .

Par ailleurs, nous avons proposé une distance pour la comparaison d'histogrammes, permettant d'utiliser des connaissances du domaine pour prendre en compte les corrélations sémantiques (potentielles) affectant les barres de ces histogrammes. Dans l'optique de pouvoir corriger les résultats de classification obtenus pour les différentes résolutions, nous proposons par la suite un mécanisme de remise en cause des résultats de classification. Ce dernier est principalement basé sur une post-classification des régions obtenues par l'intermédiaire de PCSC *via* l'approche de classification multirésolution proposée dans la partie précédente, combinée à la distance HSBD.

11.1.2 Post-classification des régions

A. Histogrammes de composition et sémantique

Pour prendre en compte au sein de ce processus, les connaissances extraites préalablement des images, nous définissons par la suite des histogrammes de composition liés à une sémantique. De telles structures permettront de caractériser des segments extraits d'une image à une résolution \mathbf{r}_t , en considérant des informations relatives aux régions classifiées, obtenues par l'analyse préalable de la résolution \mathbf{r}_{t+1} . La construction de tels histogrammes n'est possible qu'après un premier traitement du jeu de données par l'intermédiaire de PCSC car on doit disposer d'un résultat de classification basé régions à la résolution \mathbf{r}_{t+1} , lors du traitement de la résolution \mathbf{r}_t .

Pour ce faire, on considère deux images $\mathcal{I}^t : E^t \rightarrow V^t$ et $\mathcal{I}^{t+1} : E^{t+1} \rightarrow V^{t+1}$ avec $\mathbf{r}_t \leq \mathbf{r}_{t+1}$, ainsi qu'une fonction de correspondance $\lambda_{t \rightarrow t+1} : E^t \rightarrow E^{t+1}$ permettant de mettre en correspondance les pixels de E^t avec ceux de E^{t+1} . On considère également une image classifiée $\mathcal{I}_C^{t+1} : E^{t+1} \rightarrow [[1, k]]$ à la résolution \mathbf{r}_{t+1} , obtenue par une classification basée régions. On suppose alors que l'on peut affecter, aux étiquettes des clusters issus du résultat de classification obtenu à la résolution \mathbf{r}_{t+1} , une sémantique correspondant aux classes d'intérêt recherchées (*e.g.*, Cluster 1 \rightarrow Toits d'habitations rouges, ..., Cluster 7 \rightarrow Parcelles de végétation).

Il devient alors possible de construire un histogramme de composition $H(\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X))$ pour modéliser la distribution des pixels d'un ensemble $X \subseteq E^t$ en fonction des labels des classes composant l'image classifiée basée régions \mathcal{I}_C^{t+1} . Un tel histogramme peut être construit par le biais de la fonction suivante

$$\mathcal{H}_{\mathcal{I}_C^{t+1}, \lambda_{t \rightarrow t+1}(X)} : \begin{array}{ccc} [[1, k]] & \rightarrow & \mathbb{N} \\ i & \mapsto & |\cup_{\mathbf{x} \in X} \lambda_{t \rightarrow t+1}(\mathbf{x}) \cap (\mathcal{I}_C^{t+1})^{-1}(\{i\})| \end{array} \quad (11.1)$$

Cet histogramme de composition permet d'associer à chaque label i de l'image classifiée \mathcal{I}_C^{t+1} , le nombre de pixels ayant le label i et correspondant à un pixel de l'ensemble X dans l'image \mathcal{I}^t . Un exemple d'histogramme de composition basé régions (lié à une sémantique) est illustré par la figure 11.1.

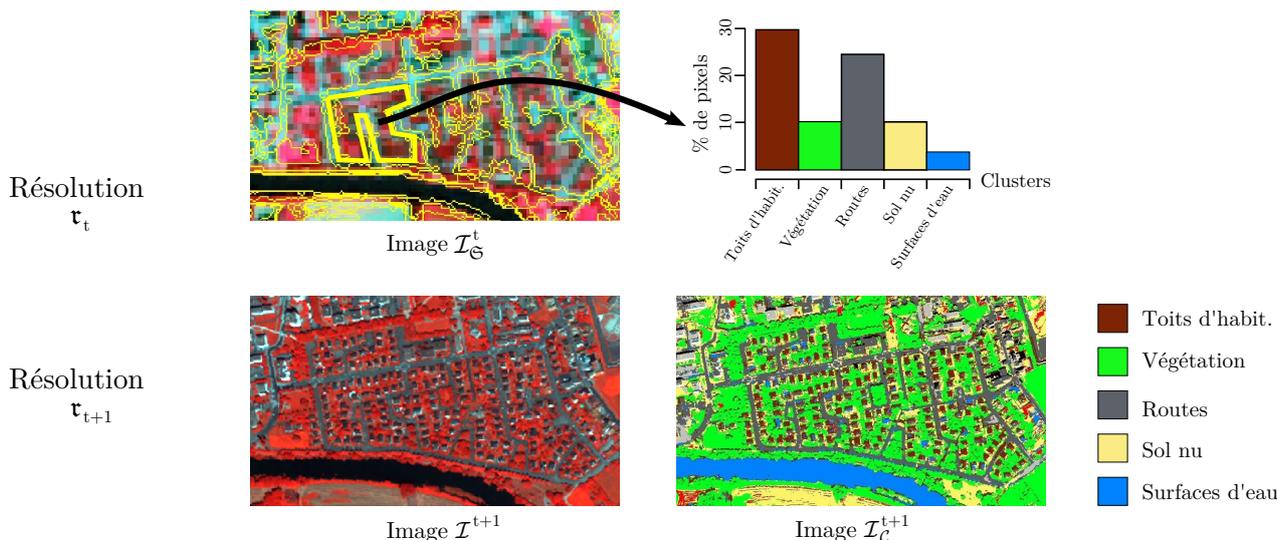


FIGURE 11.1 – Connaissances multirésolutions liées aux images : l’un des segments de la partition, délimité en jaune sur l’image I^t , est caractérisé par un histogramme de composition modélisant la répartition de ses pixels en termes de labels de clusters dans l’image classifiée I_C^{t+1} . Dans cet exemple une sémantique issue des connaissances du domaine a été associée aux clusters issues de l’image I_C^{t+1} .

Ces nouveaux histogrammes de composition reflètent ainsi d’une manière plus pertinente la composition sémantique des segments extraits à partir d’une résolution r_t et permettent de prendre en considération des informations/connaissances basées régions relatives au contenu d’une images à une résolution r_{t+1} .

B. Intégration de la distance HSBD dans l’approche de classification multirésolution

Pour comparer et classifier des régions décrites par ces histogrammes, nous proposons d’intégrer la distance HSBD dans l’algorithme de classification (utilisé dans l’approche de classification multirésolution). Pour comparer des régions extraites d’une image à une résolution r_t avec HSBD, l’utilisateur doit fournir une matrice de dissimilarité \mathcal{M}^{dis} qui modélise les relations sémantiques entre chaque instance du concept sémantique relatif aux clusters extraits de l’image à la résolution r_{t+1} .

Dans ces travaux, l’algorithme de classification utilisé est l’algorithme K -MEANS. Cet algorithme nécessite une fonction permettant de définir les objets moyens, caractérisant les centroïdes (liés aux clusters) sur lesquels l’algorithme itère. L’objet moyen d’un ensemble d’objets peut ici être considéré comme l’objet minimisant la somme des distances élevées au carré, relativement à l’ensemble des objets présents dans cet ensemble. En général, lors de l’utilisation de distances barre-à-barre, la fonction utilisée pour la redéfinition des histogrammes moyens (associés aux centroïdes) correspond à la moyenne arithmétique des barres correspondantes. Cependant lors de l’utilisation de distances entre histogrammes plus complexes, la moyenne arithmétique peut s’avérer insuffisante. Par exemple, dans le cas de l’utilisation de la distance temporelle DTW pour la comparaison d’histogrammes, de nombreux travaux utilisent des moyennes spécifiques [Gupta et al., 1996; Petitjean et al., 2011], spécialement conçues pour minimiser (*e.g.*, par le biais d’heuristiques) la somme des distances au carré des objets affectés à chaque centroïde. Les premières expériences que nous avons menées ont montré que la moyenne arithmétique s’avérait suffisante pour définir les histogrammes moyens attachés aux centroïdes lors du processus itératif de l’algorithme K -MEANS. De plus, nous avons observé expérimentalement que l’utilisation de cette moyenne n’empêche pas la convergence de l’algorithme. En effet, HSBD peut être considérée comme une somme pondérée de distances barre-à-barre. La moyenne arithmétique semble ainsi compatible avec cette dernière. Néanmoins, une perspective peut consister à étudier d’une manière plus formelle les moyennes compatibles avec HSBD.

L’intégration de la distance HSBD au sein de l’approche de classification multirésolution permet ainsi de prendre en compte des connaissances du domaine (apportées par l’utilisateur à une résolution r_{t+1}) de haut

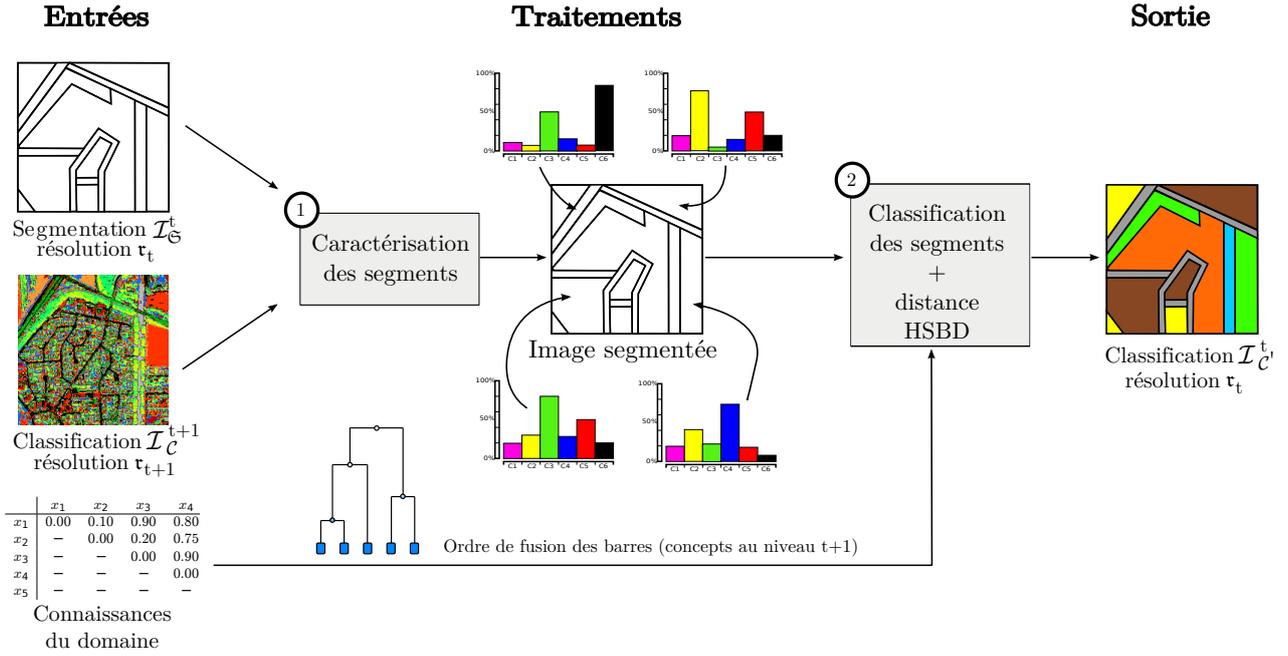


FIGURE 11.2 – Illustration du processus de post-classification monorésolution appliqué au traitement d'un niveau de résolution.

niveau pour améliorer la comparaison de ces données et ainsi les résultats de classification obtenus à une résolution τ_t .

C. Processus de post-classification mononiveau

Pour remettre en cause un résultat de classification particulier obtenu à une résolution τ_t par l'intermédiaire de PCSC, nous proposons de ré-appliquer l'approche de classification multirésolution, en caractérisant cette fois-ci les segments construits précédemment par des histogrammes de composition basés régions (et non plus basés pixels). Le principe de ce processus de post-classification mononiveau est détaillé ci-après. Les lecteurs peuvent aussi se référer à la figure 11.2 pour une description schématique de ce dernier.

Pour ce faire, on considère la segmentation \mathcal{S}_t obtenue initialement à la résolution τ_t ainsi que l'image classifiée $\mathcal{I}_{\mathcal{C}}^{t+1} : E^{t+1} \rightarrow \llbracket 1, k_{t+1} \rrbracket$. Cette image classifiée correspond au résultat de classification obtenu préalablement par l'intermédiaire de PCSC (pendant la phase descendante). On fait alors l'hypothèse que l'on est capable d'affecter aux étiquettes des clusters issus de ce résultat de classification, une sémantique correspondant aux classes d'intérêt recherchées. On considère également une fonction de correspondance $\lambda_{t \rightarrow t+1} : E^t \rightarrow E^{t+1}$ permettant de mettre en correspondance les pixels des images aux différentes résolutions.

Le principe consiste alors à caractériser chaque segment de la segmentation \mathcal{S}_t par le biais d'un histogramme de composition basé régions, calculé en considérant l'image classifiée basée régions $\mathcal{I}_{\mathcal{C}}^{t+1}$, obtenue à la résolution τ_{t+1} (voir la figure 11.2-①). Il devient alors possible de ré-appliquer l'approche de classification non-supervisée multirésolution présentée au cours du chapitre 6 (combinée à la distance HSB) pour catégoriser les régions caractérisées par ces histogrammes (voir la figure 11.2-②). Ce processus conduit à la création d'un ensemble de k_t clusters $\{K'_i\}_{i=1}^{k_t}$, associé à un ensemble de k_t centroïdes. Chaque centroïde u'_i est représenté par l'histogramme de composition moyen (normalisé) $\overline{H'(\mathcal{I}_{\mathcal{C}}^{t+1}, \lambda_{t \rightarrow t+1}(K'_i))}$ du cluster K'_i .

Les premières expériences réalisées ont montré que les clusters obtenus sont plus pertinents que ceux générés au cours de la descente et permettent de regrouper des régions décrites par des histogrammes sémantiquement similaires (mais pouvant être statistiquement différents).

11.2 Schéma méthodologique de C-PCSC

Pour remettre en cause les différents résultats de classification obtenus par l'intermédiaire de PCSC, nous proposons d'appliquer le processus de post-classification mononiveau défini précédemment pour chaque résolution considérée. Ce processus correcteur est nommé C-PCSC.

11.2.1 Entrées et sorties

Pour ce faire, on considère les n segmentations \mathfrak{S}_t obtenues initialement aux n résolutions \mathfrak{r}_t ainsi que les n images classifiées $\mathcal{I}_C^{t+1} : E^{t+1} \rightarrow [[1, k_{t+1}]]$. Ces n images classifiées sont les n résultats de classification obtenus préalablement par l'intermédiaire de PCSC. On fait alors l'hypothèse que l'on est capable d'affecter une sémantique aux étiquettes des clusters issus de ces n résultats de classification, et de produire n matrices de dissimilarité \mathcal{M}^{dis} qui modélisent les relations entre les instances considérées. On considère également $n-1$ fonctions de correspondance $\lambda_{t \rightarrow t+1} : E^t \rightarrow E^{t+1}$ permettant de mettre en correspondance les pixels des images aux différentes résolutions.

Le processus fournit alors n images classifiées \mathcal{I}_C^t , ainsi que n cartes de conflits permettant d'évaluer les différences de classification entre les résultats obtenus avec PCSC et les nouveaux résultats de post-classification. Nous verrons par la suite que ces cartes peuvent être employées pour la remise en cause des résultats de segmentation.

11.2.2 Schéma méthodologique

Le principe consiste alors à appliquer la stratégie de post-classification mononiveau définie précédemment pour chacune des n résolutions considérées. Il est à noter que ce processus de post-classification peut être appliqué aux résultats de PCSC en adoptant deux stratégies :

1. une stratégie descendante à travers la résolution : le processus de post-classification est appliqué dans un premier temps sur les régions des images aux résolutions les plus grossières pour terminer par les résolutions les plus fines ;
2. une stratégie ascendante à travers la résolution : le processus de post-classification est appliqué dans un premier temps sur les régions des images aux résolutions les plus fines pour terminer par les résolutions les plus grossières.

Comme les résultats de segmentation et de classification, obtenus à une résolution \mathfrak{r}_t , sont dépendant des résultats obtenus à une résolution \mathfrak{r}_{t-1} , nous proposons dans ces travaux que ce post-processus correcteur soit basé sur une approche ascendante.

Les n images classifiées résultantes \mathcal{I}_C^t permettent alors à l'utilisateur une analyse de la scène à plusieurs niveaux de sémantique.

11.2.3 Correction des résultats de classification

Une fois la post-classification des n images réalisée, on dispose alors pour chaque image \mathcal{I}^t de deux résultats de classification : la classification \mathcal{I}_C^t obtenue *via* PCSC et la classification \mathcal{I}_C^t , obtenue *via* C-PCSC. Il devient possible d'utiliser ces nouveaux résultats pour remettre en cause les résultats de classification préalablement obtenus par l'intermédiaire de PCSC. L'idée consiste à comparer ces deux classifications pour les corriger mutuellement.

Pour ce faire, trois étapes sont nécessaires :

1. mettre en correspondance les clusters issus des deux partitions (*i.e.*, classifications) obtenues ;
2. affecter des nouveaux labels aux régions affectées par la post-classification ;
3. générer des cartes de conflits afin de connaître les régions n'ayant pas pu être post-classifiées.

Nous détaillons par la suite chacune de ces trois étapes.

A. Mise en correspondance de partitions

Pour connaître les régions dont les labels ont changé entre la classification initiale et la post-classification, il est nécessaire de pouvoir mettre en correspondance les nouveaux résultats de classification avec les résultats initialement produits.

De nombreuses approches sont possibles pour mettre en correspondance deux classifications [Burge, 1981; Hubert et Arabie, 1985]. Parmi celles-ci, nous avons choisi d'effectuer la mise en correspondance des résultats de classification en prenant en compte, d'une part les modèles d'apprentissage relatifs aux classifications réalisées (*i.e.*, en appariant les centroïdes de la classification initiale à ceux de la nouvelle classification), et d'autre part les partitions relatives aux résultats de classification obtenus (*i.e.*, en appariant la partition formée dans l'espace des caractéristiques par les clusters issus de la classification initiale à la partition formée par les clusters issus de la post-classification).

Considérons deux images classifiées $\mathcal{I}_C^t : E \rightarrow \llbracket 1, k \rrbracket$, $\mathcal{I}_{C'}^t : E \rightarrow \llbracket 1, k \rrbracket$ ainsi que les deux ensembles de clusters associés \mathfrak{K} et \mathfrak{K}' . Dans le but de modéliser la correspondance entre les clusters des images \mathcal{I}_C^t et $\mathcal{I}_{C'}^t$, on définit deux fonctions de correspondance

$$\begin{aligned} \mu_{C \rightarrow C'} : \mathfrak{K} &\rightarrow 2^{\mathfrak{K}'} & \text{et} & \mu_{C' \rightarrow C} : \mathfrak{K}' &\rightarrow 2^{\mathfrak{K}} \\ K &\mapsto \{K_i, \dots, K_j\} & & K &\mapsto \{K_i, \dots, K_j\} \end{aligned} \quad (11.2)$$

qui associent à tout cluster de l'un des deux résultats de classification, un ensemble de clusters correspondant dans l'autre résultat de classification.

Pour illustrer leur fonctionnement, considérons un cluster de régions K_i , formé par PCSC à une résolution τ_i . Ce cluster regroupe par exemple des régions correspondant à des toits d'habitations individuelles. Considérons maintenant une région S , extraite de ce cluster. On fait alors l'hypothèse que l'étape de post-classification basée sur des connaissances a affecté la région S au cluster K_j . Pour savoir effectivement si le cluster K_j auquel à été attribué la région S lors de la post-classification, est sémantiquement différent du cluster K_i , on peut alors utiliser les fonctions de correspondance calculées précédemment. En effet, si les clusters K_i et K_j sont en correspondance directe, alors le label de la région n'a pas été modifié par le processus de post-classification. Dans le cas contraire, on peut alors affecter un nouveau label à la région S ce qui permet de remettre en cause et de corriger le résultat de classification précédent.

B. Affectation des nouveaux labels aux régions

Une fois cette mise en correspondance effectuée, on peut alors chercher, pour chaque niveau de classification considéré : (1) les régions dont les labels sont en correspondance entre la classification initiale et la post-classification et (2) les régions dont les labels ne sont pas en correspondance entre la classification initiale et la post-classification.

Pour chacune des régions dont les labels sont en correspondance, on peut alors estimer que la région avait été correctement classifiée par PCSC. Cependant, suivant le type de la correspondance, deux cas de figure peuvent apparaître :

1. la correspondance est directe (*i.e.*, $1 \rightarrow 1$) : le cluster auquel la région a été affectée pendant l'étape de classification initiale correspond à un et un seul cluster résultant de l'étape de post-classification (*e.g.*, un cluster regroupant des régions de forêt). La confiance est alors totale et la région conserve son label de cluster initial;
2. la correspondance est multiple (*i.e.*, $1 \rightarrow n$) : le cluster auquel la région a été affectée pendant l'étape de classification initiale (*e.g.*, un cluster regroupant des blocs urbains d'habitations individuelles) correspond à plusieurs clusters résultant de l'étape de post-classification (*e.g.*, un cluster regroupant des blocs urbains d'habitations individuelles aux toits gris, un cluster regroupant des blocs urbains d'habitations individuelles aux toits rouges). On peut alors s'interroger sur la pertinence du cluster auquel la région avait été initialement affectée. Une raison potentielle au phénomène de scission du cluster en plusieurs sous-clusters peut être liée au fait que les connaissances utilisées pendant l'étape de post-classification permettent de regrouper les régions à classifier d'une manière plus pertinente. De plus, les premières expériences réalisées ont montré que ces sous-clusters correspondent généralement à des groupes composés de régions plus spécialisées. La région est alors étiquetée avec le nouveau label de cluster obtenu

(*e.g.*, le label du cluster regroupant des blocs urbains d’habitations individuelles aux toits gris ou le label du cluster regroupant des blocs urbains d’habitations individuelles aux toits rouges). Cependant, on peut alors s’interroger quant à la pertinence du cluster initial formé par PCSC.

Pour chacune des régions dont les labels ne sont pas en correspondance, on peut estimer qu’un conflit est détecté : l’avis donné par la classification initiale, obtenue par l’intermédiaire de PCSC, est différent de celui donné par la post-classification, obtenue par l’intermédiaire de C-PCSC. On fait alors l’hypothèse que la nouvelle classification est plus pertinente que la précédente et l’on attribue alors à la région le nouveau label de cluster obtenu.

C. Création des cartes de conflits

En stockant, pour chaque niveau de post-classification, les régions dont les labels ne sont pas en correspondance, on peut obtenir n cartes de conflits apportant des informations supplémentaires relatives à la remise en cause des régions. Ces cartes permettent alors d’évaluer les différences de classification entre les résultats obtenus avec PCSC et les nouveaux résultats de post-classification.

Un label de région conduisant à un conflit peut provenir du fait qu’une erreur a été commise dans la segmentation de cette région et qu’il sera nécessaire par la suite de corriger les frontières du segment lui étant associé. Nous introduisons dans la section suivante la perspective d’un post-processus permettant d’utiliser une telle information pour remettre en cause l’étape de segmentation.

11.3 Vers un enrichissement itératif des résultats

11.3.1 Enrichissement de la segmentation par la classification

Le processus C-PCSC, présenté dans la section précédente, consiste à remonter dans la hiérarchie des régions formées par PCSC en considérant des connaissances relatives au contenu des images (générées pendant la descente) et des connaissances du domaine (apportées par l’utilisateur). Ce processus ascendant permet ainsi de corriger les labels de certains clusters erronés, affectés à certaines régions lors des classifications initiales obtenues au cours du processus descendant.

Dans le cas où des clusters (obtenus pendant la phase descendante) sont améliorés/modifiés, on peut alors s’interroger quant à la pertinence des modèles de segmentation utilisés au cours de la descente, pour segmenter les régions issues de ces clusters. En effet, comme l’étape de segmentation hiérarchique est guidée par cette classification, un changement dans les clusters implique une remise en cause de l’étape de segmentation.

En effet, considérons une région extraite par l’approche de segmentation hiérarchique (au cours du processus descendant) d’une image à une résolution τ_t . Faisons maintenant l’hypothèse que cette région est labellisée par l’étape de classification initiale, avec le label d’un cluster regroupant des régions de blocs urbains d’habitations individuelles. Cette région est ensuite projetée dans l’image à la résolution τ_{t+1} pour y être décomposée en suivant le modèle de segmentation défini pour former les régions correspondant, à la résolution τ_t , à des blocs urbains d’habitations individuelles.

Dans le cas où cette région se voit affecter un nouveau label de cluster, à la résolution τ_t , par l’étape de post-classification C-PCSC (*e.g.*, le label d’un cluster correspondant à des zones de végétation), il est alors nécessaire de prendre en compte ce changement lors de la segmentation de cette région à la résolution τ_{t+1} . Pour ce faire, on peut remettre en cause la segmentation de cette région à la résolution τ_{t+1} en re-partitionnant cette dernière suivant le modèle de segmentation défini pour segmenter les imagerie de type zones de végétation.

Le processus de post-classification C-PCSC permet, dans ce cas précis, d’enrichir les modèles de segmentation hiérarchique développés par l’intermédiaire de PCSC. Les connaissances utilisées par C-PCSC peuvent ainsi être également utilisées pour remettre en cause la segmentation.

11.3.2 Vers une approche ascendante itérative

Dans le cas où la post-classification des régions extraites à une résolution τ_t permet de remettre en cause et de corriger le modèle de segmentation à une résolution τ_{t+1} , les segments de cette résolution s’en trouvent

pouvait apporter une solution à ces problèmes. De telles structures peuvent être construites préalablement en fonction des connaissances du domaine de l'utilisateur, puis instanciées au cours de la descente par les connaissances extraites *via* PCSC, relatives au contenu des images. Finalement, ces ontologies peuvent être utilisées pour guider et contrôler un tel post-processus ascendant itératif, reposant sur C-PCSC. Les informations extraites des images par ce post-processus peuvent alors être employées pour compléter et mettre-à-jour les connaissances contenues dans ces ontologies, et d'une manière plus générale dans d'autres bases de connaissances.

11.4 Bilan scientifique

Ce chapitre a présenté un post-processus permettant une correction des résultats de classification obtenus par le biais de PCSC aux différents niveaux de résolution. Ce post-processus, nommé C-PCSC, repose sur un mécanisme de remise en cause des résultats de classification. Pour ce faire, les segments issus des différentes segmentations sont caractérisés par des histogrammes de composition basés régions, en considérant d'une part, des connaissances relatives au contenu des images traitées (obtenues après le traitement initial des données par le biais de PCSC), et d'autre part, des connaissances de haut niveau (fournies par l'utilisateur). Les régions résultantes sont ensuite catégorisées par une approche classique de classification non-supervisée, qui utilise la distance HSBD pour comparer les données. Ce processus correctif opère d'une manière ascendante à travers la résolution, en considérant en premier lieu les segments issus des images aux résolutions les plus fines, pour terminer par les segments issus des images aux résolutions les plus grossières.

Ces travaux ont finalement conduit à la perspective d'un post-processus ascendant itératif, basé sur le mécanisme de remise en cause présenté précédemment, permettant de corriger les résultats de segmentation obtenus par l'intermédiaire de PCSC.

Validations expérimentales

Sommaire

12.1 Protocole de validation	147
12.1.1 Contexte : post-classification des régions issues de PCSC	147
12.1.2 Post-classification à différents niveaux sémantiques	148
12.1.3 Description du protocole	148
12.2 Matériel	148
12.2.1 Jeux de données	148
12.2.2 Logiciel	149
12.3 Expérimentations, résultats et analyses	149
12.3.1 Validation de HSB D	149
12.3.2 Validation de C-PCSC	155
12.3.3 Robustesse	161
12.4 Bilan scientifique	163
12.4.1 Contributions	163
12.4.2 Valorisation scientifique	163

Ce chapitre présente les expérimentations réalisées pour valider la distance entre histogrammes HSB D ainsi que le processus C-PCSC, dans le contexte de l'analyse de scènes urbaines à partir d'images satellites multirésolutions. Nous commençons par décrire le protocole de validation proposé (Section 12.1). Nous détaillons ensuite les jeux de données employés ainsi que les développements effectués au sein d'une bibliothèque de classification, nécessaires pour réaliser ces expérimentations (Section 12.2). Nous présentons par la suite les expériences réalisées ainsi que les résultats obtenus, suivis d'une discussion sur ces derniers (Section 12.3). Enfin, nous dressons le bilan scientifique de ces travaux (Section 12.4).

12.1 Protocole de validation

12.1.1 Contexte : post-classification des régions issues de PCSC

Au cours des trois chapitres précédents, nous avons montré que les segments et les clusters formés au sein des ensembles d'images multirésolutions traitées par PCSC, étaient susceptibles de comporter des erreurs. De telles erreurs sont notamment liées à la propriété descendante de PCSC et à sa nature principalement non-supervisée. Ces dernières peuvent ensuite générer des imprécisions dans les cartes de classification résultantes, qui deviendront alors moins pertinentes pour un utilisateur.

Pour apporter une solution à ces problèmes, nous avons proposé d'intégrer dans le cadre d'un post-processus correcteur ascendant, des connaissances relatives aux images (acquises durant la descente par PCSC) ainsi que des connaissances du domaine. Ce post-processus, nommé C-PCSC est principalement basé sur une post-classification des régions initialement créées par PCSC (par l'intermédiaire de l'approche de segmentation hiérarchique basée sur des exemples). Cette post-classification est réalisée par l'intermédiaire de l'approche de classification multirésolution proposée précédemment, combinée à la distance HSB D intégrant des connaissances liées aux images et des connaissances du domaine.

Pour valider C-PCSC (et par la même occasion la distance HSB D), nous avons appliqué ce processus à la post-classification des régions obtenues par l'application préalable de PCSC. Nous présentons par la suite le

protocole de validation, suivi des résultats expérimentaux obtenus dans le cadre de l'analyse de scènes urbaines à partir d'images satellites multirésolutions. L'idée sous-jacente à ces expérimentations est de montrer qu'il est possible d'utiliser un post-processus automatique ascendant, guidé par les connaissances, permettant : (1) de corriger une partie des erreurs commises par un processus semi-automatique pendant la descente à travers la résolution, et (2) d'améliorer la précision des cartes produites, afin de permettre une meilleure compréhension de l'organisation spatiale des environnements urbains.

12.1.2 Post-classification à différents niveaux sémantiques

Nous proposons de valider C-PCSC et HSBD dans le cadre de la post-classification des régions issues des résultats fournis par l'application de PCSC. Ces résultats préalables permettent une analyse des territoires urbains à trois niveaux de sémantique, définis par la nomenclature hiérarchique présentée précédemment (Table 8.1). Ainsi, pour chaque jeu de données, les trois résultats de segmentation/classification (correspondant à trois niveaux d'analyse distincts) peuvent être remis en cause et corrigés.

Nous rappelons ci-dessous ces trois niveaux de sémantique, permettant de cartographier des surfaces urbanisées :

1. le niveau des quartiers urbains (Table 8.1, niveau 1) ;
2. le niveau des blocs urbains (Table 8.1, niveau 2) ;
3. le niveau des objets urbains simples (Table 8.1, niveau 3).

Nous proposons de remettre en cause et de corriger les résultats de segmentation et de classification liés à ces trois niveaux d'analyse, obtenus dans le cadre des expérimentations présentées dans le chapitre 8.

12.1.3 Description du protocole

Pour évaluer le processus d'analyse C-PCSC, ainsi que la distance entre histogrammes HSBD, nous avons développé un protocole expérimental reposant sur trois critères principaux :

1. **intérêt et pertinence de la distance HSBD** : nous proposons d'évaluer l'intérêt et la pertinence de la distance HSBD pour la comparaison d'histogrammes nominaux. Pour ce faire, une étude comparative a été réalisée sur les résultats de post-classification obtenus avec (ou sans) l'utilisation de cette distance. Les expériences menées sont présentées dans la section 12.3.1 ;
2. **précision et qualité de C-PCSC** : nous proposons ensuite d'évaluer l'intérêt de C-PCSC dans le cadre de la post-classification des régions de l'image. Dans cette optique, nous avons calculé le degré de concordance entre les objets d'intérêt extraits des images par C-PCSC, et les cartes de vérité terrain fournies par les experts en cartographie urbaine. Les expériences réalisées pour évaluer qualitativement les résultats produits par C-PCSC sont détaillées dans la section 12.3.2 ;
3. **robustesse** : finalement, nous avons étudié le degré de robustesse du processus. Pour ce faire, nous avons évalué l'impact de la précision des valeurs de proximité sémantique fournies par l'expert pour le calcul de la distance HSBD au cours de l'étape de post-classification de C-PCSC. Les résultats de ces expérimentations sont présentés dans la section 12.3.3.

12.2 Matériel

12.2.1 Jeux de données

Les expériences ont été réalisées sur les trois jeux de données nommés **ELSAU**, **ILLKIRCH** et **HAUTE-PIERRE** (Figure 8.1), présentés dans la partie précédente.

Pour chacun de ces jeux de données, on dispose (après l'application préalable de PCSC, voir le chapitre 8) de n résultats de segmentation et de classification pour l'ensemble des n images considérées (correspondant aux trois niveaux de sémantique évoqués précédemment). Afin de valider le processus C-PCSC (et la distance HSBD), ce dernier a été appliqué à la post-classification des segments issus de PCSC. Pour chaque niveau de résolution r_i à post-classifier, on considère :

Jeu de données	Image	Nombre de bandes	Résolution spatiale	Nombre de segments
ELSAU	MRS	4 bandes	9,6 m	859
	HRS	4 bandes	2,4 m	4 206
	THRS	1 bande	60 cm	30 546
ILLKIRCH	MRS	4 bandes	9,6 m	938
	HRS	4 bandes	2,4 m	5 632
	THRS	1 bande	60 cm	39 854
HAUTEPIERRE	MRS	4 bandes	9,6 m	632
	HRS	4 bandes	2,4 m	2 763
	THRS	1 bande	60 cm	20 698

TABLE 12.1 – Synthèse des jeux de données utilisés pour les expériences.

- le résultat de segmentation \mathcal{I}_G^t obtenu à la résolution \mathbf{r}_t ;
- le résultat de classification basée régions \mathcal{I}_C^{t+1} obtenu à la résolution \mathbf{r}_{t+1} .

Pour chaque résultat de classification basée régions \mathcal{I}_C^{t+1} , les experts ont alors affecté une sémantique (correspondant aux classes d'intérêt recherchées) aux étiquettes des clusters extraits. La table 8.3 illustre les classes thématiques associées aux clusters extraits des images par l'application préalable de PCSC.

Les jeux de données considérés ainsi que les nombres de segments (créés par PCSC) à post-classifier par l'intermédiaire de C-PCSC sont présentés par la table 12.1. Les résultats de segmentation et de classification considérés sont quant à eux illustrés par les figures 8.8, 8.9, 8.10.

12.2.2 Logiciel

Afin de pouvoir classifier les régions issues de PCSC, par l'intermédiaire de la distance HSBD, pendant l'étape de post-classification du processus C-PCSC, cette distance a été implantée dans la bibliothèque de classification JCL, issue des travaux de notre équipe de recherche. Cette bibliothèque est suffisamment générique pour supporter l'intégration de nouveaux types d'attributs ou encore de nouvelles distances pour la comparaison des données à classifier. Elle est en cours d'intégration dans la bibliothèque Java MUSTIC permettant l'analyse d'images satellites. Cette dernière est détaillée en annexe de ce manuscrit (Annexe D). Nous planifions d'intégrer totalement le processus C-PCSC au sein de ces bibliothèques et de distribuer ce logiciel sous une licence libre.

12.3 Expérimentations, résultats et analyses

12.3.1 Validation de HSBD

Dans la première partie de ces expérimentations, nous proposons d'évaluer l'intérêt de la distance HSBD pour la comparaison d'histogrammes nominaux. Pour ce faire, nous avons comparé la qualité des résultats de post-classification obtenus avec ou sans l'utilisation de cette distance, au sein de l'algorithme de classification considéré.

A. Protocole expérimental

Expériences menées Pour valider la distance proposée, nous l'avons intégrée à l'approche de post-classification, utilisée pour classifier les segments extraits des images aux niveaux des blocs urbains. Nous rappelons que ces derniers peuvent être définis comme les ensembles urbains minimaux, fermés par des voies de communication. La principale originalité de cette tâche est de classifier des ensembles de blocs urbains

x_i	Ombre	Eau	Route	Voie ferrée	Sol nu	Végét. chloro.	Forêt	Toit ardoise	Toit tuil. rouge.	Toit tuil. gris.	Toit métal.
Ombre	0,00	0,10	1,00	1,00	1,00	1,00	0,80	0,80	0,80	0,80	0,80
Eau	-	0,00	0,50	0,50	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Route	-	-	0,00	0,10	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Voie ferrée	-	-	-	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Sol nu	-	-	-	-	0,00	0,10	0,80	1,00	1,00	1,00	1,00
Végét. chloro.	-	-	-	-	-	0,00	0,80	1,00	1,00	1,00	1,00
Forêt	-	-	-	-	-	-	0,00	1,00	1,00	1,00	1,00
Toit ardoise	-	-	-	-	-	-	-	0,00	0,10	0,40	0,90
Toit tuiles rouges	-	-	-	-	-	-	-	-	0,00	0,40	0,90
Toit tuiles grises	-	-	-	-	-	-	-	-	-	0,00	0,10
Toit métallique	-	-	-	-	-	-	-	-	-	-	0,00

TABLE 12.2 – Matrice de dissimilarité \mathcal{M}^{dis} associée aux instances du concept OBJET URBAIN.

dans lesquels chaque bloc est caractérisé par sa composition en termes d’objets urbains simples (*e.g.*, maisons individuelles, routes).

Dans le cadre de ces expérimentations, chaque segment peut être caractérisé par un histogramme de composition basé régions (lié au concept sémantique OBJET URBAIN). Ainsi, un bloc urbain U_i peut être caractérisé par un histogramme $H_i(\#Toit tuiles rouges, \#Toit ardoise, \dots, \#Végétation chlorophyllienne)$ où Toit tuiles rouges, Toit ardoise, \dots , Végétation chlorophyllienne, sont des instances du concept sémantique OBJET URBAIN.

La principale difficulté est de parvenir à regrouper, dans un même cluster (ou classe), différents objets qui ne sont pas caractérisés par des histogrammes statistiquement similaires. L’exemple suivant illustre ce problème. Considérons le bloc U_i caractérisé par un histogramme $H_i(15, 3, \dots, 10)$ (*i.e.*, 15 toits en tuiles rouges, 3 toits en ardoise, \dots , 10 parcelles de végétation) et un bloc U_j caractérisé par un histogramme $H_j(3, 22, \dots, 10)$ (*i.e.*, 3 toits en tuiles rouges, 22 toits en ardoise, \dots , 10 parcelles de végétation). Du point de vue de l’utilisateur, ces deux blocs doivent être groupés dans la même classe « Blocs d’habitations individuelles », car ils sont tous les deux composés de maisons individuelles (avec des toits à tuiles rouges ou en ardoise) et des parcelles de végétation. Une solution pour résoudre ce problème consiste à utiliser un processus de classification associé à une distance prenant en compte les corrélations sémantiques des données. Pour valider l’utilité de la distance HSBD, nous l’avons intégrée dans un algorithme de classification qui est lui-même une brique de base de l’approche de classification multirésolution proposée au cours de cette thèse.

Nous disposons des trois jeux de données **ELSAU**, **ILLKIRCH** et **HAUTEPIERRE**, définis précédemment. Pour chacun de ces jeux de données, la segmentation réalisée sur l’image à la résolution τ_2 permet d’obtenir un ensemble de segments pouvant correspondre aux blocs urbains de la scène. Chaque segment S_i a ainsi été caractérisé par un histogramme de composition basé régions (lié au concept sémantique OBJET URBAIN) $H(S_i)$ qui modélise la décomposition de ce segment en fonction de la distribution des onze instances du concept

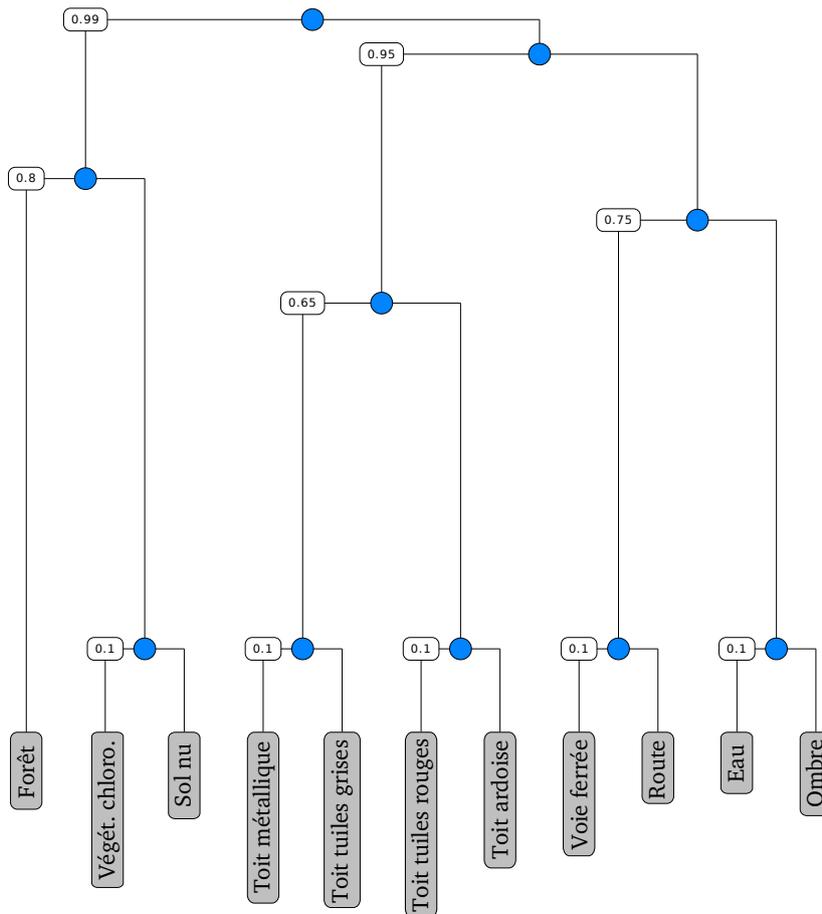


FIGURE 12.1 – Dendrogramme associé à la matrice de dissimilarité présentée dans la table 12.2. Le critère *average-link* a été utilisé. Les instances initiales sont représentées par des rectangles gris tandis que les instances de plus hauts niveaux sémantiques sont représentées par des disques bleus.

sémantique OBJET URBAIN dans l'image à la résolution τ_3 . Ces onze instances sont listées en table 8.3.

Construction de l'ordre de fusion Pour modéliser les relations sémantiques entre les barres de ces histogrammes, une matrice de dissimilarité \mathcal{M}^{dis} de dimension 11×11 a été définie par un photo-interprète (Table 12.2). Cette matrice reflète les connaissances du domaine relatives au niveau d'analyse des objets urbains simples. Les valeurs de dissimilarité modélisées par cette matrice permettent de considérer les proximités sémantiques entre les barres de ces histogrammes (et les classes thématiques liées à chacune de ces barres).

À partir de cette matrice, un dendrogramme peut être construit pour modéliser l'ordre de fusion des instances du concept sémantique considéré (Figure 12.1). La construction de ce dendrogramme repose sur l'algorithme AHC (voir la Section 10.3.2). Cet algorithme construit hiérarchiquement des clusters d'instances minimisant une mesure d'inertie intra-cluster. Pour calculer les valeurs de dissimilarité entre chaque couple de groupes, il est nécessaire de choisir un critère de fusion. Trois critères de fusion principaux peuvent généralement être considérés :

1. le critère *single-link* retient le minimum des distances entre les instances composant deux groupes ;
2. le critère *complete-link* retient le maximum des distances entre les instances les plus éloignées composant deux groupes ;
3. le critère *average-link* consiste à calculer la moyenne géométrique des distances entre les instances composant deux groupes.

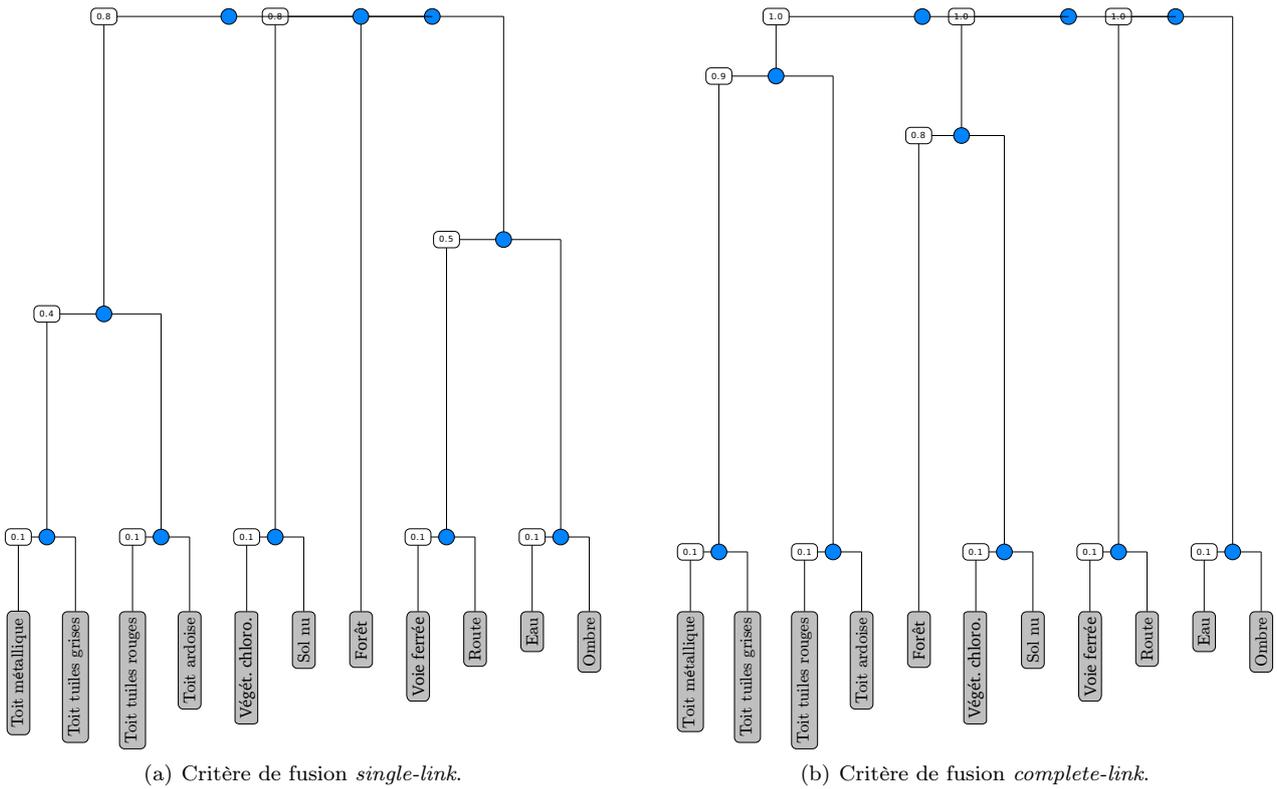


FIGURE 12.2 – Dendrogrammes associés à la matrice de dissimilarité présentée dans la table 12.2. Les instances initiales sont représentées par des rectangles gris tandis que les instances de plus hauts niveaux sémantiques sont représentées par des disques bleus.

Pour évaluer l’impact du choix de ce critère de fusion sur la pertinence de la distance HSBD, des expériences ont été réalisées pour construire différents dendrogrammes *via* l’utilisation de ces critères.

Trois dendrogrammes différents ont ainsi été générés (Figures 12.1, 12.2(a–b)), puis présentés à un expert en cartographie urbaine, qui a évalué visuellement leur pertinence. Les résultats de cette analyse ont montré que le critère *average-link* fournissait les résultats les plus en accord avec les notions sémantiques associées aux instances de haut niveau, construites artificiellement par le processus de fusion. En conséquence, le critère *average-link* a été choisi pour la suite de ces expérimentations.

Choix de la sous-distance barre-à-barre Le principe de la distance HSBD est de calculer hiérarchiquement une sous-distance barre-à-barre D_{bin} entre deux histogrammes à différents niveaux de sémantique. Pour évaluer l’influence de la distance barre-à-barre choisie D_{bin} sur la qualité des résultats obtenus, la distance HSBD a été directement intégrée dans l’algorithme *K-MEANS* afin de comparer les histogrammes classifiés. Nous avons appliqué l’algorithme *K-MEANS* en associant la distance HSBD à différentes sous-distances barre-à-barre ($HSBD_{L_1}$, $HSBD_{L_2}$, et $HSBD_{\chi^2}$). Pour comparer HSBD à d’autres distances existantes, l’algorithme *K-MEANS* a aussi été appliqué en utilisant des distances classiques barre-à-barre. Ces comparaisons permettent d’évaluer les avantages apportés par l’utilisation de la distance HSBD par rapport à l’utilisation d’une distance barre-à-barre :

1. $HSBD_{L_1}$ vs. D_{L_1} ;
2. $HSBD_{L_2}$ vs. D_{L_2} ;
3. $HSBD_{\chi^2}$ vs. D_{χ^2} .

Les mêmes nombres de clusters que dans les expérimentations relatives à la validation de PCSC ont été employés (voir la section 8.3). Les résultats fournis par l’algorithme *K-MEANS* étant sensibles à son initialisation,

Jeu de données	Distance barre-à-barre	$\overline{\mathcal{F}} \pm \sigma$		$\mathcal{K} \pm \sigma$	
		D_{bin}	HSBD $_{bin}$	D_{bin}	HSBD $_{bin}$
ELSAU	L_1	0,65 ± 0,02	0,71 ± 0,01	0,37 ± 0,02	0,41 ± 0,01
	L_2	0,63 ± 0,03	0,69 ± 0,02	0,36 ± 0,03	0,39 ± 0,02
	χ^2	0,59 ± 0,02	0,65 ± 0,02	0,33 ± 0,03	0,36 ± 0,02
ILLKIRCH	L_1	0,67 ± 0,02	0,72 ± 0,01	0,38 ± 0,02	0,47 ± 0,02
	L_2	0,64 ± 0,01	0,70 ± 0,02	0,37 ± 0,02	0,44 ± 0,01
	χ^2	0,61 ± 0,02	0,68 ± 0,03	0,33 ± 0,01	0,39 ± 0,02
HAUTEPIERRE	L_1	0,63 ± 0,01	0,66 ± 0,01	0,34 ± 0,01	0,38 ± 0,01
	L_2	0,60 ± 0,03	0,63 ± 0,01	0,34 ± 0,02	0,37 ± 0,01
	χ^2	0,58 ± 0,02	0,62 ± 0,02	0,31 ± 0,01	0,35 ± 0,02

TABLE 12.3 – Résultats d'évaluation globaux sur les jeux de données **ELSAU**, **ILLKIRCH**, **HAUTEPIERRE**, au niveau d'analyse des blocs urbains.

chaque test a été répété dix fois en faisant varier les « graines » de l'algorithme. Nous avons alors calculé la variance σ obtenue pour chaque série de tests (et pour chaque indice d'évaluation considéré).

B. Évaluation des résultats

Les résultats de classification obtenus par l'intermédiaire de K -MEANS combiné à la distance HSBD, ont été évalués qualitativement par des comparaisons avec des cartes de vérité terrain certifiées (au niveau sémantique des blocs urbains), extraites de différentes bases de données relatives à la couverture des sols. Ces cartes ont été présentées précédemment dans la section 8.3.

Pour comparer les résultats de classification obtenus aux cartes de vérité terrain, nous avons calculé des indices de qualité à un niveau local et à un niveau global. Ces indices ont été détaillés dans la section 8.3 et sont récapitulés dans la table 8.4.

C. Analyse des résultats

Les résultats d'évaluation locaux, obtenus au niveau des blocs urbains sur les trois jeux de données considérés, sont présentés par la figure 12.3. En analysant ces graphes, on peut remarquer que les scores de F-mesure, obtenus pour chaque classe thématique extraite, sont toujours plus importants quand l'algorithme de classification K -MEANS est combiné à la distance HSBD que lorsque ce dernier est combiné à la distance barre-à-barre correspondante D_{bin} . En particulier, les meilleurs scores ont été obtenus quand la distance HSBD est utilisée avec la sous-distance barre-à-barre de Manhattan D_{L_1} , tandis que les scores les plus faibles ont été obtenus quand la distance HSBD est utilisée avec la sous-distance barre-à-barre D_{χ^2} . Cependant, les scores obtenus avec la sous-distance barre-à-barre D_{χ^2} demeurent plus élevés que ceux obtenus par l'algorithme K -MEANS, combiné à une distance barre-à-barre classique. Ces différents scores d'évaluation locaux montrent que les résultats obtenus par l'intermédiaire de la distance HSBD permettent d'augmenter la précision et le rappel des résultats de classification.

Les résultats d'évaluation globaux, obtenus au niveau des blocs urbains sur les trois jeux de données considérés, sont présentés par la table 12.3. Ces derniers conduisent aux mêmes observations que les résultats d'évaluation locaux. Les scores globaux obtenus sont toujours supérieurs quand l'algorithme de classification K -MEANS est combiné à la distance HSBD que lorsque cet algorithme est combiné à la distance barre-à-barre correspondante D_{bin} . Par ailleurs, les meilleurs scores ont été obtenus lorsque la distance HSBD est combinée avec la sous-distance de Manhattan D_{L_1} .

Les résultats d'évaluation obtenus par le biais de ces expérimentations montrent que la combinaison de l'algorithme K -MEANS, à la distance HSBD, produit de meilleurs résultats de classification que la combinaison

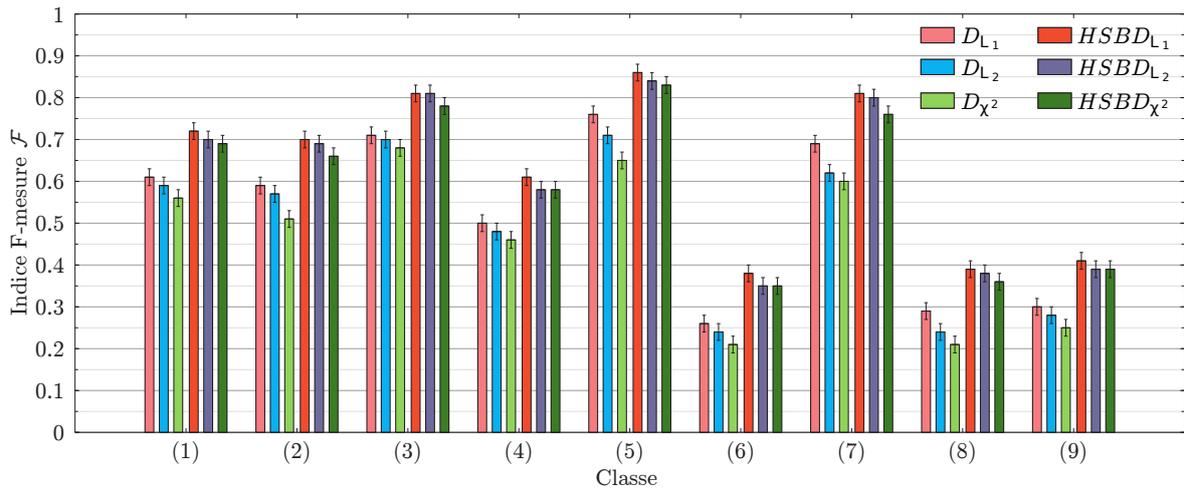
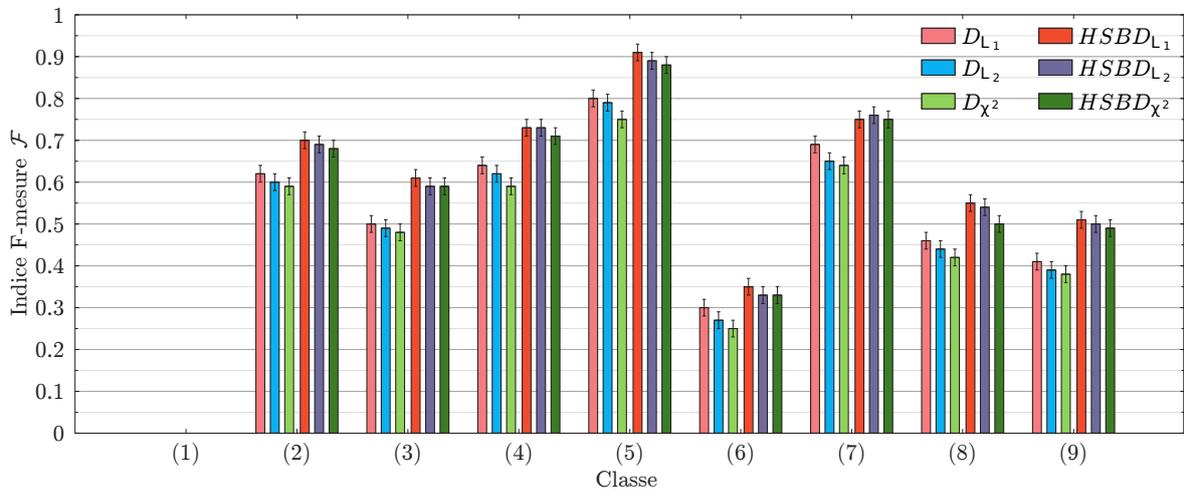
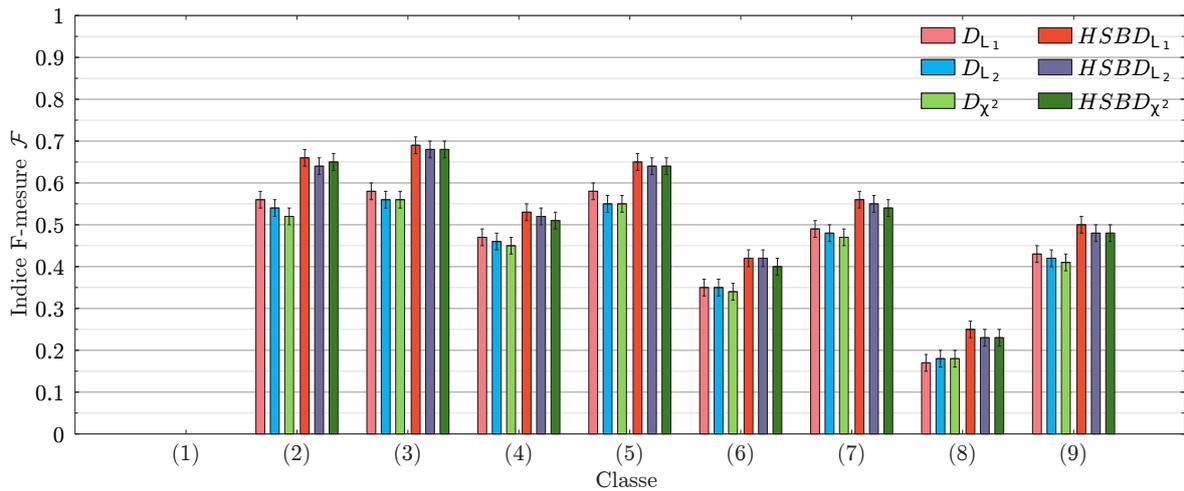
(a) Évaluation du jeu de données **ELSAU**.(b) Évaluation du jeu de données **ILLKIRCH**.(c) Évaluation du jeu de données **HAUTEPIERRE**.

FIGURE 12.3 – Résultats d'évaluation locaux sur les jeux de données **ELSAU**, **ILLKIRCH** et **HAUTEPIERRE**. Les indices des classes correspondent à ceux définis dans la table 8.3, au niveau des blocs urbains. Il est à noter que la classe (1) n'est pas disponible dans les jeux de données **ILLKIRCH** et **HAUTEPIERRE**.

x_i		Blocs collec.	Blocs ind.	Blocs fort. dens.	Blocs indus.	Surfaces d'eau	Routes	Zones agricoles	Zones de végét.	Forêts
Blocs collec.		0,00	0,40	0,60	0,90	1,00	1,00	1,00	1,00	1,00
Blocs ind.		-	0,00	0,50	0,80	1,00	1,00	1,00	1,00	1,00
Blocs fort. dens.		-	-	0,00	0,80	1,00	1,00	1,00	1,00	1,00
Blocs indus.		-	-	-	0,00	1,00	1,00	1,00	1,00	1,00
Surfaces d'eau		-	-	-	-	0,00	1,00	1,00	1,00	1,00
Routes		-	-	-	-	-	0,00	1,00	1,00	1,00
Zones agricoles		-	-	-	-	-	-	0,00	0,20	0,80
Zones de végét.		-	-	-	-	-	-	-	0,00	0,70
Forêts		-	-	-	-	-	-	-	-	0,00

TABLE 12.4 – Matrice de dissimilarité \mathcal{M}^{dis} associée aux instances du concept BLOCS URBAIN.

de cet algorithme à une distance classique barre-à-barre, lors de la comparaison d'histogrammes nominaux portant des valeurs sémantiques. En conséquence, la distance HSBD peut être considérée comme plus pertinente qu'une distance barre-à-barre classique dans le cadre de la comparaison d'histogrammes nominaux. Finalement, ces validations, dans le contexte de la classification de données géographiques, mettent en évidence l'intérêt de la distance HSBD pour des tâches plus générales relatives à la fouille de données.

12.3.2 Validation de C-PCSC

Dans la seconde partie de ces expérimentations, nous proposons d'évaluer la précision et l'intérêt du processus C-PCSC pour la post-classification des régions formées préalablement par PCSC. Pour ce faire nous avons comparé les résultats de post-classification obtenus par l'intermédiaire de cette approche corrective ascendante aux résultats initialement obtenus lors de la phase descendante par le biais de comparaisons à des cartes de vérité terrain.

A. Protocole expérimental

Pour valider ce post-processus correcteur, nous l'avons appliqué à la post-classification des segments/régions extraits des images aux différentes résolutions considérées. Pour classifier ces segments, l'approche de classification multirésolution, couplée à la distance entre histogrammes HSBD, a été employée. Dans le cadre de ces expérimentations, chaque segment peut être caractérisé par un histogramme de composition basé régions (lié à un concept sémantique particulier). Le processus C-PCSC reposant sur une approche ascendante, nous l'avons appliqué à la classification des segments/régions issus :

1. des images aux résolutions τ_3 (*i.e.*, images THRS) ;
2. des images aux résolutions τ_2 (*i.e.*, images HRS) ;
3. des images aux résolutions τ_1 (*i.e.*, images MRS).

Du fait que les segments issus de l'image à la résolution τ_3 ne sont pas caractérisés par des histogrammes de composition (*i.e.*, cas particulier de PCSC), l'application de C-PCSC à cette résolution n'a aucune influence

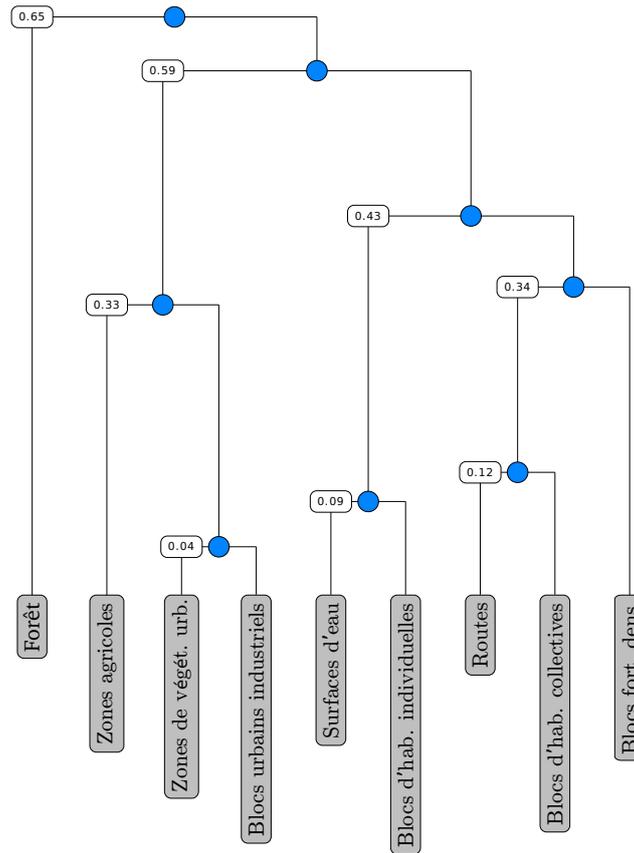


FIGURE 12.4 – Dendrogramme associé à la matrice de dissimilarité présentée dans la table 12.4. Les instances initiales sont représentées par des rectangles gris tandis que les instances de plus hauts niveaux sémantiques sont représentées par des disques bleus.

sur le résultat de classification \mathcal{I}_C^3 de l'image \mathcal{I}_3 . En conséquence, le processus C-PCSC a été validé pour la post-classification des segments/régions issus des images aux résolutions τ_2 (*i.e.*, niveau des blocs urbains) et τ_1 (*i.e.*, niveau des quartiers urbains).

Pour chacun des jeux de données considérés, la segmentation \mathcal{I}_C^2 réalisée sur l'image à la résolution τ_2 permet d'obtenir un ensemble de segments pouvant correspondre aux blocs urbains de la scène. Chaque segment S_i peut ainsi être caractérisé par un histogramme de composition basé régions (lié au concept sémantique OBJET URBAIN) $H(S_i)$, qui modélise la décomposition de ce segment en fonction de la distribution des instances du concept sémantique OBJET URBAIN dans l'image à la résolution τ_3 . Les onze instances du concept sémantique OBJET URBAIN sont listées dans la table 8.3. À partir de cette matrice, un dendrogramme à sept étapes peut être construit (voir la Section 10.3.2) afin de modéliser l'ordre de fusion des instances du concept sémantique considéré (Figure 12.1). Ces segments, caractérisés par des histogrammes de composition basés régions (liés au concept sémantique OBJET URBAIN), ont ensuite été labellisés *via* l'approche de classification non-supervisée multirésolution. Il a été choisi, en accord avec les experts, d'utiliser le même nombre de clusters $k_2 = 15$ que pendant la phase descendante liée à PCSC.

Une fois que les régions de l'image à la résolution τ_2 ont été post-classifiées pour former un nouveau résultat de classification \mathcal{I}_C^2 , (*i.e.*, au niveau des blocs urbains), il devient possible de ré-appliquer la même stratégie pour traiter l'image à la résolution τ_1 . Pour chacun des jeux de données considérés, la segmentation \mathcal{I}_C^1 réalisée sur l'image à la résolution τ_1 permet d'obtenir un ensemble de segments pouvant correspondre aux quartiers urbains et aux grandes zones de la scène. Chaque segment S_i peut ainsi être caractérisé par un histogramme de composition basé régions (lié au concept sémantique BLOC URBAIN) $H(S_i)$, qui modélise la décomposition de ce segment en fonction de la distribution des instances du concept sémantique BLOC URBAIN

Niv.	Classe	ELSAU					ILLKIRCH					HAUTEPIERRE				
		\mathcal{P}	\mathcal{R}	\mathcal{F}	$\mathcal{K} \pm \sigma$	$\overline{\mathcal{F}} \pm \sigma$	\mathcal{P}	\mathcal{R}	\mathcal{F}	$\mathcal{K} \pm \sigma$	$\overline{\mathcal{F}} \pm \sigma$	\mathcal{P}	\mathcal{R}	\mathcal{F}	$\mathcal{K} \pm \sigma$	$\overline{\mathcal{F}} \pm \sigma$
2. Blocs	(1)	0,68	0,76	0,73	0,41 ± 0,01	0,71 ± 0,01	0,60	0,70	0,65	0,47 ± 0,01	0,72 ± 0,02	0,63	0,84	0,72	0,38 ± 0,01	0,66 ± 0,01
	(2)	0,74	0,75	0,74			0,64	0,43	0,51			0,53	0,64	0,60		
	(3)	0,61	0,63	0,64			•	•	•			•	•	•		
	(4)	0,55	0,52	0,53			0,65	0,72	0,68			0,59	0,42	0,49		
	(5)	0,75	0,84	0,79			0,78	0,87	0,82			•	•	•		
	(6)	0,31	0,28	0,29			0,33	0,31	0,32			0,28	0,54	0,37		
	(7)	0,72	0,74	0,73			0,68	0,75	0,71			0,55	0,49	0,52		
	(8)	0,21	0,63	0,31			0,35	0,77	0,48			0,12	0,27	0,17		
	(9)	0,2	0,51	0,34			0,46	0,53	0,48			0,35	0,62	0,45		
1. Quartiers	(1)	0,68	0,73	0,71	0,42 ± 0,01	0,68 ± 0,01	0,57	0,82	0,63	0,43 ± 0,01	0,63 ± 0,02	0,63	0,83	0,71	0,36 ± 0,02	0,67 ± 0,01
	(2)	0,50	0,48	0,49			0,45	0,55	0,50			0,44	0,48	0,46		
	(3)	0,70	0,91	0,79			0,39	0,85	0,52			•	•	•		
	(4)	0,10	0,35	0,23			0,15	0,42	0,31			0,55	0,62	0,58		
	(5)	0,68	0,75	0,71			0,73	0,71	0,72			0,85	0,86	0,86		
	(6)	0,26	0,72	0,37			0,65	0,88	0,72			0,39	0,62	0,47		

TABLE 12.5 – Résultats d'évaluation obtenus pour les jeux de données **ELSAU**, **ILLKIRCH**, **HAUTEPIERRE**. Le symbole (•) signifie que la classe thématique considérée n'est pas disponible dans le jeu de données courant.

dans l'image à la résolution τ_2 . Les neuf instances du concept sémantique BLOC URBAIN sont listées dans la table 12.4. À partir de cette matrice, un dendrogramme à neuf étages peut être construit afin de modéliser l'ordre de fusion des instances du concept sémantique considéré (Figure 12.4). Ces segments, caractérisés par des histogrammes de composition basés régions (liés au concept sémantique BLOC URBAIN), ont ensuite été labellisés *via* l'approche de classification non-supervisée multirésolution. Il a été choisi, en accord avec les experts, d'utiliser le même nombre de clusters $k_1 = 10$ que pendant la phase descendante liée à PCSC. On obtient alors un nouveau résultat de classification $\mathcal{I}_{\mathcal{C}'}^1$ (*i.e.*, au niveau des quartiers urbains).

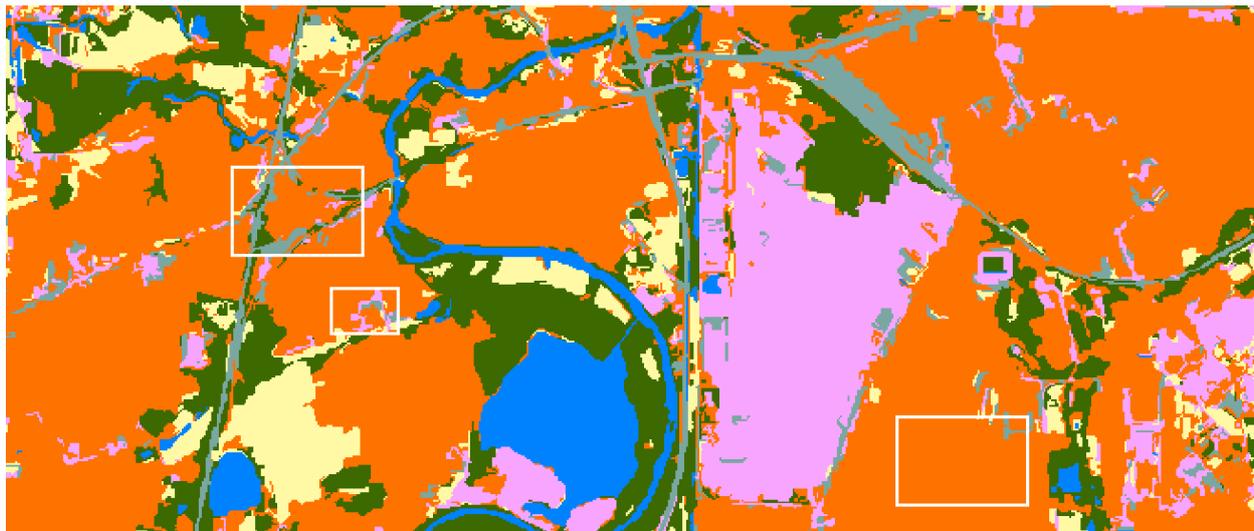
B. Évaluation des résultats

Les résultats de classification obtenus par l'intermédiaire du processus C-PCSC ont été évalués quantitativement par des comparaisons avec des cartes de vérité terrain (Section 8.3). Pour comparer les résultats de classification obtenus aux cartes de vérité terrain, nous avons calculé des indices de qualité à un niveau local et à un niveau global (Section 8.3).

C. Analyse des résultats

Nous détaillons ci-après les résultats de post-classification obtenus par l'intermédiaire de C-PCSC (au niveau des blocs et des quartiers urbains) sur les trois jeux de données considérés dans ces expérimentations.

Post-classification des régions *via* C-PCSC Le processus C-PCSC a été appliqué dans un premier temps sur les images HRS, pour classifier les blocs urbains de l'image. La table 12.5 détaille les différents scores d'évaluation globaux obtenus. Les résultats de post-classification obtenus (sur les jeux de données **ELSAU** et **ILLKIRCH**) sont présentés par les figures 12.5(b), 12.6(b). Après l'étape de post-classification, les comparaisons entre les régions classifiées résultantes et les cartes de vérité terrain permettent d'obtenir des valeurs de Kappa et de F-mesure égales à ($\mathcal{K} = 0,41$ et $\overline{\mathcal{F}} = 0,71$) pour le jeu de données **ELSAU**, ($\mathcal{K} = 0,47$ et $\overline{\mathcal{F}} = 0,72$) pour le jeu de données **ILLKIRCH**, et ($\mathcal{K} = 0,38$ et $\overline{\mathcal{F}} = 0,66$) pour le jeu de données **HAUTEPIERRE**. Les classes relatives aux blocs urbains industriels et d'habitations (classes (1–4))



(a) Post-classification MRS – niveau des quartiers.



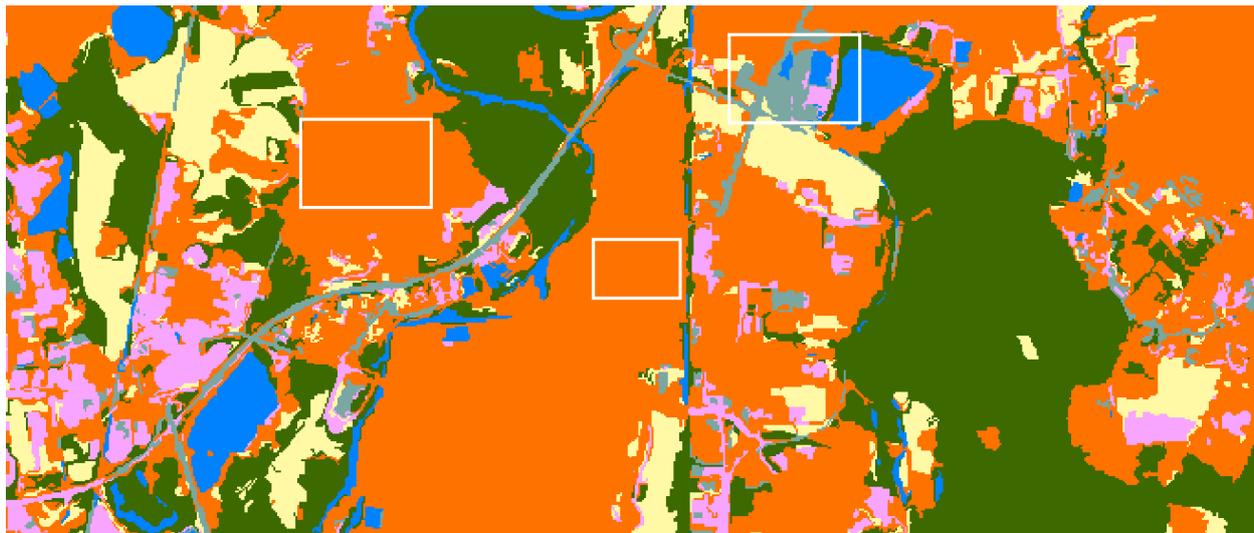
(b) Post-classification HRS – niveau des blocs.

FIGURE 12.5 – Résultats de post-classification pour le jeu de données **ELSAU** (6 576 m × 2 793 m). Les zones qui diffèrent entre la classification initiale et la post-classification sont encadrées par un rectangle blanc.

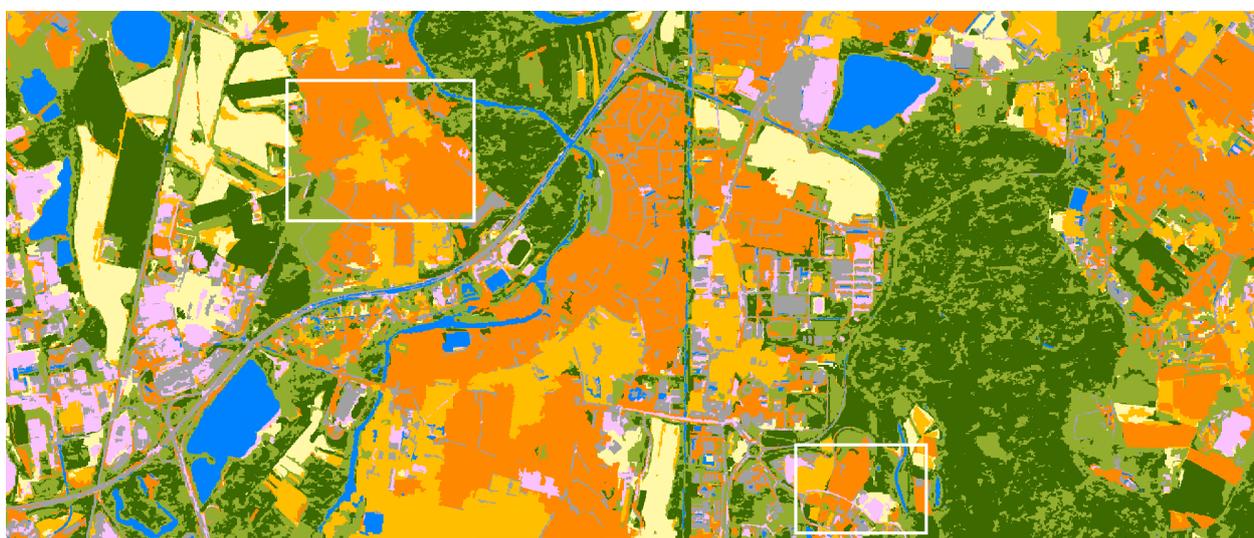
sont particulièrement bien extraites. Les valeurs de F-mesure associées à ces classes sont proches de 0,60 pour tous les jeux de données. En moyenne, on observe une augmentation de 5% des valeurs des indices de qualité Kappa et F-mesure, relativement aux résultats de classification obtenus avec PCSC. En particulier, la classe thématique correspondant aux blocs urbains à fortes densités, est extraite d'une manière plus pertinente avec C-PCSC qu'avec PCSC.

Le processus C-PCSC a été appliqué dans un second temps sur les images MRS, pour classifier les quartiers urbains de l'image. Les résultats de post-classification obtenus (sur les jeux de données **ELSAU** et **ILLKIRCH**) sont présentés par les figures 12.5(a), 12.6(a). En moyenne, on observe ici aussi une augmentation de 6% des valeurs des indices de qualité Kappa et F-mesure, relativement aux résultats de classification obtenus avec PCSC. En particulier, les classes thématiques correspondant aux quartiers résidentiels et aux quartiers spécifiques sont extraites d'une manière plus pertinente avec C-PCSC qu'avec PCSC.

L'analyse de ces résultats de post-classification permet de mettre en évidence l'intérêt d'un post-processus ascendant de classification, pour corriger et remettre en cause les résultats de classification issus du processus descendant PCSC. Ainsi, l'ajout de connaissances (du domaine et liées aux images) au sein de ce processus



(a) Post-classification MRS – niveau des quartiers.



(b) Post-classification HRS – niveau des blocs.

FIGURE 12.6 – Résultats de post-classification pour le jeu de données **ILLKIRCH** (6 576 m × 2 803 m). Les zones qui diffèrent entre la classification initiale et la post-classification sont encadrées par un rectangle blanc.

d'analyse permet d'améliorer les résultats de classification obtenus par l'intermédiaire d'une approche non-supervisée, tout en évitant à l'utilisateur les contraintes liées à l'emploi d'approches de classification supervisée.

Découverte de connaissances multirésolutions à partir des images Nous avons vu précédemment que l'application du processus C-PCSC aux différentes résolutions implique la définition d'histogrammes de composition basés régions, liés à des concepts sémantiques particuliers. Ces structures, nécessaires au déroulement du processus, peuvent aussi être considérées comme des connaissances multirésolutions relatives au contenu des images étudiées (voir la section 9.2). De tels histogrammes peuvent être utiles pour les utilisateurs, afin de mieux comprendre l'organisation hiérarchique et structurelle des environnements urbains.

Dans le cadre de ces expériences, nous présentons en figure 12.7 et en figure 12.8 les histogrammes de composition basés régions (liés au concept sémantique des objets urbains simples) obtenus pour les clusters extraits de l'image \mathcal{I}_C^2 , (*i.e.*, au niveau des blocs urbains) du jeu de données **ELSAU**. Ces histogrammes permettent d'étudier la composition des blocs urbains et de certaines structures urbaines, extraits à partir d'images HRS

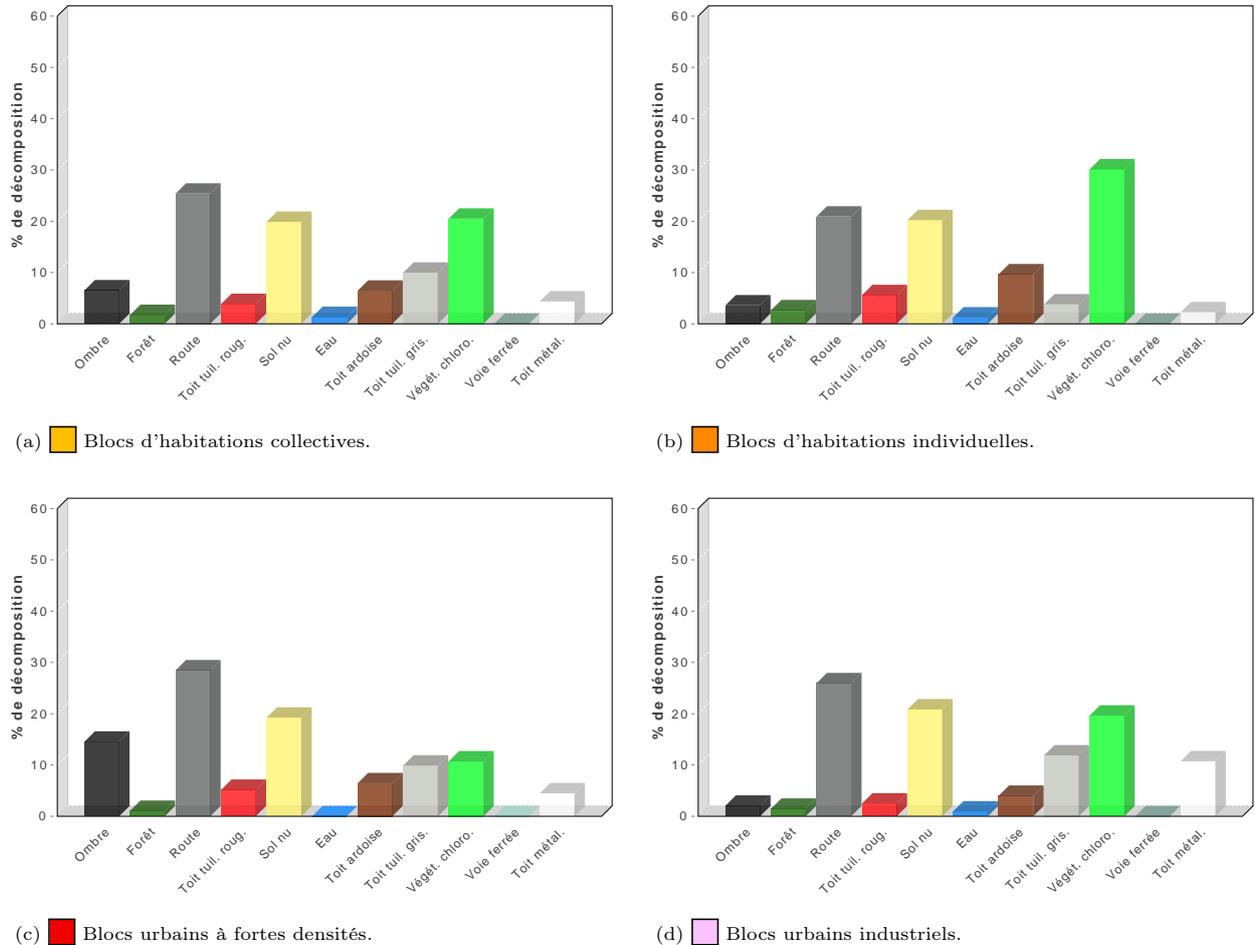


FIGURE 12.7 – Histogrammes de composition basés régions (liés au concept sémantique des objets urbains) obtenus pour les clusters de l'image \mathcal{I}_C^2 (*i.e.*, niveau des blocs urbains) du jeu de données **ELSAU**.

en termes d'objets urbains simples (extraits d'images THRS). Considérons par exemple l'histogramme (b) de la figure 12.7. Cet histogramme représente la composition moyenne des régions de l'image HRS correspondant à des blocs urbains d'habitations individuelles. Considérons maintenant l'histogramme (c) de la figure 12.7. Il représente la composition moyenne des régions de l'image HRS correspondant à des blocs urbains d'habitations à fortes densités. À partir de ces deux structures, on peut remarquer que l'une des différences majeures entre ces deux types d'environnements urbains est liée au fait que les blocs urbains d'habitations à fortes densités sont, en général, composés d'une proportion plus importante de régions correspondant à des ombres portées de bâtiments (*i.e.*, la barre noire est plus haute). Ceci peut s'expliquer par le fait que les bâtiments d'habitats collectifs sont, en général, plus hauts que les bâtiments d'habitats individuels. Le même type de raisonnement peut s'appliquer sur la proportion de zones de végétation (*i.e.*, barre vert clair) : plus la densité d'habitation augmente, plus le nombre de zones de végétation diminue.

Carte des conflits Pour évaluer visuellement les différences de classification entre les résultats de classification obtenus par PCSC et ceux obtenus par C-PCSC, nous présentons deux des cartes de conflits issues de ces expérimentations. De telles cartes permettent d'évaluer les zones de l'image sur lesquelles le processus PCSC avait commis des erreurs de segmentation/classification importantes durant le processus descendant (voir la section 11.2.3). Ces conflits peuvent être mis en évidence *via* la mise en correspondance de la partition issue de la classification de PCSC et celle issue de la post-classification de C-PCSC.

Dans le cadre de ces expériences, nous présentons en figure 12.9 les cartes des conflits obtenues pour la résolution r_2 (*i.e.*, niveau des blocs urbains) sur les jeux de données **ELSAU** et **ILLKIRCH**. On remarque que

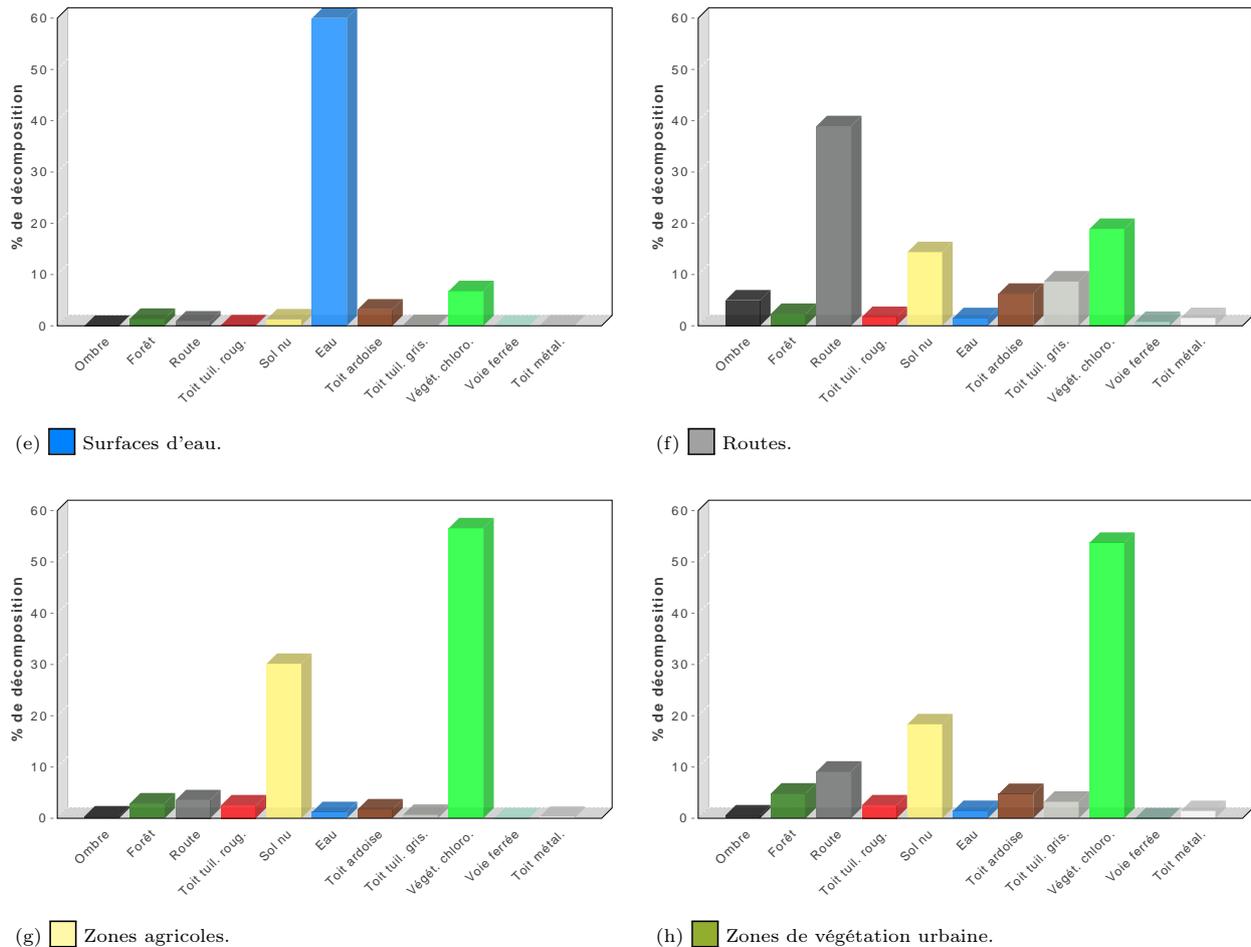


FIGURE 12.8 – Histogrammes de composition basés régions (liés au concept sémantique des objets urbains) obtenus pour les clusters de l'image \mathcal{I}_C^2 (*i.e.*, niveau des blocs urbains) du jeu de données **ELSAU**.

les conflits les plus importants apparaissent au niveau des blocs urbains à fortes densités et d'habitations collectives. L'extraction de ce type de blocs pose certains problèmes de segmentation dus à la complexité et l'hétérogénéité de ces environnements urbains. De plus, la caractérisation de ce type de blocs par le biais d'histogrammes de composition est limitée par l'absence de prise compte de l'organisation spatiale des objets urbains simples composant ces structures urbaines.

Une solution future, pour corriger les labels des régions associées à un conflit, consistera à considérer une approche ascendante itérative permettant la remise en cause et la correction de modèle de segmentation et de classification d'une image à une résolution particulière (voir les perspectives évoquées dans la section 11.3.2).

12.3.3 Robustesse

Les résultats de post-classification obtenus par l'intermédiaire du processus C-PCSC sont fortement dépendants de la pertinence de la distance HSB, utilisée pour comparer les données à classifier. Nous avons vu précédemment que cette distance est basée sur un mécanisme fusionnant les barres des histogrammes les plus corrélées sémantiquement. L'ordre de fusion de ces barres est induit par des valeurs de proximité sémantique, fournies par l'utilisateur (*via* ses connaissances du domaine), par l'intermédiaire d'une matrice de dissimilarité \mathcal{M}^{dis} . En conséquence, la qualité des résultats de post-classification obtenus est dépendante de la pertinence des valeurs de proximité sémantique liées aux connaissances du domaine de l'utilisateur.

Pour évaluer la robustesse du processus C-PCSC (et par la même occasion, la robustesse de la distance HSB) face à ces valeurs de proximité sémantique, nous avons étudié l'impact de perturbations aléatoires,

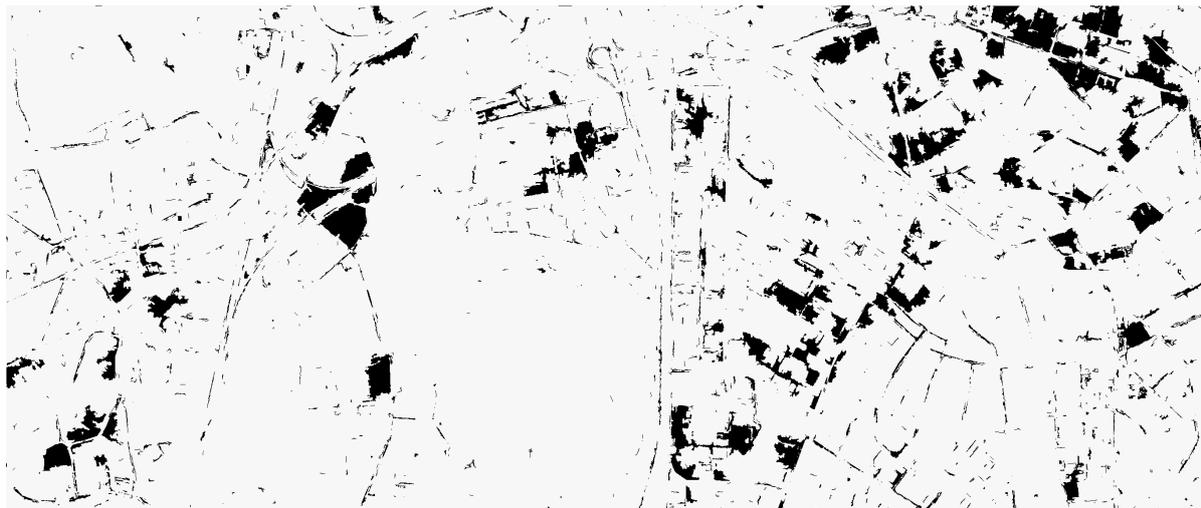
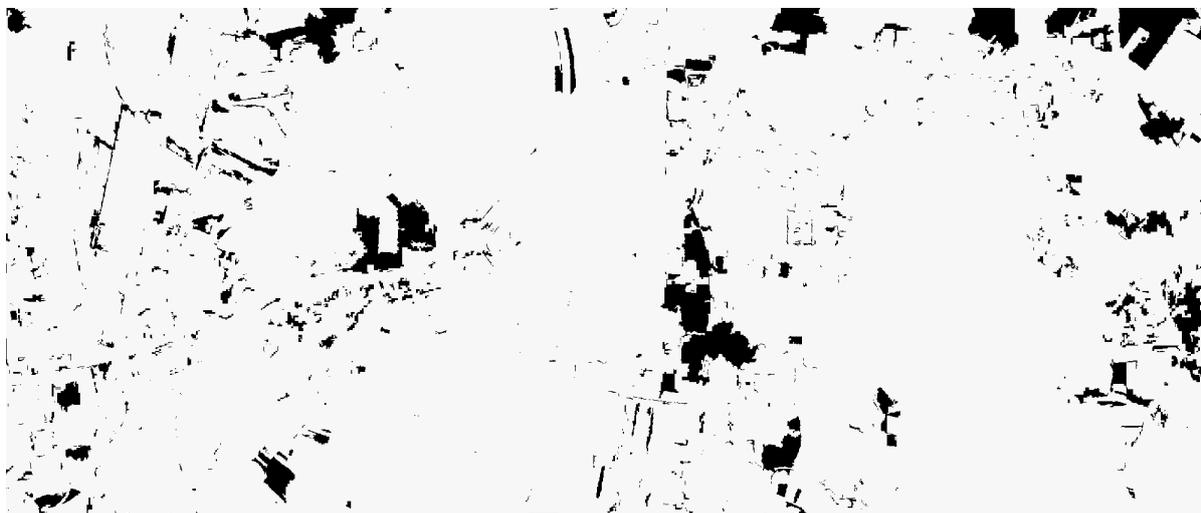
(a) Jeu de données **ELSAU**.(b) Jeu de données **ILLKIRCH**.

FIGURE 12.9 – Cartes des conflits au niveau d’analyse des blocs urbains (à partir d’images HRS). Ces conflits sont issus de la mise en correspondance de la classification obtenue par le processus descendant PCSC avec la classification obtenue par le processus ascendant C-PCSC.

apportées à ces valeurs, sur la qualité globale des résultats de classification. Pour ce faire, les valeurs de proximité sémantique, associées à la matrice \mathcal{M}^{dis} (modélisant les relations sémantiques entre les instances du concept sémantique OBJET URBAIN) ont été modifiées successivement de $x\%$, $x \in \{0, 5, 10, 15, \dots, 100\}$. Chaque matrice résultante a été utilisée par HSBd pour construire l’ordre de fusion des barres. Cette distance a alors été employée par C-PCSC pour la post-classification des régions issues de la segmentation de l’image à la résolution r_2 (pour le jeu de données **ELSAU**). Chaque résultat de classification obtenu a ensuite été comparé à la carte de vérité terrain correspondante *via* les indices d’évaluation \mathcal{K} et $\overline{\mathcal{F}}$.

Dans le cadre de ces expériences, nous présentons en figure 12.10, l’impact des perturbations appliquées aux valeurs de proximité sémantique de la matrice de dissimilarité \mathcal{M}^{dis} , sur la qualité des résultats de post-classification de C-PCSC. On remarque qu’en dessous de 10% de perturbations sur les poids de \mathcal{M}^{dis} , la qualité des résultats de classification n’est que très peu affectée. Au-delà de 15% de perturbations, la qualité des résultats de classification, relativement aux cartes de vérité terrain, se trouve considérablement diminuée.

L’analyse de ces résultats permet de montrer que la distance HSBd (et par conséquent, le post-processus C-PCSC) est relativement robuste aux valeurs de proximité sémantique fournies par l’utilisateur. Dans le

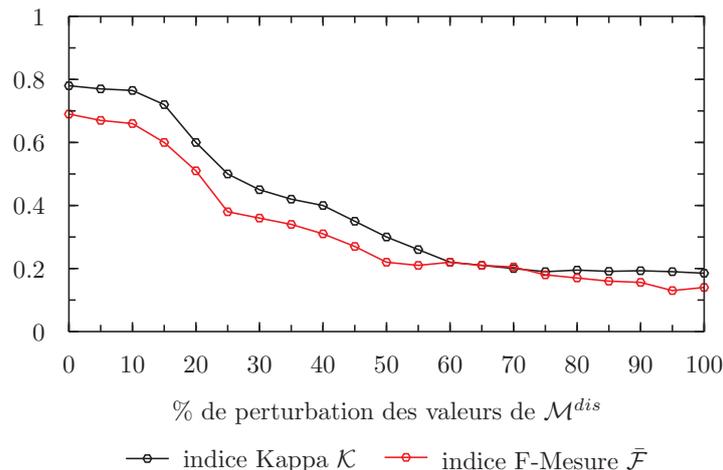


FIGURE 12.10 – Impact des perturbations appliquées sur les valeurs de proximité sémantique de la matrice de dissimilarité \mathcal{M}^{dis} (relative aux instances du concept OBJET URBAIN), sur la qualité des résultats de post-classification de C-PCSC (jeu de données **ELSAU**). L'évolution de l'indice de qualité \mathcal{K} (respectivement l'indice $\bar{\mathcal{F}}$) est représentée en noir (respectivement en rouge).

cas où les connaissances du domaine de ce dernier sont imprécises (en dessous de 10% d'imprécision), la distance HSBD peut être employée pour la classification de données et conduire à des résultats pertinents. Cette assertion peut être expliquée du fait qu'une perturbation dans les valeurs de proximité de la matrice affectera la hauteur des étages et des branches du dendrogramme mais pas nécessairement la structure de ce dernier. Dans le cas où la structure du dendrogramme n'est pas modifiée, l'ordre de fusion des barres des histogrammes ne se trouvera pas modifié.

12.4 Bilan scientifique

12.4.1 Contributions

Ce chapitre a présenté les expériences effectuées dans le cadre de la validation de la distance entre histogrammes HSBD et du processus de post-classification C-PCSC, tous deux proposés dans ces travaux. Ces expérimentations ont montré que les résultats de segmentation et de classification, obtenus par le biais du processus descendant non-supervisée PCSC, peuvent être corrigés et améliorés par un post-processus ascendant, intégrant des connaissances relatives au contenu des images (acquises durant la descente) et des connaissances du domaine, apportées par un utilisateur. De plus, les résultats de ces expériences ont permis de mettre en évidence l'intérêt de la distance HSBD pour la classification de données décrites par des histogrammes, et d'une manière plus générale, l'intérêt de cette dernière pour les approches de fouille de données.

12.4.2 Valorisation scientifique

Ces différentes expérimentations ont donné lieu à une contribution [Kurtz et al., 2012a] dans la conférence internationale de traitement d'images *IEEE ICIP 2012, International Conference on Image Processing* et publiée par *IEEE Signal Processing Society*.

Conclusion

Conclusion

« *Un poème ne se termine pas mais s'abandonne.* » 

— Paul Valéry (1871–1945)

Bilan

Au cours de cette thèse, plusieurs aspects des processus d'analyse d'images par le biais d'approches empruntées au domaine de la fouille de données ont été étudiés. Le point de départ de ces travaux était le développement de nouvelles approches d'analyse d'images basées régions. L'objectif était de proposer des approches permettant d'extraire et de catégoriser des objets d'intérêt géographiques, à plusieurs niveaux d'échelle, capables de tirer parti de la complémentarité des images multisources. Les recherches menées ont notamment permis de lever certains verrous concernant le traitement de données complexes à (très) hautes résolutions spatiales.

Dans le cadre de cette thèse, nous nous sommes intéressés, en particulier, à deux axes de recherche principaux. Le premier a été dédié au développement d'approches basées régions non-supervisées pour l'extraction d'objets d'intérêt complexes à partir d'images multirésolutions. Le second a porté sur l'intégration de connaissances dans le processus d'analyse pour pallier certains problèmes soulevés par l'utilisation d'approches non-supervisées pour le traitement de données complexes.

Contributions

Ces deux axes de recherche ont ainsi donné lieu à différentes contributions dans les domaines de la fouille de données et du traitement d'images. Nous proposons ci-après une synthèse de ces contributions.

Axe 1 : extraction d'objets d'intérêt complexes à partir d'images multirésolutions

La premier axe de recherche de cette thèse a été consacré à l'étude d'un processus d'analyse basé régions capable de traiter des ensembles d'images multirésolutions. Dans un premier temps, nous nous sommes intéressés à définir comment un algorithme de segmentation hiérarchique pouvait être combiné à un algorithme de classification pour analyser des images multirésolutions. Nous avons proposé un processus collaboratif générique basé sur un mécanisme segmentation–classification, capable d'extraire différentes catégories d'objets d'intérêt, relatifs à différents niveaux d'analyse, à partir des différentes résolutions disponibles. L'originalité principale de ce processus est d'opérer de façon descendante à travers la résolution. Des verrous relatifs à la mise en correspondance de régions, extraites de différentes images à des résolutions spatiales différentes, ont ainsi été levés.

Pour instancier un tel processus, nous avons dans un deuxième temps, défini une approche de segmentation basée sur des concepts issus de la morphologie mathématique, adaptée à l'extraction de structures urbaines complexes. Pour ce faire nous avons proposé une approche de segmentation hiérarchique reposant sur la notion d'arbres binaires de partitions et d'exemples de coupes. La principale originalité de cette dernière est de pouvoir traiter des images à (très) hautes résolutions spatiales en s'abstrayant des problèmes d'échelles. Nous nous sommes ensuite intéressés à la classification non-supervisée des segments ainsi produits. Nous avons notamment proposé une approche originale pour la caractérisation des segments à classifier, reposant sur la notion d'attributs multirésolutions, modélisés sous la forme d'histogrammes de composition.

Axe 2 : intégration de connaissances dans le processus d'analyse

Bien que la propriété descendante du processus défini précédemment présente de nombreux avantages, elle induit aussi un inconvénient principal : les erreurs de segmentation et de classification commises durant le traitement d'une résolution particulière sont propagées dans les résolutions suivantes. Des études menées ces dernières années ont montré qu'une solution potentielle pour pallier ces problèmes consiste à intégrer, dans le processus d'analyse, des connaissances du domaine fournies par les experts. Ces connaissances doivent permettre de guider les algorithmes vers des solutions plus pertinentes. Dans ce second axe de recherche, nous nous sommes ainsi consacrés à l'étude de mécanismes permettant d'intégrer des connaissances de haut niveau pour la remise en cause des résultats de segmentation et de classification, à travers la résolution. Nous avons alors proposé un post-processus générique, permettant de corriger les résultats de segmentation et de classification obtenus précédemment, pour chaque niveau de résolution. Ce dernier permet de remettre en cause les résultats d'analyse par le biais de consensus multirésolutions en adoptant une stratégie ascendante à travers la résolution.

Pour instancier un tel processus, il est nécessaire de définir quels types de connaissances utiliser et comment intégrer ces dernières dans les algorithmes de segmentation et de classification. Nous avons été amenés à définir une nouvelle distance pour comparer les données à classifier décrites par des attributs multirésolutions, modélisés sous la forme d'histogrammes de composition. Il est à noter que le problème de la distance entre histogrammes est un problème qui n'est pas spécifique à l'analyse d'images. De ce fait, la définition de cette distance, basée sur les connaissances du domaine de l'utilisateur, est une contribution plus large au domaine de la fouille de données. D'une manière plus générale, nous avons montré qu'il était possible d'appliquer un post-processus correcteur ascendant, basé sur des connaissances du domaine, pour corriger les résultats de segmentation et de classification fournis par le processus collaboratif descendant défini précédemment.

Perspectives de recherche

Les travaux théoriques et les développements logiciels présentés dans cette thèse ont également mis en évidence certaines perspectives de recherche.

Perspectives à court terme

Axe 1 : extraction d'objets d'intérêt complexes à partir d'images multirésolutions

Les recherches effectuées dans le cadre de l'extraction d'objets d'intérêt complexes à partir d'images multirésolutions, et en particulier celles relatives au développement de l'approche de segmentation hiérarchique (basée sur des exemples) ont ouvert de nouvelles voies à étudier pour pallier le problème de l'échelle, lié à l'analyse d'images (T)HRS. En effet, les travaux effectués dans cette thèse se sont principalement intéressés aux algorithmes de segmentation par arbres binaires de partitions. Cependant, comme présenté dans le chapitre 7, d'autres approches de segmentation hiérarchique peuvent être utilisées. Ainsi, une perspective à court terme serait d'étudier l'influence du choix d'un autre algorithme sur les résultats de classification obtenus. De plus, il pourrait être pertinent de réfléchir à de nouveaux moyens d'apprendre et de modéliser les exemples de coupes, interactivement définis par l'utilisateur dans le cadre de l'algorithme d'apprentissage.

Les travaux effectués sur l'approche de classification multirésolution, et en particulier ceux relatifs à l'approche de caractérisation par le biais d'attributs multirésolutions ont ouvert de nouvelles directions à explorer. Dans le cadre de ces travaux, nous nous sommes focalisés sur la caractérisation de ces données par le biais d'histogrammes de composition auxquels des connaissances sémantiques ont été associées. Ces histogrammes permettent de modéliser la composition de segments extraits à une résolution τ_t en termes de labels de classes dans une résolution τ_{t+1} . Cependant, de telles structures ne permettent pas de prendre en compte les relations spatiales structurant les classes issues des images aux résolutions τ_{t+1} . Une direction de recherche prometteuse pourrait consister à modéliser ces informations par le biais de chaînes de Markov, permettant de prendre en compte les « relations de passage » d'une classe à une autre.

Indépendamment, il serait intéressant d'envisager la parallélisation du processus PCSC sur une grille de machines. En effet, l'architecture de PCSC est bien adaptée à la division des traitements, et une telle approche permettrait un passage à l'échelle sur des données volumineuses.

Axe 2 : intégration de connaissances dans le processus d'analyse

Les travaux effectués dans le cadre de l'axe de recherche relatif à l'intégration de connaissances dans le processus d'analyse d'images ont soulevé différentes pistes de recherche.

Le post-processus de remise en cause proposé a ouvert de nouvelles voies à étudier pour apporter des solutions à la correction de résultats multinationaux, issus d'un processus d'analyse descendante basé régions. En particulier, des réflexions doivent être menées pour définir un moyen de mettre en place le processus itératif de remise en cause des résultats de segmentation aux différentes résolutions, proposé dans les perspectives du chapitre 10. En conséquence, l'un des verrous scientifiques importants qui reste à lever est la définition de critères permettant de contrôler ce processus itératif de correction des résultats. L'utilisation d'ontologies multirésolutions semble une solution potentielle pour modéliser des connaissances du domaine permettant de contrôler l'arrêt de ce processus itératif. Une réflexion plus générale serait ainsi à mener sur le rôle des connaissances du domaine dans le processus global d'extraction d'informations. Ce type d'intégration est de plus en plus courant, mais peu d'études présentent des avancées facilement généralisables. Nos travaux se sont intéressés à l'intégration de connaissances lors de l'étape de classification non-supervisée mais d'autres étapes du processus d'analyse pourraient être améliorées *via* ces connaissances.

Il serait également intéressant d'étudier de manière plus approfondie la distance proposée pour la comparaison d'histogrammes. Ainsi, une perspective à court terme serait d'étudier l'influence du choix de la distance barre-à-barre choisie sur les résultats de classification obtenus. De plus, il pourrait être pertinent de réfléchir à des moyens permettant d'aider l'utilisateur à déterminer les valeurs de proximité sémantique, nécessaires à la construction de l'ordre de fusion. Une idée réside dans le fait d'inférer ces valeurs à partir des données par le biais d'approches de classification semi-supervisée.

Perspectives à plus long terme

Finalement, la modélisation et l'utilisation de connaissances dans les processus d'interprétation d'images satellites ont ouvert de nouvelles perspectives de recherche. L'utilisation de connaissances spatiales (*e.g.*, des règles de fusion) ou encore de connaissances multirésolutions sont notamment à l'étude pour améliorer la classification des objets géographiques. Des travaux ont débuté récemment au sein de notre équipe pour représenter ces connaissances sous la forme de règles basées sur la logique floue et d'ontologies du domaine. Il serait intéressant, à plus long terme, d'étudier comment ces contraintes et ces connaissances pourraient être intégrées dans un processus d'analyse d'images basé régions, autorisant une remise en cause et une correction des résultats obtenus par l'application d'approches non-supervisées.

Parallèlement, l'extension des approches d'analyse basées régions au traitement d'images multisources (en particulier multirésolutions) a offert de nombreuses voies à explorer. Un mécanisme collaboratif de construction des régions des images pourrait être mis en place afin de pallier au problème de mise en correspondance des données à classifier. La prise en compte de l'information apportée par la complémentarité des images étudiées et des connaissances multirésolutions évoquées précédemment nous semble être un point de départ potentiel à ces recherches.

Pérennisation des travaux

En conclusion, bien que ces travaux de recherche aient levé un certain nombre de verrous et ouvert plusieurs pistes de recherche, il reste à notre avis de nombreux travaux à accomplir avant d'arriver à un processus d'analyse d'images multinationaux, capable à la fois de traiter toutes les données disponibles (quelque soit leurs spécificités), et à la fois de maximiser l'apport de l'intégration des connaissances disponibles. Pour tenter de lever certains des verrous énoncés dans ces perspectives, un sujet de thèse a récemment été proposé dans notre équipe de recherche. La problématique de la modélisation des connaissances multirésolutions, conjointe aux approches d'analyse d'images basées régions multinationaux y est à l'étude. Ces travaux peuvent être considérés comme le début d'une longue réflexion, nécessaire à l'élaboration d'un tel processus global d'analyse d'images.

Bibliographie

- Akçay, H. G. et Aksoy, S. (2008). Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2097–2111. (Cit  en pages 36, 61, 64, 103 et 104.)
- Akçay, H. G., Aksoy, S., et Soille, P. (2010). Hierarchical segmentation of complex structures. In *Proceedings of the International Conference on Pattern Recognition – ICPR 2010*, pages 1120–1123. IEEE society. (Cit  en page 36.)
- Aksoy, S. et Cinbis, R. G. (2010). Image mining using directional spatial constraints. *IEEE Geoscience and Remote Sensing Letters*, 7(1):33–37. (Cit  en page 33.)
- Aksoy, S., Cinbis, R. G., et Akçay, H. G. (2011). *Signal and image processing for remote sensing*, chapter V – Image classification and object detection using spatial contextual constraints, pages 128–149. CRC Press. (Cit  en page 26.)
- Al-Omari, F. A. et Al-Jarrah, M. A. (2005). Query by image and video content: A colored-based stochastic model approach. *Data & Knowledge Engineering*, 52(3):313–332. (Cit  en page 122.)
- Alpaydin, E. (2004). *Introduction to Machine Learning*. MIT Press. (Cit  en page 12.)
- Athanasiadis, T., Mylonas, P., Avrithis, Y., et Kollias, S. (2007). Semantic image segmentation and object labeling. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(3):298–312. (Cit  en page 28.)
- Baatz, M. et Schape, A. (2000). Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. In *Proceedings of the Angewandte Geographische Informations – AGIT XII*, number 3, pages 12–23. (Cit  en page 36.)
- Bahirat, K., Bovolo, F., Bruzzone, L., et Chaudhuri, S. (2012). A novel domain adaptation Bayesian classifier for updating land-cover maps with class differences in source and target domains. *IEEE Transactions on Geoscience and Remote Sensing*, (In Press). (Cit  en page 105.)
- Baltsavias, E. P. (2004). Object extraction and revision by image analysis using existing geodata and knowledge: Current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(4):129–151. (Cit  en page 114.)
- Barnsley, M. J. et Barr, S. L. (1997). Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system. *Computers, Environment and Urban Systems*, 21(3):209–225. (Cit  en page 33.)
- Baudry, J. P. (2010). *S lection de mod le pour la classification non supervis e. Choix du nombre de classes*. Th se de Doctorat, Universit  Paris Sud - Paris XI (France). (Cit  en page 14.)
- Beaulieu, J. M. et Goldberg, M. (1989). Hierarchy in picture segmentation: A stepwise optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):150–163. (Cit  en page 36.)
- Belarte, B. et Wemmert, C. (2012). Apport des relations spatiales dans l’extraction automatique d’informations   partir d’images. In *Proceedings of the Atelier Fouille de Donn es Complexes, Journ es Francophones Extraction et Gestion des Connaissances (EGC 2012) – FDC 2012*, pages 143–152. (Cit  en page 34.)
- Benediktsson, J. A. et Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1367–1377. (Cit  en page 44.)

- Benediktsson, J. A., Palmason, J. A., et Sveinsson, J. R. (2005). Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491. (Cité en page 35.)
- Benediktsson, J. A., Pesaresi, M., et Amason, K. (2003). Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949. (Cité en page 35.)
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., et Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3):239–258. (Cité en pages 24, 25 et 29.)
- Bhanu, B., Sungkee, L., et Das, S. (1995). Adaptive image segmentation using genetic and hybrid search methods. *IEEE Transactions on Aerospace and Electronic Systems*, 31(4):1268–1291. (Cité en page 26.)
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2–16. (Cité en pages 4 et 17.)
- Bressan, M., Guillet, D., et Vitrià, J. (2003). Using an ICA representation of local color histograms for object recognition. *Pattern Recognition*, 36(3):691–701. (Cité en page 122.)
- Brunelli, R. et Mich, O. (2001). Histograms analysis for image retrieval. *Pattern Recognition*, 34(8):1625–1637. (Cité en page 122.)
- Bruzzone, L. et Carlin, L. (2006). A multilevel context-based system for classification of Very High Spatial Resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2587–2600. (Cité en pages 26, 37 et 191.)
- Bruzzone, L., Cossu, R., et Vernazza, G. (2002). Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images. *Image fusion*, 3(1):289–297. (Cité en page 45.)
- Bruzzone, L., Prieto, D., et Serpico, S. (1999). A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1350–1359. (Cité en page 105.)
- Burge, W. H. (1981). A correspondence between partitions related to generalizations of the Rogers-Ramanujan identities. *Discrete Mathematics*, 34(1):9–15. (Cité en page 142.)
- Capdevila, M. et Florez, O. W. M. (2009). A communication perspective on automatic text categorization. *IEEE Transactions on Knowledge and Data Engineering*, 21(7):1027–1041. (Cité en page 122.)
- Carleer, A. P., Debeir, O., et Wolff, E. (2005). Assessment of Very High Spatial Resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, 71(11):1285–1294. (Cité en page 24.)
- Carleer, A. P. et Wolff, E. (2006). Urban land cover multilevel region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, 27(6):1035–1051. (Cité en page 26.)
- Carlson, T. N. et Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62(3):241–252. (Cité en page 16.)
- Cha, S. H. (2008). Taxonomy of nominal type histogram distance measures. In *Proceedings of the American Conference on Applied Mathematics – ACAM 2008*, pages 325–330. World Scientific and Engineering Academy and Society. (Cité en page 126.)
- Cha, S. H. et Srihari, S. N. (2002). On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355–1370. (Cité en pages 92, 122, 125 et 127.)

- Chang, Y., Liang, L., Han, C., et Fang, Y. (2007). Multisource data fusion for landslide classification using generalized positive Boolean functions. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1697–1710. (Cit  en page 43.)
- Chen, Q., Zhou, C., Luo, J., et Ming, D. (2005). Fast segmentation of High-Resolution satellite images using watershed transform combined with an efficient region merging approach. In Klette, R. and Zuni, J., editor, *Proceedings of the International Workshop on Combinatorial Image Analysis – IWCIA 2005*, volume 3322 of *Lecture Notes in Computer Science*, pages 621–630. Springer. (Cit  en page 24.)
- Chibani, Y. (2005). Selective synthetic aperture Radar and Panchromatic image fusion by using the   trous wavelet decomposition. *EURASIP Journal on Applied Signal Processing*, 34(14):2207–2214. (Cit  en page 43.)
- Collet, C., Caloz, R., et Rochon, G. (2001). *Pr cis de t l d tection : traitements num riques d’images de t l d tection*, volume 3. Puq. (Cit  en pages 10 et 86.)
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46. (Cit  en page 96.)
- Cornu jols, A. et Miclet, L. (2010). *Apprentissage Artificiel. Concepts et Algorithmes*. Eyrolles. (Cit  en page 11.)
- Couprie, C., Grady, L. J., Najman, L., et Talbot, H. (2011). Power watershed: A unifying graph-based optimization framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1384–1399. (Cit  en page 24.)
- Cousty, J., Bertrand, G., Couprie, M., et Najman, L. (2008). Fusion graphs: Merging properties and watersheds. *Journal of Mathematical Imaging and Vision*, 30(1):87–104. (Cit  en page 37.)
- Cousty, J., Bertrand, G., Najman, L., et Couprie, M. (2009). Watershed cuts: Minimum spanning forests and the drop of water principle. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1362–1374. (Cit  en page 37.)
- Crespo, J., Schafer, R., Serra, J., Gratin, C., et Meyer, F. (1997). The flat zone approach: A general low-level region merging segmentation method. *Signal Processing*, 62(1):37–60. (Cit  en page 24.)
- Cross, A. M., Mason, D. C., et Dury, S. J. (1988). Segmentation of remotely-sensed images by a split-and-merge process. *International Journal of Remote Sensing*, 9(8):1329–1345. (Cit  en page 24.)
- Daum  III, H. et Marcu, D. (2006). Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26(1):101–126. (Cit  en page 105.)
- de Lussy, F., Kubik, P., Greslou, D., Pascal, V., Gigord, P., et Cantou, J. (2005). PLEIADES-HR image system products and geometric accuracy. In *Proceedings of the ISPRS International Society for Photogrammetry and Remote Sensing Congress – ISPRS 2005*, volume 1, pages 50–57. (Cit  en page 86.)
- Derivaux, S., Forestier, G., Wemmert, C., et Lef vre, S. (2010). Supervised image segmentation using watershed transform, fuzzy classification and evolutionary computation. *Pattern Recognition Letters*, 31(15):2364–2374. (Cit  en pages 26 et 46.)
- Dogrusoz, E. et Aksoy, S. (2007). Modeling urban structures using graph-based spatial patterns. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2007*, pages 4826–4829. IEEE Geoscience and Remote Sensing Society. (Cit  en page 33.)
- Donoghue, D. N. M. (2000). Remote sensing: Sensors and applications. *Progress in Physical Geography*, 24(3):407–414. (Cit  en page 30.)

- Dou, W., Chen, Y., Li, X., et Sui, D. Z. (2007). A general framework for component substitution image fusion: An implementation using the fast image fusion method. *Computers & Geosciences*, 33(2):219–228. (Cité en page 44.)
- Dubuisson, M. et Jain, A. (1995). Contour extraction of moving objects in complex outdoor scenes. *International Journal of Computer Vision*, 14(1):83–105. (Cité en page 44.)
- Fabrizio, S. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, (34):1–47. (Cité en page 122.)
- Felzenszwalb, P. et Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59:167–181. (Cité en page 59.)
- Forestier, G. (2010). *Connaissances et clustering collaboratif d'objets complexes multisources*. Thèse de Doctorat, Université de Strasbourg (France). (Cité en pages 14, 25, 26, 27, 45 et 114.)
- Forestier, G., Puissant, A., Wemmert, C., et Gañçarski, P. (2012). Knowledge-based region labeling for remote sensing image interpretation. *Computers, Environment and Urban Systems*, (In Press). (Cité en page 28.)
- Forestier, G., Wemmert, C., et Gañçarski, P. (2008). Multi-source images analysis using collaborative clustering. *EURASIP Journal on Advances in Signal Processing*, 2008(1):1542–1553. (Cité en pages 26, 45 et 72.)
- Frawley, W. J., Piatetsky-Shapiro, G., et Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, 13(3):3–57. (Cité en page 10.)
- Gaetano, R., Scarpa, G., et Poggi, G. (2009). Hierarchical texture-based segmentation of multiresolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2129–2141. (Cité en pages 34, 36, 53, 103, 104, 115 et 117.)
- Gañçarski, P. et Blansché, A. (2008). Darwinian, Lamarckian, and Baldwinian (co)evolutionary approaches for feature weighting in K-means-based algorithms. *IEEE Transactions on Evolutionary Computation*, 12(5):617–629. (Cité en page 15.)
- Gañçarski, P., Blansché, A., et Wania, A. (2008). Comparison between two coevolutionary feature weighting algorithms in clustering. *Pattern Recognition*, 41(3):983–994. (Cité en page 15.)
- Gañçarski, P. et Wemmert, C. (2007). Collaborative multi-step mono-level multi-strategy classification. *Journal on Multimedia Tools and Applications*, 35(1):1–27. (Cité en page 14.)
- Garrido, L., Salembier, P., et Garcia, D. (1998). Extensive operators in partition lattices for image sequence analysis. *Signal Processing*, 66(2):157–180. (Cité en page 58.)
- Germain, M., Voorons, M., Boucher, J., Béné, G. B., et Beaudry, E. (2004). Multisource image fusion algorithm based on a new evidential reasoning approach. In *Proceedings of the ISPRS International Society for Photogrammetry and Remote Sensing Congress – ISPRS 2004*, volume 20, pages 129–138. (Cité en page 44.)
- Goffe, R., Brun, L., et Damiand, G. (2011). Tiled top-down combinatorial pyramids for large images representation. *International Journal of Imaging Systems and Technology*, 21(1):28–36. (Cité en pages 35, 37, 54 et 107.)
- Gonzalez, R. C. et Woods, R. E. (2002). *Digital Image Processing*. Prentice Hall Press. (Cité en pages 11 et 20.)
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human Computer Studies*, 43(5):907–928. (Cité en page 27.)

- Gueguen, L., Le Men, C., et Datcu, M. (2006). Analysis of satellite image time series based on information bottleneck. In *Proceedings of the International Conference on Bayesian Inference and Maximum Entropy Methods In Science and Engineering – ICBIM 2006*, pages 367–374. AIP. (Cité en page 189.)
- Gueguen, L., Pesaresi, M., et Soille, P. (2011). An interactive image mining tool handling gigapixel images. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2011*, pages 1581–1584. IEEE Geoscience and Remote Sensing Society. (Cité en page 30.)
- Gueguen, L., Soille, P., et Pesaresi, M. (2010). Differential morphological decomposition segmentation: A multi-scale object based image description. In *Proceedings of the International Conference on Pattern Recognition – ICPR 2010*, pages 938–941. IEEE society. (Cité en page 35.)
- Guigues, L., Le Men, H., et Cocquerez, J. P. (2003). The hierarchy of the cocoons of a graph and its application to image segmentation. *Pattern Recognition Letters*, 24(8):1059–1066. (Cité en page 33.)
- Guo, D., Xiong, H., Atluri, V., et Adam, N. R. (2009). Object discovery in High-Resolution remote sensing images: A semantic perspective. *Knowledge and Information Systems*, 19(2):211–233. (Cité en page 33.)
- Gupta, L., Molfese, D. L., Tammana, R., et Simos, P. G. (1996). Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Transactions on Biomedical Engineering*, 43(4):348–356. (Cité en page 139.)
- Hadamard, J. (1902). *Sur les Problèmes aux Dérivées Partielles et leur Signification Physique*. Princeton University Press. (Cité en page 22.)
- Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M., et Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736. (Cité en page 126.)
- Herold, M., Liu, X. H., et Clarke, K. C. (2003). Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing*, 69(9):991–1001. (Cité en page 20.)
- Hubert, L. et Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2:193–218. (Cité en page 142.)
- Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63. (Cité en page 44.)
- Inglada, J. (2007). Automatic recognition of man-made objects in High Resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(3):236–248. (Cité en page 103.)
- Inglada, J. et Michel, J. (2009). Qualitative spatial reasoning for High-Resolution remote sensing image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):599–612. (Cité en page 34.)
- Jähne, B. (2005). *Digital Image Processing*. Springer. (Cité en page 16.)
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666. (Cité en page 13.)
- Jain, A. K. et Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall. (Cité en page 14.)
- Jain, A. K., Murty, M. N., et Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323. (Cité en pages 13 et 122.)
- Jasani, B., Pesaresi, M., Schneiderbauer, S., et Zeug, G., editors (2009). *Remote Sensing from Space: Supporting International Peace and Security*, volume 297 of *Studies in Remote Sensing*. Springer. (Cité en page 16.)
- Jensen, J. R. et Lulla, K. (1987). Introductory digital image processing: A remote sensing perspective. *Geocarto International*, 2(1):65–65. (Cité en page 15.)

- Kim, J. B. et Kim, H. J. (2003). Multiresolution-based watersheds for efficient image segmentation. *Pattern Recognition Letters*, 24(1–3):473–488. (Cité en page 44.)
- Kittler, J. et Foglein, J. (1984). Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2(1):13–29. (Cité en page 17.)
- Kopanas, I., Avouris, N. M., et Daskalaki, S. (2002). The role of domain knowledge in a large scale data mining project. In Vlahavas, I. et Spyropoulos, C., editors, *Methods and Applications of Artificial Intelligence*, volume 2308 of *Lecture Notes in Computer Science*, pages 746–746. Springer. (Cité en page 114.)
- Lee, J. A. et Verleysen, M. (2007). *Nonlinear Dimensionality Reduction*. Springer. (Cité en page 74.)
- Levina, E. et Bickel, P. (2001). The Earth Mover’s Distance is the Mallows distance: Some insights from statistics. In *Proceedings of the IEEE International Conference on Computer Vision – ICCV 2001*, volume 2, pages 251–256. IEEE Society. (Cité en page 128.)
- Li, Z. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13(1):25–50. (Cité en page 17.)
- Ling, H. et Okada, K. (2007). An Efficient Earth Mover’s Distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853. (Cité en page 127.)
- Liu, Y., Zhang, D., Lu, G., et Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282. (Cité en page 122.)
- Ma, Y., Gu, X., et Wang, Y. (2010). Histogram similarity measure using variable bin size distance. *Computer Vision and Image Understanding*, 114(8):981–989. (Cité en page 128.)
- Macqueen, J. B. (1967). Some methods of classification and analysis of multivariate observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability – BSMSP 1967*, volume 1, pages 281–297. (Cité en page 13.)
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693. (Cité en page 44.)
- Mallat, S. G. (1996). Wavelets for a vision. *Proceedings of the IEEE*, 84(4):604–614. (Cité en pages 44 et 54.)
- Mallows, C. L. (1972). A note on asymptotic joint normality. *Annals of Mathematical Statistics*, 43(2):508–515. (Cité en page 128.)
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press. (Cité en page 33.)
- McGuinness, K. et O’Connor, N. E. (2010). A comparative evaluation of interactive segmentation algorithms. *Pattern Recognition*, 43(2):434–444. (Cité en pages 61 et 93.)
- Meinel, G. et Neubert, M. (2004). A comparison of segmentation programs for High Resolution remote sensing data. In *Proceedings of the ISPRS International Society for Photogrammetry and Remote Sensing Congress – ISPRS 2004*, volume 35, pages 1097–1105. (Cité en page 24.)
- Meyer, F. et Petros, M. (1999). Morphological scale-space representation with levelings. In Nielsen, M., Johansen, P., Olsen, O., et Weickert, J., editors, *Proceedings of the International Conference on Scale-Space Theories in Computer Vision – SSTCV 1999*, volume 1682 of *Lecture Notes in Computer Science*, pages 187–198. Springer. (Cité en page 25.)
- Miller, R. et Small, C. (2003). Cities from space: Potential applications of remote sensing in urban environmental research and policy. *Environmental Science & Policy*, 6(2):129–137. (Cité en page 40.)
- Monasse, P. et Guichard, F. (2000). Scale-space from a level lines tree. *Journal of Visual Communication and Image Representation*, 11(2):224–236. (Cité en page 37.)

- Moscheni, F., Bhattacharjee, S., et Kunt, M. (1998). Spatio-temporal segmentation based on region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):897–915. (Cité en page 189.)
- Mueller, M., Segl, K., et Kaufmann, H. (2004). Edge- and region-based segmentation technique for the extraction of large, man-made objects in High-Resolution satellite imagery. *Pattern Recognition*, 37(8):1619–1628. (Cité en page 24.)
- Nagao, N., Matsuyama, T., et Ikeda, Y. (1979). Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, 10(3):195–223. (Cité en page 25.)
- Najman, L. (2011). On the equivalence between hierarchical segmentations and ultrametric watersheds. *Journal of Mathematical Imaging and Vision*, 40(3):231–247. (Cité en page 37.)
- Nascimento, M. A., Tousidou, E., Chitkara, V., et Manolopoulos, Y. (2002). Image indexing and retrieval using signature trees. *Data & Knowledge Engineering*, 43(1):57–77. (Cité en page 122.)
- Ng, E. K. K., Fu, A. W. C., et Wong, R. C. W. (2005). Projective clustering by histograms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):369–383. (Cité en page 122.)
- Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D., Yanker, P., Faloutsos, C., et Taubin, G. (1993). QBIC project: Querying Images By content, Using Color, Texture and Shape. In *Proceedings of the International Conference on Storage and Retrieval for Image and Video Databases – ICSRIVD 1993*, number 1, pages 173–187. (Cité en page 126.)
- Ogier, J. M., Mullot, R., Labiche, J., et Lecourtier, Y. (2000). Semantic coherency: The basis of an image interpretation device-application to the cadastral map interpretation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(2):322–338. (Cité en page 28.)
- Özdemir, B., Aksoy, S., Eckert, S., Pesaresi, M., et Ehrlich, D. (2010). Performance measures for object detection evaluation. *Pattern Recognition Letters*, 31(10):1128–1137. (Cité en pages 103 et 104.)
- Pal, N. R. et Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294. (Cité en page 22.)
- Pal, S. K., Ghosh, A., et Shankar, B. U. (2000). Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. *International Journal of Remote Sensing*, 21(11):2269–2300. (Cité en page 23.)
- Passat, N., Naegel, B., Rousseau, F., Koob, M., et Dietemann, J. L. (2011). Interactive segmentation based on component-trees. *Pattern Recognition*, 44(10–11):2539–2554. (Cité en page 61.)
- Pavlidis, T. (1977). *Structural Pattern Recognition*. Springer. (Cité en page 20.)
- Pesaresi, M. et Benediktsson, J. A. (2001). A new approach for the morphological segmentation of High-Resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):309–320. (Cité en page 35.)
- Petitjean, F., Ketterlin, A., et Gançarski, P. (2011). A global averaging method for Dynamic Time Warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693. (Cité en pages 93, 127 et 139.)
- Pietikainen, M. et Rosenfeld, A. (1981). Image segmentation by texture using pyramid node linking. *IEEE Transactions on Systems, Man and Cybernetics*, 11(12):822–825. (Cité en page 35.)
- Pignalberi, G., Cucchiara, R., Cinque, L., et Levialdi, S. (2003). Tuning range image segmentation by genetic algorithm. *EURASIP Journal on Applied Signal Processing*, 2003(1):780–790. (Cité en page 26.)
- Plaza, A. J. et Tilton, J. C. (2005). Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2005*, volume 7, pages 4946–4949. IEEE Geoscience and Remote Sensing Society. (Cité en page 61.)

- Puissant, A. (2003). *Information géographique et image à Très Haute Résolution*. Thèse de Doctorat, Université Strasbourg 1 (France). (Cit  en pages 4, 26 et 30.)
- Puissant, A. et Weber, C. (2002). The utility of Very High Spatial Resolution images to identify urban objects. *Geocarto International*, 17(1):33–44. (Cit  en page 17.)
- Rath, T. M. et Manmatha, R. (2003). Word image matching using Dynamic Time Warping. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition – CVPR 2003*, volume 2, pages 521–527. IEEE Computer Society. (Cit  en page 127.)
- Rezaee, M. R., van der Zwet, P. M. J., Lelieveldt, B. P. E., van der Geest, R. J., et Reiber, J. H. C. (2000). A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. *IEEE Transactions on Image Processing*, 9(7):1238–1248. (Cit  en page 44.)
- Richards, J. A. (1986). *Remote Sensing: Digital Image Analysis*. Springer. (Cit  en page 16.)
- Roelfsema, P. R., Lamme, V. A. F., et Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395(6700):376–381. (Cit  en page 17.)
- Rubner, Y., Puzicha, J., Tomasi, C., et Buhmann, J. M. (2001). Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, 84(1):25–43. (Cit  en page 127.)
- Rubner, Y., Tomasi, C., et Guibas, L. J. (2000). The Earth Mover’s Distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121. (Cit  en page 127.)
- Sakoe, H. et Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49. (Cit  en page 92.)
- Salembier, P. et Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, 9(4):561–576. (Cit  en pages 37, 56 et 57.)
- Salembier, P., Oliveras, A., et Garrido, L. (1998). Antiextensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing*, 7(4):555–570. (Cit  en page 37.)
- Salembier, P. et Wilkinson, M. H. F. (2009). Connected operators: A review of region-based morphological image processing techniques. *IEEE Signal Processing Magazine*, 26(6):136–157. (Cit  en page 37.)
- Scarpa, G., Haindl, M., et Zerubia, J. (2007). A hierarchical finite-state model for texture segmentation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP 2007*, volume 1, pages 1209–1212. IEEE Signal Processing Society. (Cit  en page 36.)
- Scheunders, P. et Sijbers, J. (2002). Multiscale watershed segmentation of multivalued images. In *Proceedings of the International Conference on Pattern Recognition – ICPR 2002*, volume 3, pages 855–858. IEEE society. (Cit  en page 44.)
- Sch olkopf, B. et Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press. (Cit  en page 12.)
- Serra, J. (2011). Hierarchies and optima. In *Proceedings of the International Symposium on Discrete Geometry for Computer Imagery – DGCI 2011*, volume 6607 of *Lecture Notes in Computer Science*, pages 35–46. Springer. (Cit  en page 61.)
- Serra, J. C. et Salembier, P. (1993). Connected operators and pyramids. In Dougherty, E. R., Gader, P. D., et Serra, J. C., editors, *Proceedings of the International Image Algebra and Morphological Image Processing Symposium – IIAMIP 1993*, volume 2030, pages 65–76. SPIE. (Cit  en page 37.)
- Serratosa, F. et Sanfeliu, A. (2006). Signatures versus histograms: Definitions, distances and algorithms. *Pattern Recognition*, 39(5):921–934. (Cit  en page 127.)

- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, First edition. (Cité en page 44.)
- Sheikholeslami, G., Chatterjee, S., et Zhang, A. (2000). WaveCluster: A wavelet-based clustering approach for spatial data in very large databases. *International Journal on Very Large Data Bases*, 8(3):289–304. (Cité en page 44.)
- Shekhar, S., Schrater, P. R., Vatsavai, R. R., Wu, W., et Chawla, S. (2002). Spatial contextual classification and prediction models for mining geospatial data. *IEEE Transactions on Multimedia*, 4(2):174–188. (Cité en page 17.)
- Shen, H. C. et Wong, A. K. C. (1983). Generalized texture representation and metric. *Computer Vision, Graphics, and Image Processing*, 23(2):187–206. (Cité en page 127.)
- Shi, J. et Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905. (Cité en page 35.)
- Soille, P. (2006). Morphological image compositing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):673–683. (Cité en page 30.)
- Soille, P. (2008). Constrained connectivity for hierarchical image decomposition and simplification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1132–1145. (Cité en pages 25 et 37.)
- Soille, P. (2010). Constrained connectivity for the processing of Very-High resolution satellite images. *International Journal of Remote Sensing*, 31(22):5879–5893. (Cité en pages 25 et 37.)
- Soille, P. (2011). Preventing chaining through transitions while favouring it within homogeneous regions. In Soille, P., Pesaresi, M., et Ouzounis, G. K., editors, *Proceedings of the International Symposium on Mathematical Morphology – ISMM 2011*, volume 6671 of *Lecture Notes in Computer Science*, pages 96–107. Springer. (Cité en page 37.)
- Soille, P. et Najman, L. (2012). On morphological hierarchical representations for image processing and spatial data clustering. In *Proceedings of the International Workshop on Applications of Discrete Geometry and Mathematical Morphology – WADGMM 2010*, volume 7346 of *Lecture Notes in Computer Science*, pages 43–67. Springer. (Cité en page 129.)
- Strelkov, V. V. (2008). A new similarity measure for histogram comparison and its application in time series analysis. *Pattern Recognition Letters*, 29(13):1768–1774. (Cité en pages 122 et 127.)
- Stricker, M. A. et Orengo, M. (1995). Similarity of color images. In *Proceedings of the International Conference on Storage and Retrieval for Image and Video Databases – ICSRIVD 1995*, volume 1, pages 381–392. SPIE. (Cité en page 126.)
- Sun, W., Heidt, V., Gong, P., et Xu, G. (2003). Information fusion for rural land-use classification with High-Resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4):883–890. (Cité en page 54.)
- Tarabalka, Y., Chanussot, J., et Benediktsson, J. A. (2010). Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367–2379. (Cité en page 23.)
- Tarabalka, Y., Tilton, J. C., Benediktsson, J. A., et Chanussot, J. (2011). Marker-based hierarchical segmentation and classification approach for hyperspectral imagery. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP 2011*, pages 1089–1092. IEEE Signal Processing Society. (Cité en page 61.)
- Tilton, J. C. (2003). Analysis of hierarchically related image segmentations. In *Proceedings of the IEEE International Workshop on Advances in Techniques for Analysis of Remotely Sensed Data – IWATARS D 2003*, volume 2, pages 60–69. IEEE Geoscience and Remote Sensing Society. (Cité en page 36.)

- Tousch, A. M., Herbin, S., et Audibert, J. Y. (2012). Semantic hierarchies for image annotation: A survey. *Pattern Recognition*, 45(1):333–345. (Cité en page 128.)
- Valero, S., Chanussot, J., Benediktsson, J. A., Talbot, H., et Waske, B. (2010a). Advanced directional mathematical morphology for the detection of the road network in Very High Resolution remote sensing images. *Pattern Recognition Letters*, 31(10):1120–1127. (Cité en page 59.)
- Valero, S., Salembier, P., et Chanussot, J. (2010b). New hyperspectral data representation using binary partition tree. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2010*, volume 2, pages 80–83. IEEE Geoscience and Remote Sensing Society. (Cité en page 37.)
- Vilaplana, V., Marques, F., et Salembier, P. (2008). Binary partition trees for object detection. *IEEE Transactions on Image Processing*, 17(11):2201–2216. (Cité en page 37.)
- Vincent, L. et Soille, P. (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598. (Cité en page 23.)
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 301(58):236–244. (Cité en pages 33, 129 et 134.)
- Weber, J. (2011). *Segmentation morphologique interactive pour la fouille de séquences vidéo*. Thèse de Doctorat, Université de Strasbourg (France). (Cité en page 22.)
- Wemmert, C. (2000). *Classification hybride distribuée par collaboration de méthodes non supervisées*. Thèse de Doctorat, Université Strasbourg 1 (France). (Cité en pages 14 et 45.)
- Wemmert, C., Puissant, A., Forestier, G., et Gançarski, P. (2009). Multiresolution remote sensing image clustering. *IEEE Geoscience and Remote Sensing Letters*, 6(3):533–537. (Cité en pages 4, 44, 46 et 77.)
- Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, 117(1):34–49. (Cité en page 17.)
- Werman, M., Peleg, S., et Rosenfeld, A. (1985). A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing*, 32(3):328–336. (Cité en page 127.)
- Weszka, J. S., Nagel, R. N., et Rosenfeld, A. (1974). A threshold selection technique. *IEEE Transactions on Computers*, 23(12):1322–1326. (Cité en page 23.)
- Yoon, S., Henschen, L. J., Park, E. K., et Makki, S. (1999). Using domain knowledge in knowledge discovery. In *Proceedings of the International Conference on Information and Knowledge Management – CIKM 1999*, pages 243–250. ACM. (Cité en page 114.)
- Zamalieva, D., Aksoy, S., et Tilton, J. C. (2009). Finding compound structures in images using image segmentation and graph-based knowledge discovery. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2009*, volume 5, pages 252–255. IEEE Geoscience and Remote Sensing Society. (Cité en page 33.)
- Zamperoni, P. (1992). An automatic low-level segmentation procedure for remote sensing images. *Multidimensional Systems and Signal Processing*, 3(1):29–44. (Cité en page 23.)
- Zighed, D. J., Tsumoto, S., Ras, Z. W., et Hacid, H., editors (2009). *Mining Complex Data*, volume 165 of *Studies in Computational Intelligence*. Springer. (Cité en page 40.)
- Zucker, S. W. (1976). Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*, 5(3):382–399. (Cité en page 21.)

Annexes

Publications liées aux travaux



Articles en revues internationales

1. Petitjean, F., **Kurtz, C.**, Passat, N., and Gañçarski, P. (2012). Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, 33(13):1805–1815. (Cit  en page 192.)
2. **Kurtz, C.**, Passat, N., Gañçarski, P., and Puissant, A. (2012). Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology. *Pattern Recognition*, 45(2):685–706. (Cit  en page 84.)
3. **Kurtz, C.**, Passat, N., Gañçarski, P., and Puissant, A. (2010). Multiresolution region-based clustering for urban analysis. *International Journal of Remote Sensing*, 31(22):5941–5973. (Cit  en page 78.)



Articles dans les actes de conf rences internationales

1. **Kurtz, C.**, Passat, N., Gañçarski, P., and Puissant, A. (2012a). A histogram semantic-based distance for multiresolution image classification. In *Proceedings of the IEEE International Conference on Image Processing – ICIP 2012* (sous presse). IEEE Signal Processing Society. (Cit  en page 163.)
2. **Kurtz, C.**, Puissant, A., Passat, N., and Gañçarski, P. (2012b). Domain adaptation for the extraction of complex urban patterns from multiresolution satellite images. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2012* (sous presse). IEEE Remote Sensing Society. (Cit  en page 108.)
3. **Kurtz, C.**, Passat, N., Puissant, A., and Gañçarski, P. (2011a). Hierarchical segmentation of multiresolution remote sensing images. In Soille, P., Pesaresi, M., and Ouzounis, G. K., editors, *Proceedings of the International Symposium on Mathematical Morphology – ISMM 2011*, volume 6671 of *Lecture Notes in Computer Science*, pages 343–354. Springer. (Cit  en page 69.)
4. **Kurtz, C.**, Petitjean, F., and Gañçarski, P. (2011b). A context-based approach for the classification of satellite image time series. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium – IGARSS 2011*, pages 495–498. IEEE Remote Sensing Society. (Cit  en page 192.)
5. **Kurtz, C.**, Puissant, A., Passat, N., and Gañçarski, P. (2011c). An interactive approach for extraction of urban patterns from multisource images. In *Proceedings of the IEEE Joint Urban Remote Sensing Event – JURSE 2011*, pages 321–324. IEEE Geoscience and Remote Sensing Society. (Cit  en page 69.)



Articles dans les actes de conf rences nationales

1. **Kurtz, C.** (2012). Une distance hi rarchique bas e sur la s mantique pour la comparaison d’histogrammes nominaux. Dans *Actes de la Conf rence Internationale Francophone sur l’Extraction et la Gestion de Connaissance – EGC 2012*, pages 77–88. (Cit  en page 135.)
2. **Kurtz, C.** (2011a). Extraction de connaissances hi rarchis es   partir d’images multir solutions : application   la t l d tection. Dans *Actes des rencontres de la Soci t  Francophone de Classification – SFC 2011*, pages 56–62. (Cit  en page 84.)
3. **Kurtz, C.** (2011b). Classification non-supervis e de donn es satellites multir solutions. Dans *Actes de l’atelier Fouille de Donn es Complexes, Journ es Francophones Extraction et Gestion des Connaissances (EGC 2011) – FDC 2011*, pages A143–A152. (Cit  en page 78.)

ANNEXE B

Matériel

Les images satellites et les cartes de vérité terrain issues des jeux de données **ILLKIRCH** et **HAUTEPIERRE** sont présentées ci-après.



(a) MRS – 685 × 292 pixels – quatre bandes spectrales.



(b) HRS – 2 740 × 1 168 pixels – quatre bandes spectrales.



(c) THRS – 10 960 × 4 672 pixels – une bande spectrale.

FIGURE B.1 – Présentation du jeu de données **ILLKIRCH** (6 576 m × 2 803 m). (a) image MRS (9,6 m). (b) image HRS (2,4 m). (c) image THRS (60 cm).



(a) MRS – 343×234 pixels – quatre bandes spectrales.



(b) HRS – $1\,372 \times 936$ pixels – quatre bandes spectrales.

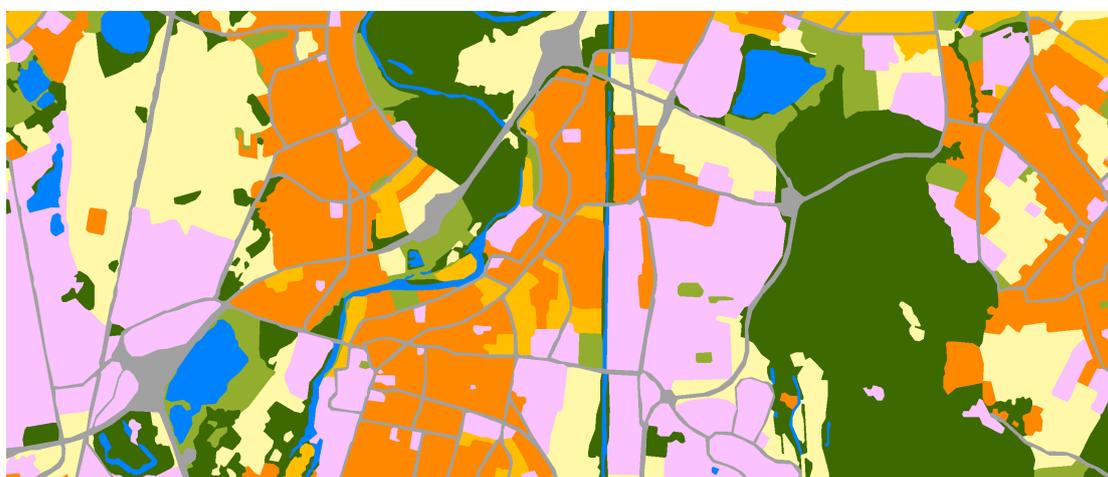


(c) THRS – $5\,488 \times 3\,744$ pixels – une bande spectrale.

FIGURE B.2 – Présentation du jeu de données **HAUTEPIERRE** ($3\,293 \text{ m} \times 2\,246 \text{ m}$). (a) image MRS (9,6 m). (b) image HRS (2,4 m). (c) image THRS (60 cm).



(a) Niveau des quartiers urbains (6 classes sémantiques).

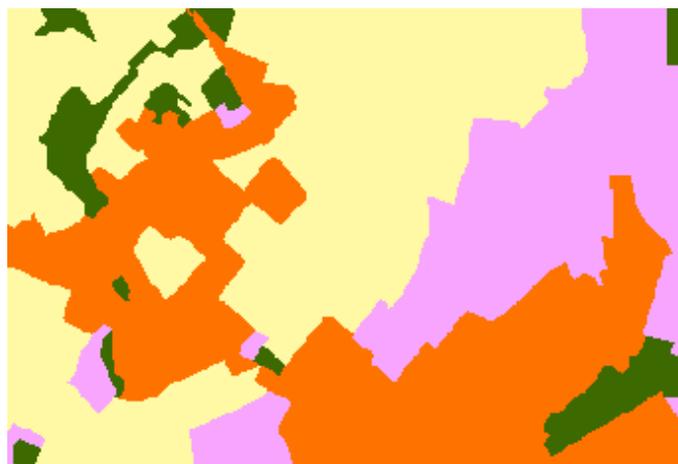


(b) Niveau des blocs urbains (9 classes sémantiques).

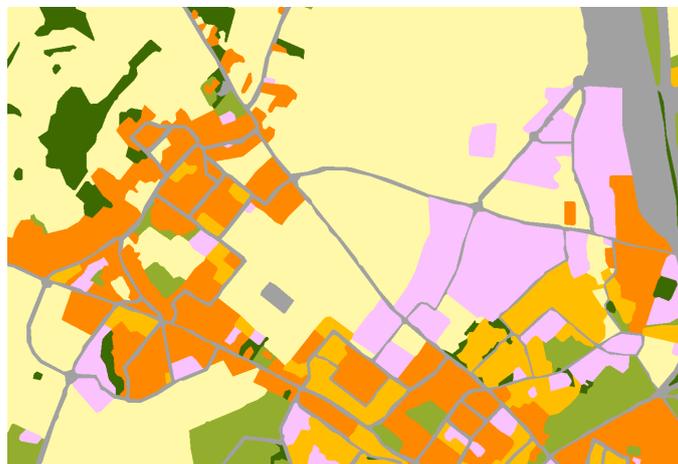


(c) Niveau des objets urbains (1 classe sémantique).

FIGURE B.3 – Cartes de vérité terrain associées aux images du jeu de données **ILLKIRCH**. Voir la table 8.3 pour obtenir des informations relatives aux couleurs et aux concepts sémantiques associés aux classes thématiques représentées.



(a) Niveau des quartiers urbains (6 classes sémantiques).



(b) Niveau des blocs urbains (9 classes sémantiques).



(c) Niveau des objets urbains (1 classe sémantique).

FIGURE B.4 – Cartes de vérité terrain associées aux images du jeu de données **HAUTEPIERRE**. Voir la table 8.3 pour obtenir des informations relatives aux couleurs et aux concepts sémantiques associés aux classes thématiques représentées.

Classification de séries temporelles d'images satellites

Parallèlement aux travaux principaux de cette thèse, une collaboration a été menée avec François Petitjean dans le cadre de sa thèse. Ces travaux ont porté sur le développement d'une nouvelle méthode automatique d'analyse de séries temporelles d'images de télédétection. Cette annexe s'articule de la manière suivante. Le contexte de ces travaux ainsi que les motivations applicatives sont présentés en section A. En section B, nous proposons une méthodologie permettant d'extraire et d'analyser des clusters de pixels ayant des comportements radiométriques similaires à travers le temps. La section C introduit les validations expérimentales réalisées pour évaluer la pertinence de l'approche proposée. Dans une dernière section (Section D), nous dressons le bilan scientifique de ces travaux.

A. Contexte

Avec la disponibilité croissante des images satellites, il devient courant pour les experts, d'utiliser des séries temporelles d'images pour suivre et comprendre le développement et l'évolution des territoires.

Pour extraire des connaissances utiles à partir de ces données, les techniques d'analyse classiques consistent à étudier l'évolution radiométrique des pixels à travers le temps et à les regrouper dans des ensembles homogènes et cohérents, interprétables par les experts. Ces approches, basées sur le paradigme « pixels » ne tiennent pas compte des informations contextuelles des pixels analysés.

Dans les années à venir, des images à hautes résolutions spatiales et temporelles vont devenir largement accessibles apportant ainsi un niveau de détail et de précision très supérieur aux anciennes générations d'images. Pour profiter pleinement de cette richesse, il devient ainsi nécessaire d'utiliser des approches basées régions. Cependant, ces approches requièrent une première étape de segmentation, nécessaire à la prise en compte du contexte spatial et de voisinage des pixels pendant leur analyse. Bien que ce genre d'approche ait été largement étudié dans le cadre de l'analyse d'image mono-temporelle, la généralisation des méthodes de segmentation à l'analyse de séries temporelles reste un problème ouvert.

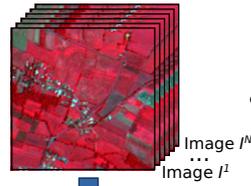
Différentes méthodes ont déjà été proposées dans le but de projeter des segments d'une image à une autre [Gueguen et al., 2006], ou encore dans le but de directement construire des segments spatio-temporels [Moscheni et al., 1998]. Bien que ces approches soient prometteuses, leur application à des séries d'images complexes et hautement résolues reste problématique en raison des perturbations locales (temporelles et spatiales) présentes au sein de ces séries.

B. Proposition

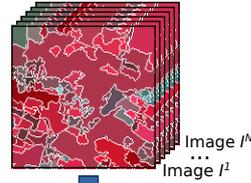
Dans ce contexte, nous proposons de tirer parti des avantages des méthodes basées régions pour analyser des séries d'images satellites, tout en contournant le problème de la segmentation multitemporelle. Nous proposons ainsi de continuer à classifier ces séries d'images en considérant l'évolution radiométrique des pixels à travers le temps. Cependant, dans le but de tenir compte des propriétés spatiales de ces données, nous proposons de caractériser chaque pixel avec des attributs contextuels. En conservant ainsi un paradigme pixels, cette formulation permet de prendre en compte les caractéristiques géométriques de ces données pendant leur analyse, tout en s'abstrayant d'une correspondance entre les segments. Cette formulation peut ainsi permettre d'extraire des cartes de comportements plus précises et donc plus facilement exploitables par les experts, car les détails géométriques des structures d'intérêt y sont mieux conservés. Ce paradigme, utilisant

Entrées

Série temporelles
d'images satellites

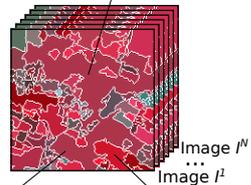


A. Segmentation
des images



B. Caractérisation
des régions

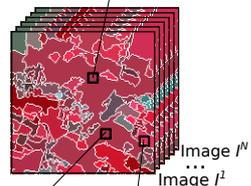
$$R_4^1(e=0.2, s=0.4, a=0.3)$$



$$R_2^1(e=0.2, s=0.8, a=0.9) \quad R_{18}^1(e=0.7, s=0.1, a=0.3)$$

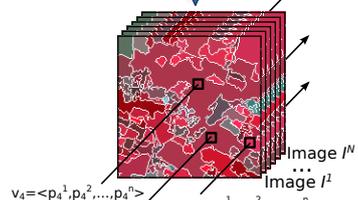
C. Caractérisation
des pixels

$$p_4^1(e=0.2, s=0.4, a=0.3)$$



$$p_2^1(e=0.2, s=0.8, a=0.9) \quad p_{18}^1(e=0.7, s=0.1, a=0.3)$$

D. Construction
des séries temporelles



$$v_4 = \langle p_4^1, p_4^2, \dots, p_4^n \rangle \quad v_{18} = \langle p_{18}^1, p_{18}^2, \dots, p_{18}^n \rangle$$

$$v_2 = \langle p_2^1, p_2^2, \dots, p_2^n \rangle$$

E. Classification
des séries temporelles

Sorties

Carte de
classification

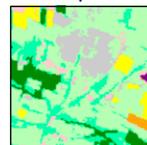


FIGURE C.1 – Illustration de l'approche de classification spatio-temporelle proposée.

des pixels spatialement caractérisés, a été introduit dans le cadre de la segmentation multiéchelle monoimage dans [Bruzzone et Carlin, 2006] et a montré des résultats prometteurs.

La méthode proposée est divisée en cinq étapes principales décrites ci-dessous (et illustrées par la figure C.1); dans le cas standard, la méthode peut traiter n images représentant la même zone géographique à différentes dates.

1. **Segmentation indépendante des n images :** dans un premier temps, les n images sont segmentées indépendamment *via* l'algorithme de segmentation Mean-Shift produisant ainsi n partitions indépendantes.
2. **Caractérisation des régions des n images :** pour chacune des n images segmentées, les segments obtenus sont ensuite caractérisés par des attributs (radiométriques et géométriques) calculés directement sur ces segments. Dans ces travaux, nous avons choisi de caractériser chaque région par : sa moyenne spectrale dans les différentes bandes radiométriques, son élongation, son aire, et la rugosité de son périmètre.
3. **Caractérisation des pixels des n images :** chacun des pixels des n images segmentées peut alors être caractérisé par deux types d'attributs : (1) les valeurs radiométriques directement associées au pixel; (2) les attributs « régions » calculés dans l'étape précédente sur le segment auquel ce pixel est associé. Des vecteurs d'attributs sont ainsi construits et associés à l'ensemble des pixels des n images.
4. **Construction des séries temporelles de pixels caractérisés :** les séries temporelles de pixels peuvent ensuite être formées en concaténant, pour un pixel donné de coordonnées (x, y) , l'ensemble des vecteurs formés précédemment (de coordonnées (x, y)) présents dans les n images considérées.
5. **Classification des séries temporelles de pixels caractérisés :** basée sur ces compositions, une classification non-supervisée des séries temporelles de pixels caractérisés peut être réalisée. Dans ces travaux, nous avons choisi d'utiliser l'algorithme K -MEANS.

C. Validations expérimentales

Cette méthodologie a été validée dans le cadre de l'analyse de paysages agricoles afin d'extraire des classes de couverture des sols reflétant des comportements temporels particuliers. L'approche de classification proposée a été appliquée sur une série de dix images FORMOSAT-2 multispectrales orthorectifiées acquises au cours de la même année. Les figures C.2 (a–c) présentent trois images issues de ce jeu de données. Les résultats de classification obtenus ont ensuite été comparés quantitativement et qualitativement à une carte de vérité terrain fournie par l'expert (Figure C.2 (d)). La figure C.2 (e) illustre l'un des résultats de classification.

Les premiers résultats produits par le biais de cette méthode se sont montrés encourageants. Les cartes de classification obtenues sont de meilleure qualité que les résultats obtenus par l'intermédiaire d'approches classiques basées pixels. Cette étude comparative a ainsi permis de mettre en évidence l'intérêt de cette approche dans le cadre de la classification de séries d'images HRS.

D. Bilan scientifique

Contributions

Ces travaux ont montré que la prise en compte du contexte spatial des pixels (par le biais d'indices de formes ou de textures calculés sur l'ensemble des pixels voisins) pouvait améliorer les résultats de classification des séries temporelles d'images satellites. En effet, l'emploi de certains de ces indices peut aider la discrimination des classes de couverture (et d'évolution) des sols qui n'auraient pas été directement extractibles en utilisant uniquement les informations radiométriques portées par les pixels. De plus, ces travaux ont permis de proposer une approche permettant de s'abstraire de la mise en correspondance des régions (issues des segmentations des différentes images) à travers le temps.

Cependant, il est souvent difficile d'estimer à l'avance quels indices de formes ou de textures pourront être utiles pour améliorer l'extraction d'une classe temporelle particulière de couverture des sols. En effet,

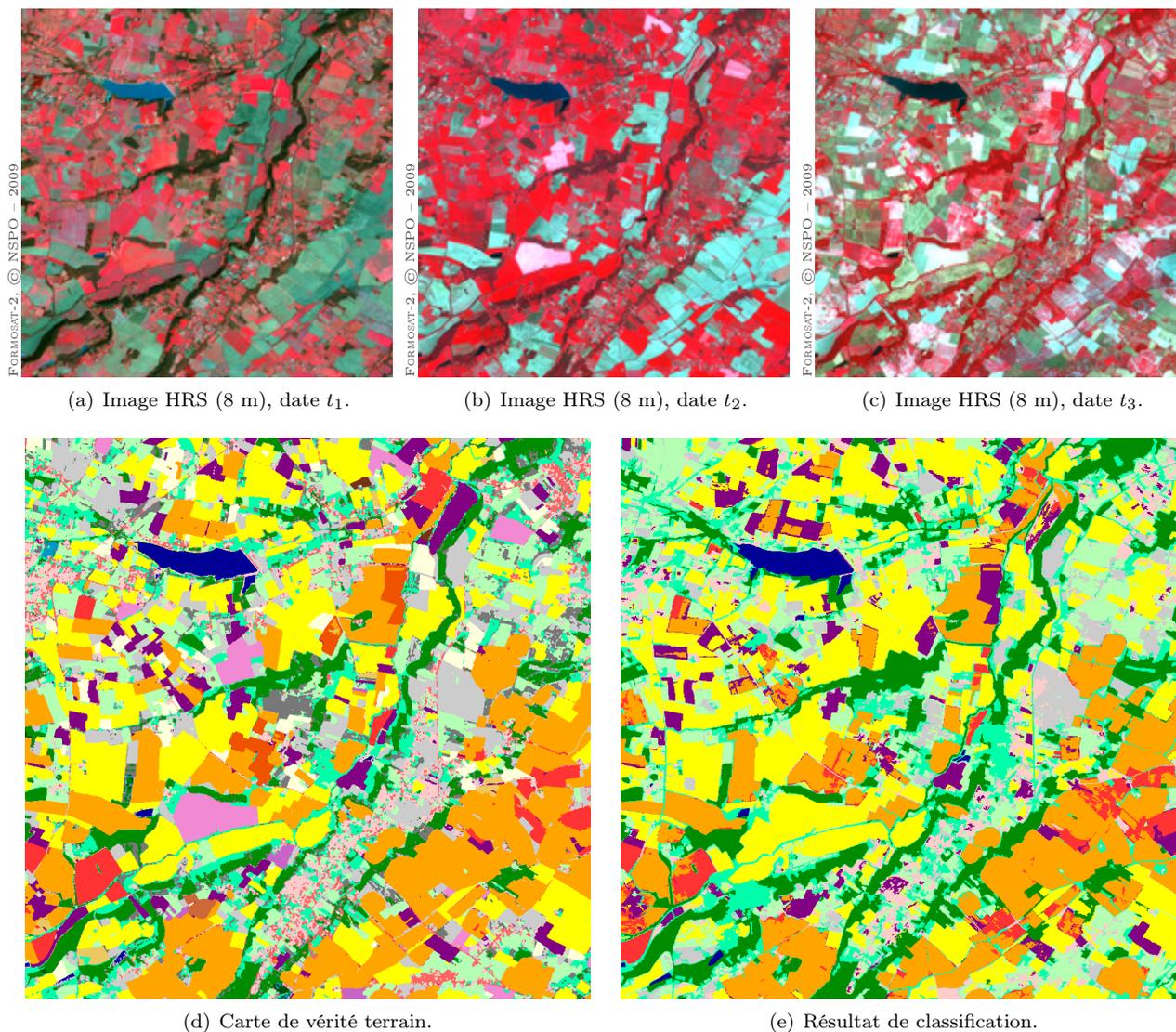


FIGURE C.2 – Illustration d’une partie du jeu de données considéré et d’un résultat de classification obtenu.

si l’utilisation d’un indice de compacité peut améliorer l’extraction d’une classe de surfaces agricoles, son utilisation pour extraire des surfaces de développement urbain peut conduire à des résultats de faible qualité. Ainsi, caractériser la totalité des pixels de ces images en utilisant les mêmes attributs contextuels peut conduire à des résultats non-pertinents. De plus, lors de la classification des séries temporelles de pixels caractérisés, il est souvent difficile d’estimer les bonnes pondérations à attribuer aux caractéristiques utilisées. Nos recherches s’orientent actuellement vers le développement de nouvelles stratégies permettant de pallier ces problèmes.

Valorisation scientifique

Ces différents travaux ont donné lieu à deux contributions scientifiques principales :

- 📌 la première contribution [Petitjean et al., 2012] est liée à un article paru dans la revue internationale *Pattern Recognition Letters*, publiée par *Elsevier* ;
- 📌 la seconde contribution [Kurtz et al., 2011b] est liée à un article dans les actes de la conférence internationale de télédétection *IGARSS 2011, International Geoscience & Remote Sensing Symposium*, publié par *IEEE Geoscience & Remote Sensing Society*.

Développements logiciels

Les travaux de recherche effectués au cours de cette thèse ont donné lieu à différentes contributions dans les logiciels et bibliothèques suivants :

☆ JSL : JAVA SEGMENTATION LIBRARY

- **Début du développement** : 2012
- **Nombre de développeurs** : 5
- **Langage de programmation** : Java
- **Licence** : Libre
- **Propriétaire** : LSIIT – Équipe BFO
- **Téléchargement** : <http://lsiit-cnrs.unistra.fr/fdbt-fr/index.php/Logiciels>
- **Description du logiciel** : JSL est une bibliothèque de traitement d'images, implantant différents algorithmes de segmentation développés au sein du LSIIT, et d'autres issus de la littérature. Parmi ceux-ci, on peut citer un algorithme à bases de LPE, l'algorithme Mean-Shift, et un algorithme relatif à la construction de BPT.

La vocation de cette bibliothèque logicielle est d'être partageable par les membres de l'équipe et de permettre une intégration efficace et rapide des travaux de segmentation développés. En particulier, les algorithmes relatifs à l'approche de segmentation hiérarchique basée sur des exemples, proposés dans cette thèse, y sont en cours d'intégration.

Pour pouvoir manipuler des images THRS de grandes dimensions, nous avons développé dans le cadre de cet outil, une bibliothèque TIFF autorisant le chargement en mémoire vive des parties restreintes des images, nécessaires aux segmentations courantes. Cette stratégie a pour effet de diminuer considérablement les ressources mémoire requises par cette application. Des travaux sont aussi en cours pour permettre d'interfacer la bibliothèque JSL à la bibliothèque Orfeo Toolbox¹ (OTB), dans l'optique de profiter d'algorithmes y étant déjà implantés.

- **Perspectives** : Nous prévoyons de continuer ces travaux d'intégration et d'améliorer la généricité des structures de données afin de permettre la représentation mémoire d'ensembles d'images multirésolutions multitemporelles.

☆ JCL : JAVA CLUSTERING LIBRARY

- **Durée de développement** : 2010–2012
- **Nombre de développeurs** : > 10
- **Langage de programmation** : Java
- **Licence** : Libre
- **Propriétaire** : LSIIT – Équipe BFO

1. Dans le cadre du programme de promotion des données issues des systèmes PLÉIADES, le CNES a développé l'ORFEO Toolbox (OTB), un ensemble de briques algorithmiques permettant l'analyse d'images THRS. L'OTB est distribuée sous la licence CeCILL, afin d'encourager les contributions des utilisateurs. La bibliothèque OTB est téléchargeable à l'adresse suivante : <http://otb.cnes.fr>.

- **Téléchargement** : <http://lsiit-cnrs.unistra.fr/fdbt-fr/index.php/Logiciels>
- **Description du logiciel** : JCL est une bibliothèque logicielle permettant la classification non-supervisée de nombreux types de données génériques. Cette bibliothèque intègre les algorithmes de classification multistratégies développés au sein du LSIIT, ainsi que de nombreux algorithmes issus de la littérature. De plus, elle permet aisément la re-définition de nouveaux types d'attributs pour la caractérisation des données, ainsi que de nouvelles distances (et moyennes) pour le traitement de ces dernières.

Dans le cadre de ces travaux, l'approche de classification multirésolution proposée a été implantée dans cette bibliothèque. Pour ce faire, nous avons développé de nouveaux attributs génériques reposant sur la notion d'histogrammes de composition. Ces derniers peuvent alors être employés pour caractériser des données quelconques, qui peuvent être classifiées par l'un des algorithmes de classification non-supervisée implantés dans cette bibliothèque logicielle. Nous avons également implanté la distance entre histogrammes HSBD pour la comparaison des histogrammes de composition.

- **Perspectives** : Des travaux futurs pourront consister à intégrer une approche permettant d'aider l'utilisateur à la définition des valeurs de proximité sémantique requises pour le calcul de la distance HSBD. Nous prévoyons d'intégrer une approche de classification semi-supervisée permettant d'inférer ces valeurs par le biais de contraintes entre les données définies par l'utilisateur.

☆ MUSTIC : MULTISTRATEGY IMAGE CLASSIFICATION

- **Durée de développement** : 2010–2012
- **Nombre de développeurs** : > 10
- **Langage de programmation** : Java
- **Licence** : Libre
- **Propriétaire** : LSIIT – Équipe BFO
- **Téléchargement** : <http://lsiit-cnrs.unistra.fr/fdbt-fr/index.php/Logiciels>
- **Description du logiciel** : La plate-forme MUSTIC est une interface graphique permettant la classification d'images de télédétection par le biais d'approches basées pixels et régions. Elle repose principalement sur l'utilisation des bibliothèques JSL et JCL. La vocation de cette plate-forme est de permettre une intégration rapide des outils nécessaires à la manipulation graphique des approches d'analyse proposées par les membres de l'équipe. La figure D.1 illustre cette interface.

Dans le cadre de cette thèse, le processus d'analyse descendant PCSC y a été implanté. Pour permettre de tester expérimentalement ce processus, une interface permettant d'exploiter l'approche de segmentation hiérarchique basée sur des exemples a été développée. Un mécanisme de seuillage, permet à l'utilisateur de parcourir interactivement chaque BPT. Grâce à cet outil, ce dernier peut facilement déterminer des exemples de segmentation (globalement, et/ou en raffinant une ou plusieurs branches) pertinents pour son application.

Pour obtenir des temps de calcul rapides (*i.e.*, moins de 30 secondes CPU) et améliorer ainsi l'interactivité avec l'utilisateur, les différentes structures de données nécessaires à la construction des BPT sont calculées lors d'une étape de pré-traitement. L'utilisateur peut alors analyser l'ensemble des coupes d'un BPT afin de choisir la plus adaptée.

Cette plate-forme a été complétée par l'intégration d'une base de données et d'une ontologie dédiées aux objets géographiques auxquelles nous avons adjoint des mécanismes d'utilisation de cette connaissance. Une interface spécialement dédiée à l'analyse interactive d'images de télédétection avec utilisation de connaissances a été développée. Un module permettant l'exportation des résultats sous la forme de cartes de classification au format « .kmz » a aussi été intégré. Ces résultats peuvent alors être échangés et visualisés par l'intermédiaire du logiciel © Google Earth (Figure D.2).

- **Perspectives** : Finalement, nous réfléchissons actuellement à l'intégration du processus itératif C-PCSC au sein de cette librairie et de distribuer ce logiciel sous une licence libre.

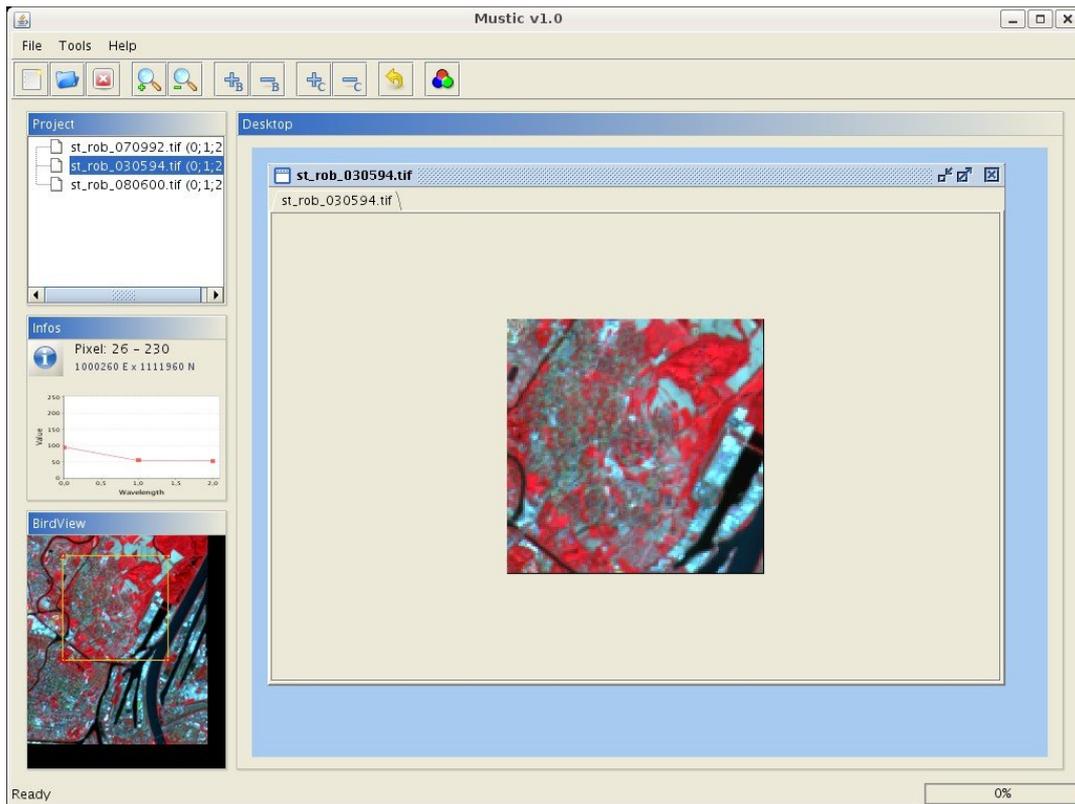


FIGURE D.1 – Illustration de l'interface graphique du logiciel MUSTIC.

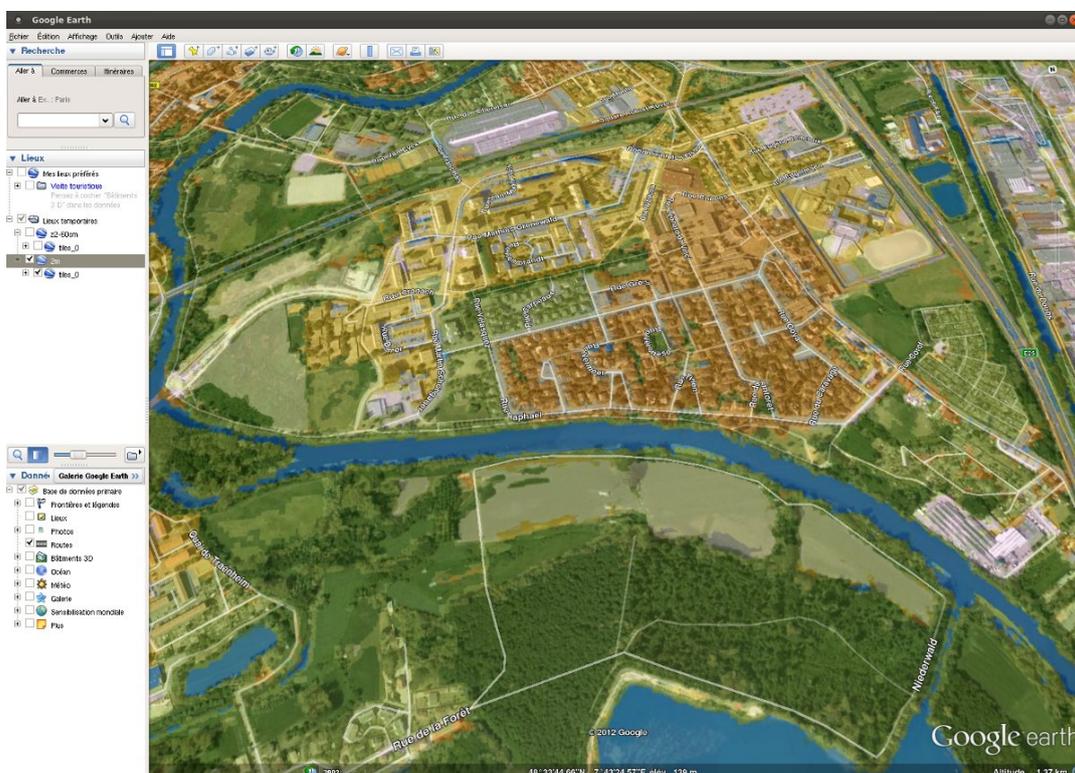


FIGURE D.2 – Illustration de l'exportation des résultats de classification dans le logiciel © Google Earth.

Approche collaborative segmentation–classification pour l’analyse descendante d’images multirésolutions

Depuis la fin des années 1990, les images optiques à très hautes résolutions spatiales issues de capteurs satellitaires sont de plus en plus accessibles par une vaste communauté d’utilisateurs. En particulier, différents systèmes satellitaires sont maintenant disponibles et produisent une quantité de données importante, utilisable pour l’observation de la Terre. En raison de cet important volume de données, les méthodes analytiques manuelles deviennent inadaptées pour un traitement efficace de ces données. Il devient donc crucial d’automatiser ces méthodes par des procédés informatiques, capables de traiter cette quantité de données hétérogènes.

Dans le cadre de cette thèse, nos recherches se sont focalisées sur le développement de nouvelles approches basées régions (*i.e.*, segmentation et classification) permettant l’extraction de plusieurs niveaux de connaissance et d’information à partir d’ensembles d’images à différentes résolutions spatiales. De telles images offrent en effet des vues différentes de la scène étudiée, ce qui peut permettre de faciliter l’extraction des objets d’intérêt. Ces derniers étant structurés sous la forme de hiérarchies d’objets complexes, nos travaux se sont naturellement tournés (1) vers l’utilisation d’approches de segmentation hiérarchique fournissant des ensembles de partitions de la scène à différents niveaux de détail et (2) vers l’intégration de connaissances de haut-niveau dans les processus de fouille de données. De manière plus générale, nous nous sommes intéressés à élaborer un outil informatique reposant sur une stratégie d’analyse descendante, similaire à celle d’un utilisateur, qui consiste à interpréter la scène en considérant, en premier lieu, les grandes zones composant les territoires (à partir des images aux résolutions les plus grossières) puis à affiner récursivement le niveau d’interprétation pour en extraire des zones plus spécialisées (à partir des images aux résolutions les plus fines).

L’ensemble de ces travaux a été implanté dans une bibliothèque logicielle et validé dans le contexte de l’analyse d’environnements urbains à partir d’ensembles d’images multirésolutions.

Mots clés : Analyse d’images, segmentation hiérarchique, classification non-supervisée, images multirésolutions, télédétection

A collaborative region-based approach for the top-down analysis of multiresolution images

In the field of remote sensing image analysis, the recognition of complex patterns from satellite images presents several challenges related to the size, the accuracy and the complexity of the considered data. Indeed, due to the large amount of ground details provided by these images, the classical photo-interpretation approaches do not provide satisfactory results. In this context, it is then relevant to develop new automatic tools adapted to the extraction of complex patterns from such data.

In this thesis, we have proposed new region-based approaches (*i.e.*, segmentation and classification) enabling to extract different levels of information from sets of images at different spatial resolutions. Indeed, such multiresolution sets of images provide different (complementary) views on the represented objects of interest and can be used to make easier the extraction process of these objects. The main principle of the proposed approach is to progressively extract and classify segments/objects of interest from the lowest to the highest resolution data, and then finally to determine complex patterns from VHSR images. This approach, inspired by the principle of photo-interpretation and human vision, merges hierarchical segmentation approaches with multiresolution clustering strategies combined to the integration of high-level background knowledge.

The proposed framework has been validated in the context of the urban mapping of complex objects. Experiments have been carried out on multiresolution sets of satellite images sensed over different cities. The results obtained have shown that the quality and the accuracy of the extracted patterns seem sufficient to further accurately perform both classification or object detection in an operational context.

Keywords : Image analysis, hierarchical segmentation, clustering, multiresolution images, remote sensing
