



HAL
open science

Approche collaborative et qualité des données et des connaissances en analyse multi-paradigme d'images de télédétection

Andrès Troya-Galvis

► **To cite this version:**

Andrès Troya-Galvis. Approche collaborative et qualité des données et des connaissances en analyse multi-paradigme d'images de télédétection. Traitement du signal et de l'image [eess.SP]. Université de Strasbourg, 2016. Français. NNT : 2016STRAD040 . tel-01552626

HAL Id: tel-01552626

<https://theses.hal.science/tel-01552626>

Submitted on 3 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*ÉCOLE DOCTORALE Mathématiques, Sciences de l'Information
et de l'Ingénieur*

Laboratoire ICube UMR 7357

THÈSE présentée par :

Andrés TROYA-GALVIS

soutenue le : 6 Octobre 2016

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/Spécialité : Informatique

**Approche collaborative et qualité
des données et des connaissances
en analyse multi-paradigme
d'images de télédétection**

THÈSE dirigée par :

M. GANÇARSKI Pierre
Mme. BÉRTI-ÉQUILLE Laure

Professeur, Université de Strasbourg
Directrice de recherches, IRD

RAPPORTEURS :

M. LEFÈVRE Sébastien
Mme. SEYLER Frédérique

Professeur, Université de Bretagne-Sud
Directrice de recherches, IRD

AUTRES MEMBRES DU JURY :

M. CORNUÉJOLS Antoine
M. PASSAT Nicolas
M. WEMMERT Cédric

Professeur, AgroParisTech
Professeur, Université de Reims Champagne-Ardenne
Professeur, Université de Strasbourg

Remerciements

Je dédie ce travail à tous ceux qui ont participé d'une manière ou d'une autre à ma formation personnelle et académique. Je tiens à remercier particulièrement mon père Octavio Troya, car c'est en grande partie grâce à lui que je suis arrivé jusqu'ici. Je remercie ma sœur Angélica et ma compagne Maria Fernanda car elles ont toujours été là pour m'encourager dans les moments difficiles et pour partager la joie lors des bons moments.

Je remercie spécialement mon directeur de thèse Pierre Gançarski et ma co-directrice Laure Berti-Équille car ils ont toujours su encourager mes bonnes idées et critiquer positivement les mauvaises idées pour me remettre à temps sur le bon chemin. Je remercie Nicolas Passat pour ses sages conseils lors de nos réunions de travail. Finalement mais tout aussi important, je remercie mon directeur de stage de Master Eric Léclercq, car c'est lui qui m'a introduit au monde de la recherche et qui m'a poussé à réaliser une thèse.

Merci à l'ANR qui a rendu possible la réalisation de cette thèse grâce à son financement. Merci à tous les membres du projet COCLICO dont les nombreux échanges ont toujours été profitables à l'avancement de mes travaux.

Un grand merci à Sébastien Lefèvre et à Frédérique Seyler d'avoir accepté de rapporter cette thèse, ainsi qu'à Antoine Cornuéjols, Nicolas Passat et Cédric Wemmert de participer à l'évaluation et au jury de celle-ci.

Par ailleurs j'exprime ma gratitude à tous les membres de l'équipe SDC : permanents, doctorants, stagiaires, etc., car ils ont toujours été là pour rendre le cadre de travail agréable.

Sommaire

Remerciements	iii
Sommaire	v
Liste des tableaux	ix
Table des figures	xi
Introduction	1
I Qualité de la segmentation d’images et qualité des connaissances	5
1 Évaluation de la qualité de la segmentation d’images	7
1.1 Approches classiques pour évaluer la qualité des segmentations d’images	8
1.1.1 Évaluation supervisée	9
1.1.2 Évaluation non-supervisée	13
1.1.3 Combinaison de métriques	16
1.2 Évaluation de la segmentation en télédétection	17
1.2.1 Évaluation supervisée en télédétection	19
1.2.2 Évaluation non-supervisée en télédétection	24
1.3 UOA : une métrique quantifiant la sur- et la sous-segmentation	26
1.3.1 Évaluation locale	26
1.3.2 Évaluation globale	28
1.3.3 Propriétés	29
1.4 Étude expérimentale de UOA	30
1.4.1 Analyse des indices d’homogénéité	30
1.4.2 Analyse de sensibilité	37
1.4.3 Étude comparative avec des méthodes existantes	40
1.4.4 Intérêt de l’évaluation locale	43
1.5 Bilan scientifique	44
2 Création d’un jeu de données de référence de qualité	49
2.1 Présentation des données	50
2.2 Crowd-sourcing pour évaluer la qualité des données de référence	52
2.2.1 Analyse statistique des résultats	55
2.2.2 Analyse thématique des données	61
2.2.3 Création d’une base de données fiable	62

2.2.4	Deuxième phase de <i>crowd-sourcing</i>	63
2.3	Bilan scientifique	67
II Approche collaborative multi-paradigme pour l'extraction mono-classe		69
3	CoSC : Collaborative Segmentation and Classification	71
3.1	Approches mono-classe pour l'analyse d'image de télédétection	71
3.2	Proposition	73
3.2.1	Définitions	73
3.2.2	Le processus CoSC	74
3.2.3	Données en sortie	75
3.3	Implémentation du processus CoSC	76
3.3.1	Les différents paramètres	76
3.3.2	Étude du comportement de CoSC en fonction de ses paramètres	79
3.3.3	Bilan	89
3.4	Étude détaillée d'une configuration	90
3.4.1	Présentation des données	90
3.4.2	Instanciation de CoSC	90
3.4.3	Résultats	90
3.5	Bilan scientifique	94
III Vers une approche collaborative par agrégation d'agents collaborants		99
4	Approches collaboratives multi-classes	101
4.1	Application d'agents CoSC en cascade : CoSCade	102
4.1.1	Expériences	103
4.1.2	Bilan scientifique	113
4.2	Extraction multi-classe par l'agrégation collaborative d'agents collaboratifs	113
4.2.1	Collaborative CoSC : CoCoSC	114
4.2.2	Expérimentation	117
4.2.3	Bilan scientifique	120
Conclusion		125
Bibliographie		129
A	Segmentation et classification d'images de télédétection	137
A.1	Classification	137
A.1.1	Approches supervisées	137
A.1.2	Approches non-supervisées	139
A.1.3	Approches semi-supervisées et guidées par des connaissances	140
A.2	Segmentation d'images	140
A.2.1	Approches orientées régions	141
A.2.2	Approches orientées contours	141
A.2.3	Approches hiérarchiques	141
A.2.4	Approches par clustering	142

B Métriques supervisées classiques pour évaluer la segmentation d'images	143
C Métriques non-supervisées classiques pour évaluer la segmentation d'images	145
D Prototype développé	149

Liste des tableaux

1.1	Tableau de synthèse des métriques supervisées	12
1.2	Tableau de synthèse des métriques non-supervisées	16
1.3	Métriques supervisées définies dans [Weidner, 2008]	21
1.4	Tableau de synthèse des métriques supervisées pour la télédétection	22
1.5	Tableau de synthèse des métriques non-supervisées pour la télédétection	25
1.6	Corrélation entre les indices d’homogénéité et la taille des segments	36
1.7	Sensibilité des indices en fonction d’un seuil	38
1.8	Résultats des métriques agrégées sur 10 segmentations choisies aléatoirement	42
2.1	Équivalences des classes proposées lors du crowd-sourcing	57
2.2	Nombre, aire et périmètre des segments évalués	58
2.3	Correspondance des labels proposés lors de la deuxième phase de crowd-sourcing	64
3.1	Synthèse du comportement de CoSC avec certains paramètres	85
3.2	Synthèse du comportement de CoSC avec certains paramètres	86
3.3	Synthèse du comportement de CoSC avec certains paramètres	87
3.4	Synthèse du comportement de CoSC avec certains paramètres	88
3.5	Résultats de la classification pour l’étude préliminaire	92
3.6	Métriques de qualité standard	93
3.7	Métriques de qualité floue	94
4.1	Précision, Rappel et leur moyenne harmonique par classe pour les meilleurs résultats obtenus par CoSCade avec les deux segmentations initiales et pour SSC	106
4.2	Meilleurs et pires ordres de classe selon les scores de qualité calculés	107
4.3	Différents paramètres utilisés pour l’expérience sur CoCoSC	118

Table des figures

1	Chaîne de traitement des approches OBIA	2
1.1	Exemples de sur- et sous-segmentation.	7
1.2	Familles de méthodes d'évaluation de la segmentation.	9
1.3	Variabilité des objets identifiables à très haute résolution spatiale	18
1.4	Superposition d'une base de données géographiques avec l'image correspondante	19
1.5	Exemple illustrant l'approche d'évaluation locale	27
1.6	Fonctions pour quantifier de manière fine la qualité locale	28
1.7	Analyse de l'entropie	31
1.8	Analyse du contraste	32
1.9	Analyse du contraste dans l'espace CIELab	33
1.10	Analyse de la cohésion	34
1.11	Analyse de la variance	35
1.12	Analyse de la moyenne des indices	36
1.13	Image employée pour l'étude de <i>UOA</i>	37
1.14	Analyse de sensibilité de la sur- et sous-segmentation en fonction d'un seuil . . .	38
1.15	Analyse de sensibilité de la sur- et sous-segmentation, et taux de segments bien isolés en fonction d'un seuil pour la moyenne des 5 indices testés	39
1.16	Sous- vs. sur-segmentation avec le seuil allant de 0 à 1, pour 10 segmentations sélectionnées aléatoirement	40
1.17	Meilleures segmentations classées par les métriques <i>Z</i> , <i>SU</i> et UOA_{L_2}	41
1.18	Segmentations trouvées par UOA_{Σ}	43
1.19	Illustration des scores de qualité locale	44
1.20	Meilleur résultat trouvé par la métrique <i>SU</i>	46
1.21	Meilleur résultat trouvé par la métrique <i>Z</i>	47
1.22	Meilleur résultat trouvé par la métrique UOA_{L_2}	48
2.1	Exemple de données de référence issues de bases de données géographiques	50
2.2	Données de référence sur l'image de Strasbourg fournies par l'expert	51
2.3	Illustration de la procédure d'amélioration des données de référence	53
2.4	Exemple de segment soumis à évaluation lors du crowd-sourcing	55
2.5	Distribution spatiale des 10 000 segments évalués par la foule sur la qualité de la segmentation	56
2.6	Exemples de segments évalués	58
2.7	Distribution de la moyenne des notes par segment par rapport à leur qualité . . .	59
2.8	Distribution des labels des segments dans les données de référence par rapport à la qualité des segments	60

2.9	Distribution des labels de la foule par rapport à la qualité des segments	60
2.10	Distribution des labels attribués par la foule par rapport aux labels des données de référence	61
2.11	Distribution des classes dans les données de référence d'origine	62
2.12	Méthodologie utilisée pour la production de la base d'exemples	63
2.13	Résultats de la deuxième phase avec les segments à consensus faible et relatif . .	65
2.14	Résultats de la deuxième phase avec les segments en désaccord et désaccord total	66
3.1	Comparaison des approches d'analyse d'images	72
3.2	Schéma conceptuel de CoSC	73
3.3	Image étudiée et données de référence floues	80
3.4	Meilleures classifications mono-classes	82
3.5	Comportement de CoSC avec certains paramètres	82
3.6	Comportement de CoSC avec certains paramètres	83
3.7	Comportement de CoSC avec certains paramètres	83
3.8	Comportement de CoSC avec certains paramètres	84
3.9	Temps d'exécution de CoSC pour les différentes configurations testées	84
3.10	Précision floue pour chaque configuration testée	89
3.11	Segmentations initiales pour trois imagerie	91
3.12	Images, données de référence, et résultats pour l'analyse préliminaire	92
3.13	Segmentations initiales et finales pour l'analyse préliminaire	93
3.14	Vue détaillée des résultats obtenus avec le NDVI	95
3.15	Vue détaillée des résultats obtenus avec OBIA classique	95
3.16	Vue détaillée des résultats obtenus par CoSC	96
3.17	Image étudiée, données de référence et résultats de CoSC sur l'image entière . . .	97
4.1	Schéma de la méthode CoSCade	102
4.2	Image étudiée et données de référence associées	105
4.3	Segmentations initiales pour l'expérience sur CoSCade	105
4.4	Scores de qualité calculés pour différentes configurations	107
4.5	Résultat obtenu par la méthode SSC	108
4.6	Meilleurs résultats obtenus par CoSCade	109
4.7	Métrique UOA_{L_2} avec δ allant de 0 à 1 pour les meilleurs résultats de CoSCade et pour la méthode SSC	110
4.8	Résultat étape par étape de CoSCade : étapes 1 et 2	111
4.9	Résultat étape par étape de CoSCade : étapes 3 et 4	112
4.10	Schéma collaboratif complet	114
4.11	Schéma du processus de collaboration verticale entre agents CoSC	116
4.12	Évolution des résultats de CoCoSC	119
4.13	Analyse des résultats de CoCoSC	120
4.14	Image complète de l'agglomération strasbourgeoise	121
4.15	Données de référence pour l'image complète	122
4.16	Résultats de CoCoSC sur l'image complète	123
D.1	Visualisation de l'évaluation de la segmentation	150
D.2	Visualisation du résultat d'une classification	151
D.3	Visualisation de l'opérateur de fusion d'un segment avec l'un de ses voisins . . .	152

Abréviations

ABIA Agent Based Image Analysis.

ALI Advanced Land Imager.

BPT Binary Partition Tree.

COCLICO COllaboration CLassification Incrémentale et COnnaissances.

CoCoSC Collaborative CoSC.

CoSC Collaborative Segmentation et Classification.

CoSCade CoSC en cascade.

EM Expectation-Maximization.

FA Fuzzy Accuracy.

Fm F-measure.

JSL Java Segmentation Library.

LCE Local Consistency Error.

MLP Multi-Layer Perceptron.

NDVI Normalised Difference Vegetation Index.

OBIA Object Based Image Analysis.

OCE Object Consistency Error.

Pr Précision.

PTSDB Prague Texture Segmentation Data-generator Benchmark.

Re Recall (rappel).

RMSE Root Mean Squared Error.

SDC Sciences des Données et Connaissances.

SLIC Simple Linear Iterative Clustering.

SSC Spatial-Spectral Clustering.

SU Spatial-Unsupervised.

THRS Très Haute Résolution Spatiale.

UOA Under- and Over-segmentation Aware.

VFDT Very Fast Decision Tree.

VT Vérité-terrain.

Symboles

- R_i Segment (Un ensemble de pixels similaires selon un critère donné.)
- S Segmentation (Un ensemble de segments $\{R_i \mid 0 \leq i < M\}$ formant un partitionnement d'une image.)
- $\mathcal{N}(R_i)$ Voisinage (L'ensemble des segments R_j , $i \neq j$ adjacents à R_i .)
- $H(R_i)$ Indice d'homogénéité (Une fonction quantifiant l'homogénéité d'un segment R_i selon un critère donné.)
- H_e Entropie (Indice d'homogénéité basé sur l'entropie.)
- H_c Contraste (Indice d'homogénéité basé sur le contraste des couleurs.)
- $H_{C_{LAB}}$ Contraste CIE Lab (Indice d'homogénéité basé sur le contraste dans l'espace couleurs CIE Lab.)
- H_{co} Cohésion (Indice d'homogénéité basé sur une mesure de cohésion.)
- H_v Variance (Indice d'homogénéité basé sur la variance.)
- H_μ Indice moyen (Indice d'homogénéité basé sur la moyenne d'autres indices d'homogénéité.)
- δ Seuil d'homogénéité (Un seuil sur les valeurs d'un indice d'homogénéité H donné.)
- $\phi_\delta(R_i)$ Fonction d'évaluation locale (La fonction permettant de déterminer si un segment R_i est sur- sous- ou bien isolé selon un indice d'homogénéité H et un seuil δ .)
- over* Taux de sur-segmentation (La fonction permettant de quantifier la sur-segmentation d'un segment.)
- under* Taux de sous-segmentation (La fonction permettant de quantifier la sous-segmentation d'un segment.)
- isolation* Taux d'isolation (La fonction permettant de quantifier le taux d'isolation d'un segment.)
- $\phi_{\delta_{fine}}$ Fonction d'évaluation locale fine (La fonction permettant d'évaluer finement la qualité locale d'un segment selon un indice d'homogénéité H et un seuil δ .)
- UOA_Σ Fonction d'évaluation globale 1 (La fonction permettant d'agrèger les scores de qualité locale des segments au moyen d'une somme pondérée.)
- UOA_{L_2} Fonction d'évaluation globale 2 (La fonction permettant d'agrèger les scores de qualité locale des segments considérant la sur- et sous-segmentation comme deux composantes distinctes.)
- Ψ Sous-segmentation (Le taux de segments considérés comme sous-segmentés dans une segmentation donnée.)
- Θ Sur-segmentation (Le taux de segments considérés comme sur-segmentés dans une segmentation donnée.)
- UOA_{ok} Segments bien isolés (Le taux de segments considérés comme bien isolés par rapport à leur voisins dans une segmentation donnée.)
- consensus* Degré de consensus (Un indice permettant de calculer le taux d'accord entre plusieurs évaluateurs.)
- O_f Fréquence majoritaire (Le nombre d'occurrences du label majoritaire pour un segment.)
- \mathcal{C}_k Classe thématique (Une classe thématique d'intérêt choisie par l'expert.)
- $\mathfrak{C}_{\mathcal{C}_k}$ Classifieur (Un algorithme de classification un-contre-tous entraîné pour extraire la classe \mathcal{C}_k .)

- S_{C_k} Segmenteur spécialisé (Agent de segmentation permettant d'évaluer et modifier un segment de manière adaptée à la classe C_k .)
- $P_{C_k}(R_i)$ Probabilité d'appartenance (La probabilité qu'un segment R_i appartienne à la classe C_k .)
- C_{C_k} Extracteur de classe (Agent de classification permettant de déterminer $P_{C_k}(R_i)$.)
- SC_{C_k} Agent de collaboration (Agent permettant de gérer les interactions entre un segmenteur spécialisé et un extracteur de classe.)
- T_{\in} Seuil d'appartenance (Un seuil sur $P_{C_k}(R_i)$ pour déterminer si un segment R_i appartient à la classe C_k .)
- T_{\notin} Seuil de non appartenance (Un seuil sur $P_{C_k}(R_i)$ pour déterminer si un segment R_i n'appartient pas à la classe C_k .)
- \mathcal{O} Opérateur de modification (Un opérateur qui modifie localement un segment selon une stratégie donnée.)
- O_L Liste de modification 1 (Liste contenant des opérateurs de modification visant à réduire la sur-segmentation.)
- U_L Liste de modification 2 (Liste contenant des opérateurs de modification visant à réduire la sous-segmentation.)
- W_L Liste de modification 3 (Liste contenant des opérateurs de modification divers.)
- \mathcal{F} Opérateur de fusion (Opérateur de modification qui permet de fusionner un segment avec son voisin le plus similaire selon un critère donné.)
- \mathcal{E} Opérateur de rétrécissement (Opérateur de modification qui rétrécit un segment grâce à une érosion morphologique.)
- \mathcal{D} Opérateur de croissance (Opérateur de modification qui fait croître un segment grâce à une dilatation morphologique.)
- P_c Modèle de classification (Paramètre correspondant au classifieur employé pour l'extraction d'une classe thématique donnée.)
- P_m Stratégie de modification (La définition des différents opérateurs de modification à appliquer aux segments en fonction de leur qualité.)
- M_F Stratégie de modification sans scission (Stratégie contenant uniquement les opérateurs de modification \mathcal{F} , \mathcal{E} et \mathcal{D} .)
- M_S Stratégie de modification avec scission (Stratégie contenant les opérateurs de modification \mathcal{F} , \mathcal{E} , \mathcal{D} et un opérateur de scission.)
- P_e Critère d'évaluation (Fonction de qualité à optimiser par CoSC.)
- Q_{clsf} Critère basé sur la classification (Fonction de qualité basée sur le nombre de segments ambigus.)
- Q_{seg} Critère basé sur la segmentation (Fonction de qualité basée sur UOA_{L_2} .)
- Q_{mix} Critère d'évaluation mixte (Fonction de qualité basée sur la combinaison de Q_{clsf} et Q_{seg} .)
- P_s Stratégie de sélection (Stratégie permettant de choisir le segment candidat à être modifié lors de chaque itération de CoSC.)
- s_{alea} Sélection aléatoire (Choix du segment candidat par sélection aléatoire.)
- s_{amb}^- Sélection au moins ambigu (Choix du segment candidat le plus proche de T_{\in} ou T_{\notin} .)
- s_{amb}^+ Sélection au plus ambigu (Choix du segment candidat le plus éloigné de T_{\in} ou T_{\notin} .)

- s_{seg} Sélection au pire segmenté (Choix du segment candidat basé sur $\phi_{\delta_{fine}}$.)
- $CoSC_i$ Agent CoSC (Une instance de CoSC spécialisée dans l'extraction de la classe \mathcal{C}_i .)
- P_i Image de probabilité (Image de probabilité associé à l'extraction de la classe \mathcal{C}_i par un agent CoSC.)
- IS_{over} Segmentation initiale 1 (Segmentation choisie par un expert utilisée dans les expériences de CoSCade.)
- IS_{opt} Segmentation initiale 2 (Segmentation initiale sélectionnée en évaluant 100 segmentations différentes.)
- Q_{gl} Critère d'évaluation globale (Fonction permettant d'estimer la qualité d'une solution intermédiaire de CoCoSC.)

Introduction

La télédétection est l'ensemble des connaissances et des techniques utilisées pour déterminer les caractéristiques de la surface et de l'atmosphère de la Terre à partir de mesures effectuées par un engin aérien ou spatial. Dans le cas particulier de l'imagerie optique, la donnée captée par un système aéroporté ou satellitaire est la réponse spectrale (aussi appelée *signature spectrale*) d'une surface liée à son ensoleillement. Par exemple, chaque satellite est doté de capteurs permettant de mesurer la quantité d'énergie réfléctée par un objet sur une ou plusieurs plages de longueurs d'onde principalement dans le domaine du visible et de l'infrarouge. Ces informations radiométriques se présentent alors sous la forme de matrices de pixels. Suivant les satellites, la résolution spatiale, définie comme étant la surface couverte par un pixel, varie de plusieurs dizaines de mètres à une résolution proche du mètre pour les satellites les plus récents.

L'extraction d'information contenue dans une telle image consiste à associer à chaque pixel de celle-ci une sémantique liée au domaine d'analyse (géographie, géosciences, etc.). En général, l'objectif de cette interprétation est d'obtenir une carte thématique représentant entre autres les classes d'occupation du sol, les types de végétation ou de minéraux dans la scène. Celle-ci peut alors être employée par des experts dans des applications variées telles que le suivi d'écosystèmes [Duveiller et al., 2008, Coppin et al., 2004], la gestion de risques liés aux glissements de terrain [Jha et al., 2008], la gestion de déversements pétroliers [Promper et al., 2014], ou la planification urbaine [Pham et al., 2011, Nielsen, 2015]. Les objets ou classes d'occupation du sol recherchés dépendent directement du domaine d'application.

Cette interprétation peut être réalisée manuellement par un photo-interprète. Néanmoins, cette tâche basée sur un processus d'analyse visuelle est consommatrice de temps, d'autant plus que le volume des données augmente avec les nouvelles technologies. De plus, elle est particulièrement subjective; de ce fait, l'automatisation de ce processus devient une nécessité. Cette extraction automatique d'informations se fait alors principalement par des techniques de traitement d'images et de fouille de données, en particulier la classification.

Jusqu'à la fin des années 1990, l'un des principaux problèmes liés aux images de télédétection résidait dans le fait qu'un pixel peut représenter une zone composée de plusieurs objets de natures très différentes (et cela d'autant plus que la résolution spatiale est basse). On parle alors de *pixel mixte* ou de *mixel*. De fait, les images à faible résolution ne permettent qu'une analyse des tissus thématiques (zone bâtie dense ou clairsemée, végétation, eau, etc.) et/ou structurelle du paysage (extraction de réseaux routiers ou ferrés par exemple).

Ainsi, de nombreuses méthodes pour analyser de telles images ont été développées et ont prouvé leur efficacité. Les méthodologies utilisées par les géographes dans les approches par pixel sont principalement axées sur l'utilisation d'approches d'apprentissage supervisées se basant quasi-uniquement sur les propriétés radiométriques des pixels. Elles peuvent être précédées éventuellement par une phase de classification non-supervisée comme première analyse des données. Cela permet à l'expert d'obtenir des informations sur la structure des classes

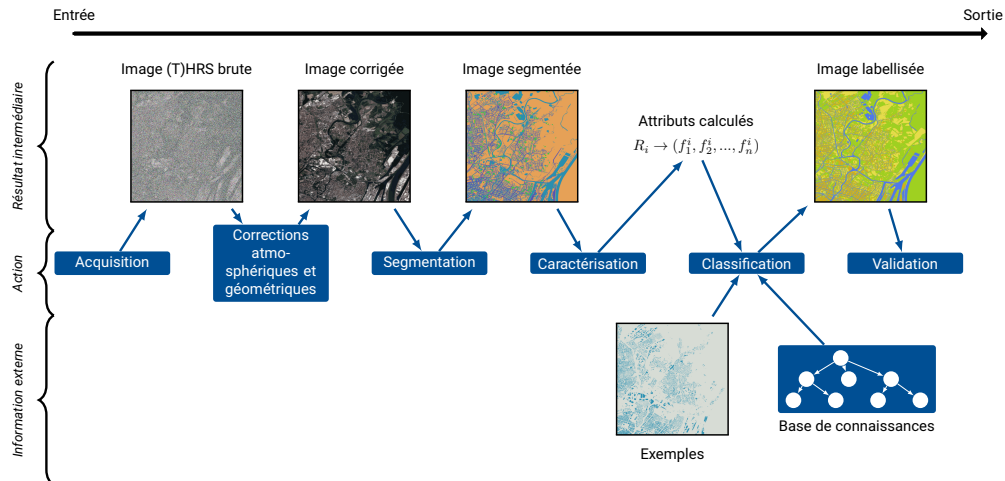


FIGURE 1 – Chaîne de traitement des approches OBIA.

et peut l'aider au paramétrage d'algorithmes supervisés. Le lecteur intéressé peut se référer à [Mather and Koch, 2011] pour approfondir ses connaissances sur ce sujet.

À partir des années 2000, avec la mise en service d'une nouvelle génération de satellites (SPOT 5, Pléiades, etc.) les images disponibles présentent une Très Haute Résolution Spatiale (THRS) allant de 60cm à 2m. De fait, les objets d'intérêt visibles sur une image changent de nature. En effet, alors que les images à basses et moyennes résolutions n'autorisent qu'une analyse des tissus artificialisés ou naturels, les images THRS permettent une analyse des objets géographiques eux-mêmes : bâtiments, routes, arbres isolés, etc. Or, il est fréquent que de tels objets soient composés de nombreux pixels pouvant éventuellement être très différents (en termes de réponse radiométrique) en fonction de la texture de l'objet. De plus, les méthodes par pixel ne prennent pas en compte naturellement des informations de plus haut niveau telles que la forme des objets, leurs structures et éventuellement leurs relations spatiales.

Des méthodes dites basées objet (Object Based Image Analysis (OBIA)) ont été proposées pour pallier ces limites [Blaschke, 2010]. Elles cherchent à construire les objets à partir des pixels et à les caractériser avant de les classifier¹.

La figure 1 illustre les trois principales phases de la chaîne de traitements typique de telles méthodes.

- Une première phase consiste, d'une part, à acquérir les données (étape 1) puis à les corriger via des techniques de corrections atmosphériques et géométriques pour obtenir une image prête à l'analyse (étape 2). Ces deux premières étapes sont communes aux approches par pixel.
- La deuxième phase, dite de vectorisation, consiste à extraire et caractériser les objets d'intérêt de l'image. Un processus de segmentation partitionne l'image en se basant principalement sur des critères radiométriques pour créer des ensembles de pixels que l'on appelle *segments*² (étape 3). Chacun de ces segments est ensuite caractérisé par un ensemble de propriétés telles que la forme, la texture, la composition ou les relations spatiales, calculées directement à partir du segment. Ce couple (segment, caractéristiques) est souvent appelé

1. Le lecteur non familier avec les concepts de classification et de segmentation d'images peut se référer à l'annexe A pour une courte introduction à ces sujets.

2. Pour éviter toute confusion, nous réserverons le terme *objet* pour faire référence aux objets géographiques.

région (étape 4).

À ce stade, chaque segment est donc décrit par un vecteur de caractéristiques que l'on peut exploiter en classification supervisée ou non.

- Enfin, une phase de classification (étape 5) permet d'obtenir une carte thématique, c'est-à-dire une image labellisée où chaque pixel (ou segment) est associé au label de sa classe thématique. Cette carte doit être validée soit par un expert géographe soit par une métrique objective en fonction de données de référence fournies par un expert (étape 6).

Cette chaîne de traitements est fortement sujette aux erreurs à tous les niveaux. Il est évident que toute erreur présente à une étape donnée se propage et se répercute sur la qualité du résultat final.

En particulier, dans la phase de vectorisation, le mécanisme de segmentation introduit un bruit fort. En effet, il n'existe pas à l'heure actuelle de mécanisme de segmentation parfait produisant une bijection entre les segments et les objets géographiques. De fait, des pixels correspondant à d'autres objets peuvent être inclus dans le segment associé à un objet. Réciproquement, des pixels correspondant à l'objet peuvent avoir été omis. Enfin, des segments peuvent ne correspondre à aucun objet physiquement présent dans la scène observée. Par ailleurs, la vectorisation des segments, c'est-à-dire l'association à chaque objet d'une description en termes de caractéristiques de forme, de texture ou autres, pose le problème de la pertinence et de la redondance de ces descripteurs. De plus, il est évident que la qualité de chacun de ces descripteurs, c'est-à-dire sa capacité à rendre compte d'une « vérité » liée à l'objet, dépend directement à la fois de la qualité du segment (couverture de l'objet par le segment principalement) et à la possibilité de calculer ce descriptif concrètement. Il est par exemple difficile de calculer certaines caractéristiques géométriques à partir d'un segment étant donné l'effet de la discrétisation (crénelage) des lignes dans les images.

De même, dans la phase de classification, le choix des descripteurs est crucial car des caractéristiques peu informatives entraînent des modèles de classification peu discriminants. De plus, la classification dépend fortement de la qualité des exemples employés pour l'entraînement des modèles et/ou de la modélisation des connaissances expertes. Or, ces données, de même que celles de validation, sont dans le cadre de la télédétection, très difficiles, voire quasi-impossible à obtenir, et sont de plus bien souvent incomplètes et entachées d'incertitude.

Les recherches actuelles pour définir et mettre en place des méthodes d'analyse d'images basée objet visent principalement à proposer :

- des méthodes de segmentation efficaces ainsi que des mécanismes d'évaluation objective de la qualité de leurs productions afin de choisir au mieux une (ou des) segmentation(s) adaptée(s) à une application donnée ;
- des algorithmes de classification efficaces ainsi que des mécanismes permettant d'estimer la représentativité, la qualité et la certitude des données d'apprentissage et de validation.

La majorité des travaux entrepris dans ce domaine considèrent souvent ces deux aspects indépendamment. Notre hypothèse est que, ces deux étapes cruciales du processus sont fortement liées :

- D'une part, il nous semble évident que les résultats de la classification dépendent de la qualité des segments : si les segments à classer sont de mauvaise qualité (peu représentatifs de la réalité ou très différents des exemples d'apprentissage), alors leur classification sera de mauvaise qualité.
- D'autre part, la qualité de la classification des segments fournit une information sur la qualité des segments eux-mêmes, qui doit pouvoir être utilisée pour ajuster ces segments afin d'augmenter la qualité de la classification.

La mise en œuvre de stratégies permettant d'utiliser ces informations afin d'améliorer conjointement la qualité de la classification et celle de la segmentation nous semble une voie prometteuse.

Afin de concrétiser cette idée, nous avons donc étudié les concepts entrant dans la définition et le développement d'un cadre générique innovant d'analyse d'images de télédétection multi-paradigme intégrant des stratégies de collaboration entre des méthodes de segmentation et de classification pour une amélioration incrémentale et conjointe de la qualité de leurs productions ; l'objectif final étant que la classification thématique proposée soit la plus pertinente possible pour l'expert.

Ainsi, dans cette thèse, nous proposons deux approches différentes basées sur la qualité des données et des connaissances, pour initialiser, guider et évaluer un processus collaboratif de manière objective :

1. Une première approche basée sur une stratégie d'extraction mono-classe qui permet de se focaliser sur les propriétés particulières d'une classe donnée afin de mieux labelliser les objets de cette classe par rapport au reste.
2. Une deuxième approche multi-classe offrant deux stratégies différentes d'agrégation d'extracteurs mono-classes qui permet l'obtention d'une image entièrement labellisée de manière automatique.

Cette thèse est donc structurée suivant trois axes.

- La première partie est consacrée à l'évaluation des données et des connaissances. En particulier, nous nous intéresserons à la qualité des résultats en segmentation d'images et aux données de référence pouvant être employées pour entraîner ou valider des modèles de classification.
- La deuxième partie est consacrée à la méthode Collaborative Segmentation et Classification (CoSC) pour l'extraction d'une seule classe thématique.
- La troisième partie est consacrée à l'étude d'approches collaboratives faisant intervenir plusieurs instances de CoSC afin d'obtenir des méthodes multi-classes pour l'analyse d'images de télédétection.

Enfin, nous concluons cette thèse par un bilan scientifique général des différents travaux effectués ainsi que par les perspectives de recherche que ces derniers laissent entrevoir.

Cette thèse s'inscrit dans le cadre du projet ANR-12-MONU-0001 Collaboration Classification Incrémentale et Connaissances (COCLICO) dont l'objectif est de proposer des méthodes génériques, prenant en compte la qualité des données et des connaissances, pour le traitement de grands volumes de données spatio-temporelles, notamment en analyse d'images de télédétection.

Première partie

Qualité de la segmentation
d'images et qualité des
connaissances

Évaluation de la qualité de la segmentation d'images

Comme nous l'avons vu auparavant, la segmentation d'images est le premier chaînon dans le processus d'analyse d'images basée objet. Dans ce chapitre, nous nous intéressons à ce problème, et plus particulièrement aux moyens d'évaluer la qualité des résultats en segmentation.

La segmentation d'images est connue comme étant un problème mal posé. En effet, un algorithme de segmentation peut, en fonction de ses paramètres, fournir une multitude de solutions, toutes aussi valides les unes que les autres. Tout dépend alors de la manière d'évaluer ces résultats ou de leur utilisation. Il est donc crucial de définir clairement l'objectif recherché par une méthode de segmentation pour réduire l'espace de solutions valables afin de pouvoir quantifier effectivement la qualité de celles-ci.

En télédétection, une segmentation idéale est celle qui fournit un partitionnement de l'image présentant une correspondance parfaite (bijection) entre les segments et les objets présents dans l'image. Malheureusement, il est en général très difficile, voire impossible, d'arriver à une telle correspondance.

En segmentation d'images, on doit couramment faire face à deux types d'erreurs :



(a) Deux bâtiments.

(b) Sous-segmentation.

(c) Sur-segmentation.

FIGURE 1.1 – Exemples de sur- et sous-segmentation.

- sur-segmentation : les segments obtenus sont trop « petits ». Un objet peut-être découpé en plusieurs segments différents ;
- sous-segmentation : les segments obtenus sont trop « grands ». Plusieurs objets peuvent se retrouver au sein du même segment.

La figure 1.1 illustre ces erreurs de segmentation. L'image de gauche représente deux bâtiments différents accolés, aux couleurs hétérogènes. Au centre, on observe un exemple de sous-segmentation où le segment central englobe les deux bâtiments dans leur totalité. À droite on observe un exemple de sur-segmentation où les bâtiments ont été découpés en plusieurs segments.

Dans la suite de ce document, nous allons nous intéresser aux manières d'identifier et de quantifier les erreurs de segmentation pour déterminer la qualité des résultats de la segmentation d'images. D'abord, nous nous intéressons aux méthodes d'évaluation de la segmentation d'image de manière générale, puis nous nous intéressons aux méthodes d'évaluation spécifiquement définies pour des applications en télédétection. Enfin, nous proposons et étudions une nouvelle métrique d'évaluation non-supervisée spécifique à la télédétection.

1.1 Approches classiques pour évaluer la qualité des segmentations d'images

Un nombre très important d'algorithmes de segmentation d'images existe dans la littérature. En revanche, de moindres efforts ont été faits pour évaluer la qualité des résultats. Néanmoins, un certain nombre de chercheurs se sont intéressés au sujet ; nous allons présenter dans cette section les avancées en la matière.

[Zhang et al., 2008] propose une hiérarchie des approches pour évaluer la qualité de la segmentation d'images. Les auteurs divisent ces approches en deux grands groupes (Figure 1.2) : *évaluation subjective* si elle est faite par un humain et *objective* sinon (niveau 1). Les approches objectives se divisent à leur tour (niveau 2) en méthodes d'*évaluation au niveau du système* qui mesurent l'influence des algorithmes de segmentation sur la performance globale des applications, et en méthodes d'*évaluation directe* qui évaluent l'algorithme lui-même. Dans les méthodes directes (niveau 3), on retrouve les *approches analytiques* qui évaluent les algorithmes d'un point de vue purement théorique en tenant compte de leurs propriétés mathématiques et algorithmiques, et les *méthodes empiriques* qui tentent d'évaluer les algorithmes directement à partir de leurs résultats. Finalement, dans les méthodes empiriques (niveau 4), on retrouve les *approches supervisées* qui réalisent l'évaluation par rapport à une segmentation de référence, et *non-supervisées* qui estiment la qualité à partir de propriétés intrinsèques de l'image et des segments.

Les approches par évaluation subjective, aussi connues sous le nom d'approches cognitives, consistent à examiner visuellement le résultat et le qualifier selon les critères personnels de l'examineur. L'avantage de ce type d'approches est également leur principale faiblesse. En effet, lorsque l'évaluation est faite par l'humain, les résultats ont plus de chances de correspondre aux attentes de l'utilisateur. Néanmoins, ces critères varient d'un être humain à l'autre. Si l'on veut réduire le biais introduit par cette subjectivité, il est nécessaire de mettre en place des expériences complexes et robustes pour assurer leur validité statistique par la répétabilité. De fait, la mise en place de ces approches est très coûteuse en temps et en énergie. Évaluer visuellement les résultats est une tâche longue et fastidieuse [Zhang et al., 2008].

Les méthodes d'évaluation au niveau système requièrent l'exécution complète de la chaîne de traitement avant de donner un score de qualité. Or, le temps d'exécution des méthodes en télédétection n'est pas négligeable et rend ces approches d'évaluation sans intérêt.

Les méthodes analytiques fournissent des indications théoriques sur la qualité d'un algorithme donné mais ne donnent pas de garantie sur la qualité des résultats de celui-ci.

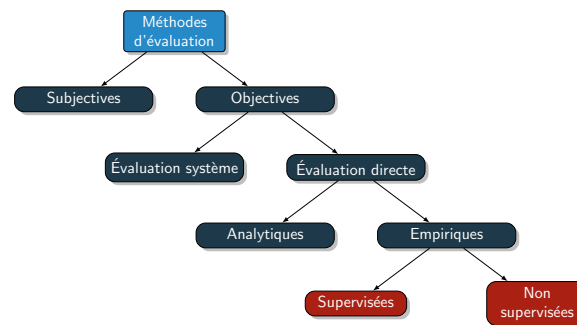


FIGURE 1.2 – Familles de méthodes d'évaluation de la segmentation d'après [Zhang et al., 2008]. En rouge, les approches étudiées dans ce document.

Enfin, les méthodes d'évaluation empiriques travaillent directement sur les résultats des algorithmes. Par conséquent, ces méthodes sont celles le plus fréquemment mises en œuvre en télédétection. Ce sont donc celles que nous allons étudier par la suite.

1.1.1 Évaluation supervisée

Aussi appelées approches de désaccord empirique (*empirical discrepancy* [Zhang, 1996]), cette famille de méthodes correspond à des métriques de distance (ou de similarité) entre deux segmentations. Elles quantifient le taux de désaccord entre deux (ou plus) segmentations données. Afin de pouvoir établir un score de qualité, il est donc nécessaire de disposer d'au moins une segmentation de référence, ou vérité-terrain, obtenue par photo-interprétation ou par d'autres moyens. Elle est alors considérée comme la segmentation idéale. Dans cette section nous présentons les méthodes existantes les plus récentes¹.

[Paglieroni, 2004] étudie les différents aspects à prendre en compte pour définir une métrique de qualité supervisée pour la segmentation d'images. D'après l'auteur, la principale caractéristique que l'on doit rechercher est la cohérence entre les résultats numériques de la métrique et ceux de l'évaluation faite par des humains. Toujours d'après l'auteur, quatre aspects devraient être pris en compte par toute métrique supervisée :

1. mesurer la disparité de deux segmentations en établissant une correspondance simple ou pondérée entre chaque point de la frontière d'une segmentation avec celle de référence ;
2. considérer un seuil de distance minimale pour la correspondance. En effet, cela permet de ne pas pénaliser les différences peu significatives (deux segmentations ne coïncident que très rarement au pixel près) ;
3. effectuer les mesures dans les deux sens, car généralement les métriques de disparité ne sont pas commutatives. Les mesures bidirectionnelles permettent de mieux estimer la similarité/disparité des deux segmentations ;
4. prendre en compte l'orientation des frontières. On peut compléter la métrique par un terme pénalisant les points dont l'orientation n'est pas la même avec son correspondant dans la frontière de référence.

De même que pour les algorithmes de segmentation, on retrouve deux grandes familles de métriques de qualité : les métriques orientées objets et les métriques orientées contours. Par exemple,

1. Pour un bref aperçu des méthodes plus anciennes le lecteur intéressé peut se référer à l'annexe B.

[Polak et al., 2009] définit la mesure Object Consistency Error (OCE) qui est une métrique orientée objet. Elle prend en compte l'erreur deux-à-deux entre les objets de deux segmentations qui se superposent par rapport à l'aire totale des deux objets (union) et leur associe un poids proportionnel à leur taille. Cette métrique est particulièrement adaptée pour évaluer la qualité des segmentations lorsque l'objectif est d'isoler des objets de forme similaire mais de taille variable. Elle s'inspire de la métrique Local Consistency Error (LCE) (Annexe B, Eq. B.2), mais cherche à pénaliser les raffinements mutuels, c'est-à-dire, des segments qui partagent partiellement les contours des segments de référence. Elle est donnée par :

$$\begin{aligned}
OCE(I_g, I_s) &= \min(E_{g,s}, E_{s,g}) \\
E_{g,s} &= \sum_{j=1}^M \left[1 - \sum_{i=1}^N \frac{|A_j \cap B_i|}{|A_j \cup B_i|} \times W_{ji} \right] W_j \\
W_{ji} &= \frac{\bar{\delta}(|A_j \cap B_i|)|B_i|}{\sum_{k=1}^N \bar{\delta}(|A_j \cap B_k|)|B_k|} \\
W_j &= \frac{A_j}{\sum_{l=1}^M |A_l|}
\end{aligned} \tag{1.1}$$

où A_i (resp. B_j) est un segment de la segmentation g (resp. s), les poids W_{ji} et W_j pénalisent les segments dont l'intersection est non nulle en fonction de leur taille. Ainsi, $E_{g,s}$ mesure la disparité entre la segmentation g et la segmentation s . Notons que cette mesure n'est pas commutative ; elle doit donc être calculée dans les deux sens. Le score minimum est alors conservé.

[Vojodi and Eftekhary Moghadam, 2012] présente également une métrique orientée objet. Elle se base sur un descripteur de forme calculé sur chaque région à partir des angles entre deux rayons partant d'un point à la frontière d'une région vers tous les autres points de cette frontière. Pour chaque point p_i de la frontière, on calcule les rayons reliant p avec les points p_{i+k} et p_{i-k} , $k \in 1, 2, \dots, N/2$ et l'on mesure l'angle entre chaque paire de rayons, puis on calcule le moment angulaire d'ordre 2. L'indice de forme de la région est donc représenté par un vecteur des moments sur chaque point de sa frontière. Afin de mesurer la similarité entre deux segmentations, on calcule la similarité des indices de forme de chaque région de la première segmentation avec toutes les régions dont l'intersection est non nulle. La valeur finale est la somme des rapports entre les valeurs minimale et maximale pour la similarité de chaque région. Cette métrique est donnée par la formule suivante :

$$\begin{aligned}
EV &= \min(E_{gs}, E_{sg}) \\
S_{s,g} &= \sum_{j=1}^{mg} \frac{\arg \min_{i \in ms} (OCS(V_j, V_i) \times \bar{\delta}(|V_j \cap V_i|))}{\arg \max_{i \in ms} OCS(V_j, V_i)} \\
E_{gs} &= 1 - \frac{S_{g,s}}{mg}
\end{aligned} \tag{1.2}$$

où OCS est une mesure de similarité entre deux segments en fonction de leur forme donnée.

Pareillement, [Pont-Tuset and Marques, 2013] présente une métrique orientée objet reposant sur le fait que la segmentation d'images est de plus en plus utilisée comme la première étape pour la reconnaissance d'objets. Dans un premier temps, le taux de chevauchement des régions provenant des deux segmentations comparées est calculé, et deux seuils sont définis pour déterminer si

un objet est pertinent ou non pour l'évaluation. En effet, si le pourcentage de chevauchement est élevé, le segment est pris en compte pour le calcul. Si le taux de chevauchement est négligeable pour un segment donné, celui-ci est considéré comme du bruit. Il est alors possible de calculer la précision et le rappel pour les objets ; ainsi :

$$\begin{aligned} P_{op} &= \frac{oc + fr + \beta pc}{|S|} \\ R_{op} &= \frac{oc' + fr' + \beta pc'}{|G|} \end{aligned} \quad (1.3)$$

où

- S est la segmentation à évaluer, G est l'ensemble des régions de toutes les vérités-terrain disponibles ;
- oc (resp. oc') est le nombre d'objets candidats dans S (resp. G) ;
- fr (resp. fr') est la somme des fragmentations de toutes les régions dans S (resp. G) ;
- pc, pc' le nombre de parties candidates ;
- β est une pondération pour ne compter les parties que partiellement afin de ne pas favoriser la sur-segmentation.

La *F-measure* peut être utilisée pour combiner ces deux indices :

$$F_{op} = \frac{P_{op} \cdot R_{op}}{(1 - \alpha)R_{op} + \alpha P_{op}} \quad (1.4)$$

[Li et al., 2013] présente une métrique supervisée orientée contour. L'idée de base est très intuitive : il s'agit de tester les pixels aux frontières en récompensant les paires de pixels qui se trouvent à la fois dans le segment évalué et celui de référence. Elle se traduit par la formule suivante :

$$\begin{aligned} E &= \sum_i \sum_{j \in N_k(i)} W_{ij} (p_i \otimes p_j) \\ \text{avec } p_i \otimes p_j &\begin{cases} 1 & R(S_1, p_i) \neq R(S_1, p_j) \\ 0 & \text{sinon} \end{cases} \\ \text{et } W_{ij} &\begin{cases} W_{ij}^+ & R(S_2, p_i) \neq R(S_2, p_j) \\ W_{ij}^- & \text{sinon} \end{cases} \end{aligned} \quad (1.5)$$

Le calcul s'effectue dans un voisinage d'ordre k . La taille de cette fenêtre permet de définir une tolérance plus ou moins forte aux frontières. L'opérateur \otimes s'assure que l'on ne teste que les pixels qui n'appartiennent pas au même segment dans la segmentation de départ (S_1) ; les poids W_{ij} favorisent les pixels appartenant à des segments différents dans la segmentation de référence (S_2). Les poids positifs et négatifs sont calculés en fonction de la distance entre les pixels testés. Tous ces calculs sont effectués dans les deux sens. Lorsque la segmentation de référence est une vérité-terrain, cela revient à calculer la précision, et lorsque la segmentation de référence est la segmentation de départ (S_1), cela revient à calculer le rappel. Une moyenne harmonique est employée pour combiner ces deux valeurs.

Un autre détail remarquable de ces travaux est la façon de représenter les vérités-terrain. En effet, elles sont organisées de manière hiérarchique suivant des niveaux de détail différents dans une structure que les auteurs appellent un arbre de perception. Ainsi, lors de l'évaluation d'une segmentation, les auteurs choisissent la vérité-terrain la plus proche par rapport au nombre de segments, ce qui rend la métrique plus ou moins tolérante aux raffinements mutuels. Soient S_1 et

S_2 des segmentations ; on dit que S_1 raffine S_2 si S_1 s'emboîte parfaitement dans S_2 ; considérons $C(S)$ l'ensemble des pixels appartenant aux contours des segments d'une segmentation S , alors $C(S_2) \subset C(S_1)$. Ce point est important car il existe une multitude de segmentations correctes à différents niveaux de détail. Dans certaines applications, on peut avoir intérêt à accepter ce type de différences.

Il existe également des métriques permettant de faire des comparaisons de 1 à n segmentations. Cela s'avère utile lorsque l'on dispose d'un jeu de plusieurs segmentations de référence, comme dans le cas de la base de données Berkeley Segmentation Dataset and Benchmark [Martin et al., 2001] où l'on dispose pour chaque image, d'un ensemble de segmentations manuelles réalisées par différents individus. [Peng and Li, 2013] présente un exemple de ces métriques. Les auteurs proposent une approche probabiliste pour mesurer la similarité d'une segmentation S par rapport à un ensemble segmentations de référence $G = \{G_1, \dots, G_n\}$ où $G_i = \{g_i^1, \dots, g_i^m\}$ et chaque g_i^j est un segment. Dans un premier temps, les segmentations de référence sont fusionnées en une seule vérité-terrain G^* en appliquant une stratégie optimiste pour faire en sorte que la similarité entre S et G^* soit maximale. La métrique considère deux hypothèses :

- la segmentation à évaluer est obtenue par une transformation géométrique quelconque sur l'ensemble de vérités terrain initiales, cette hypothèse est notée \mathcal{H}_G ;
- tous les éléments de G^* sont pris parmi l'une des vérités terrain initiales $\mathcal{H}(g_j^*, G)$.

La métrique M correspond donc à la probabilité que chaque couple (pixel, segment) de la segmentation soit observé étant donné les deux hypothèses précédentes :

$$M(S, G) = \prod_{i=1}^K \prod_{s_j \in S_i} P(s_j | \mathcal{H}_G, \mathcal{H}(g_j^*, G)) \quad (1.6)$$

En résumé, il existe une variété importante de métriques de qualité supervisées. Elles constituent un bon moyen d'évaluer les performances des algorithmes de segmentation d'images lorsque l'on dispose d'une ou plusieurs segmentations de référence. Le tableau 1.1 présente les six méthodes présentées ci-avant.

TABLEAU 1.1 – Tableau de synthèse des métriques supervisées.

Nom	Description	Avantages	Limites
<i>OCE</i> (Eq. 1.1)	Mesure la disparité entre deux segmentations en fonction des objets dont l'intersection est non nulle et de leurs tailles.	Pénalise la sur- et la sous-segmentation.	N'est pas adapté pour des applications où plusieurs niveaux de détail sont possibles.
<i>EV</i> (Eq. 1.2)	Mesure la similarité de deux segmentation en fonction de la forme des segments.	Pénalise la sur- et sous-segmentation.	Le calcul du descripteur de forme est coûteux.

Nom	Description	Avantages	Limites
Précision et rappel objets/parties (Eq. 1.3)	Met en relation chaque objet de référence avec un ou plusieurs segments dont l'intersection avec l'objet est assez importante.	Permet de quantifier la sur/sous-segmentation.	N'est pas adapté pour des applications où plusieurs niveaux de détail sont possibles.
F-measure (Eq. 1.4)	Combine la précision et rappel pour quantifier la qualité d'une segmentation par rapport à une segmentation de référence.	Donne un ordre d'idée sur la quantité de pixels mal segmentés.	N'est pas adapté pour des applications où plusieurs niveaux de détail sont possibles.
Précision et rappel aux frontières (Eq. 1.5)	Calcule la similarité aux frontières de deux segmentations dans une fenêtre de k-voisins.	Intuitive et facile à calculer.	Coût algorithmique non négligeable.
Probabilistic Rand Index (Eq. 1.6)	Mesure le nombre d'étiquetages en accord avec l'image de référence.	Prend en compte plusieurs images de référence.	Ne correspond pas forcément au jugement humain d'une bonne segmentation.

1.1.2 Évaluation non-supervisée

Aussi appelées approches de qualité empirique (*empirical goodness* [Zhang, 1996]), ces approches essaient de quantifier la qualité d'une segmentation uniquement à partir de propriétés intrinsèques de l'image et des segments. Bien qu'il n'y ait pas besoin de segmentation de référence ou de protocoles d'évaluation complexes, ces approches s'avèrent les plus difficiles à concevoir. En effet, il faut dans un premier temps définir très clairement ce qu'est une segmentation de bonne qualité et ensuite trouver les propriétés et les paramètres qui vont permettre de quantifier cette notion de qualité. Il est néanmoins intéressant d'étudier cette famille de méthodes puisqu'elles permettent, entre autres, l'automatisation du paramétrage des algorithmes de segmentation.

Il existe dans la littérature de plus en plus de tentatives pour définir des métriques de qualité non-supervisées. [Zhang et al., 2008] présente une étude de cette famille de méthodes. Les auteurs distinguent trois types d'indices qui mesurent certaines caractéristiques d'un segment et qui sont combinés entre eux afin d'obtenir une métrique globale.

1. Les indices d'uniformité intra-segment qui mesurent l'homogénéité des pixels à l'intérieur d'un segment, comme des indices de contraste, des mesures de texture ou l'entropie.
2. Les indices de disparité inter-segment qui mesurent les différences d'un segment par rapport à son voisinage comme la différence des moyennes des couleurs entre les segments, la différence de couleurs aux frontières, la distance aux barycentres ou la différence d'entropies.
3. Les indices de forme qui mesurent des propriétés concernant la forme des segments comme des indices de rectangularité ou de circularité.

Les auteurs montrent que les métriques existantes sont plus ou moins adaptées pour discerner les bonnes segmentations parmi un ensemble de segmentations générées par un algorithme donné. Néanmoins, elles ont du mal à mimer l'évaluation subjective d'une personne. Plusieurs problèmes ont été identifiés sur les métriques étudiées :

- les indices d'uniformité des segments sont trop sensibles au bruit et favorisent la sous-segmentation ;
- la plupart des métriques considèrent une distribution gaussienne des pixels sur chaque segment ;
- les indices d'uniformité et de disparité ne sont pas équilibrés ou ne se complètent pas réellement ;
- les métriques n'utilisent que des caractéristiques bas niveau. Elles ne considèrent pas la sémantique des objets.

Nous allons maintenant étudier plus en détail quelques métriques récentes².

[Chen and Wang, 2004] présente un indice de qualité non-supervisé tenant compte de la différence perceptible des couleurs. Les calculs se font dans l'espace de couleurs CIE Lab qui a la particularité de bien modéliser la perception humaine des couleurs. Comme dans la plupart des méthodes non-supervisées, le critère se compose de deux termes différents :

$$E_{intra} = \frac{\sum_{x=1}^N \sum_{y=1}^M u(\|f(x,y) - \bar{f}(x,y)\|_{Lab} - th)}{N \times M} \quad (1.7)$$

$$E_{inter} = \frac{\sum_{i=1}^R \sum_{j=1, j \neq i}^R w_{ij} \times u(th - \|\bar{f}_i - \bar{f}_j\|_{Lab})}{C \times N \times M}$$

où E_{intra} est le terme d'uniformité intra-segment permettant de quantifier la sous-segmentation, f est l'image en question et \bar{f} correspond à l'image segmentée où chaque région est remplie par la couleur moyenne des pixels qui la composent. $\|\cdot\|_{Lab}$ correspond à la différence des couleurs dans l'espace CIE Lab. Le terme de disparité inter-segment E_{inter} permet de quantifier la sur-segmentation ($u(t) = 1$ si $t > 0$ et $u(t) = 0$ sinon), th est un seuil d'homogénéité donné par l'utilisateur.

Il est montré que le terme de sous-segmentation diminue lorsque le terme de sur-segmentation augmente ; les auteurs combinent les deux termes par une simple addition et argumentent que les résultats donnés par cette approche sont plutôt satisfaisants.

[Andres et al., 2009] proposent une métrique pour mesurer l'homogénéité d'un segment. Elle se base sur des marches aléatoires dans un graphe valué et permet d'estimer si un segment est sous-segmenté ou sur-segmenté. On considère l'image comme un graphe dont les sommets sont les pixels et les arêtes connectent chaque pixel avec ses voisins. Le poids associé à chaque arête correspond à une mesure de similarité des deux pixels. Deux graphes différents sont considérés : un graphe dit géométrique, où tous les poids ont la même valeur et un graphe de similarité basé sur une fonction qui donne pour chaque paire de pixels, la probabilité qu'ils soient séparés par une frontière. On considère que le processus est markovien, où chaque sommet correspond à un état et la probabilité de passer d'un état à un autre ne dépend pas des transitions précédentes. Le temps de relaxation associé au graphe est le nombre d'itérations que le processus met avant de converger vers sa distribution stationnaire. Ainsi, on définit l'homogénéité d'un segment par :

2. Pour une brève description des méthodes classiques, le lecteur peut se référer à l'annexe C.

$$\mathcal{H} = \frac{\tau_w}{\tau_g} \quad (1.8)$$

où τ_w et τ_g correspondent aux temps de relaxation associés aux graphes de similarité et géométrique respectivement. Lorsque τ_w est égal à τ_g , alors $\mathcal{H} = 1$ et le segment peut être considéré homogène. *A contrario*, τ_w très supérieur à τ_g indique que le segment est sous-segmenté. En effet, la marche aléatoire a tendance à rester bloquée longtemps à l'endroit où il y a une frontière puisque la probabilité pour la traverser est moindre, et par conséquent, le temps de relaxation est plus élevé.

Les métriques basées sur l'entropie utilisent des concepts de théorie de l'information. Elles considèrent que la quantité d'information contenue dans un segment peut être interprétée comme le taux de désordre des pixels à l'intérieur du segment, et donc comme un indice d'homogénéité. Dans [Zhang et al., 2003], les auteurs ont été les premiers à employer cette idée comme critère d'évaluation. Ils définissent un indice d'entropie local qui permet de mesurer l'homogénéité intra-segment, ainsi qu'un indice d'entropie global sur la structure de la segmentation permettant de mesurer l'hétérogénéité inter-segment. La métrique E est définie par :

$$\begin{aligned} E &= \mathcal{H}_r(I) + \mathcal{H}_l(I) \\ \mathcal{H}_v(R_j) &= - \sum_{m \in V_j} \frac{L_j(m)}{S_j} \log \frac{L_j(m)}{S_j} \\ \mathcal{H}_r(I) &= \sum_{j=1}^N \left(\frac{S_j}{S_I} \right) \mathcal{H}_v(R_j) \\ \mathcal{H}_l(I) &= - \sum_{j=1}^N \frac{S_j}{S_I} \log \frac{S_j}{S_I} \end{aligned} \quad (1.9)$$

où \mathcal{H}_v mesure l'entropie d'un segment, S_I est le nombre de pixels dans l'image, S_j est le nombre de pixels dans le segment R_j , V_j est l'ensemble des valeurs possibles d'un pixel et $\frac{L_j(m)}{S_j}$ est la probabilité qu'un pixel dans le segment ait la valeur m . \mathcal{H}_r est le terme d'uniformité intra-segment ; c'est la somme des entropies de tous les segments, pondérée par la taille des segments. \mathcal{H}_l est le terme de disparité inter-segment. Il prend en compte la probabilité qu'un pixel appartienne au segment j . Ces deux termes sont finalement additionnés pour obtenir un score global.

[Khan and Bhuiyan, 2014] proposent une métrique basée sur l'entropie. Les auteurs proposent d'améliorer les métriques similaires existantes grâce à une pondération. La métrique est donnée par la formule :

$$\mathcal{H}_W(j) = - \sum_{m \in V_j} \frac{|m - \bar{m}|}{n_i} \frac{L_j(m)}{S_j} \log \frac{L_j(m)}{S_j} \quad (1.10)$$

où \bar{m} est la moyenne des valeurs des $m \in V_j$, n_i est le nombre de valeurs possibles pour m , $L_j(m)$ est le nombre de pixels ayant la valeur m , et S_j est le nombre de pixels dans la région j .

Cette pondération permet de mieux quantifier le désordre à l'intérieur d'une région. Les auteurs calculent alors une matrice d'entropies mutuelles entre les régions deux-à-deux ; la diagonale correspond à l'entropie de chaque région. Dans le cas idéal, l'entropie intra-région doit être minimale (0 si une région est complètement homogène). Une haute valeur d'entropie intra-région est alors interprétée comme la sous-segmentation de cette région. En revanche lorsque l'entropie d'une région est trop petite, il est nécessaire de vérifier que l'entropie inter-région soit élevée pour s'assurer qu'il ne s'agit pas de sur-segmentation. Pour déterminer la qualité d'une segmentation, les auteurs prennent la valeur maximale trouvée pour l'entropie intra-région et la valeur minimale

trouvée pour l'entropie inter-région, le but étant de minimiser la première et de maximiser la seconde.

En résumé, les métriques non-supervisées modélisent les notions de sur- et sous-segmentation en mesurant l'homogénéité intra-segment et l'hétérogénéité inter-segment avec des stratégies variées. Le tableau 1.2 présente les quatre métriques introduites précédemment.

TABLEAU 1.2 – Tableau de synthèse des métriques non-supervisées.

Nom	Description	Avantages	Limites
Entropy based (Eq. 1.9)	Emploie la notion d'entropie pour représenter le désordre à l'intérieur des régions.	Est générique, peut être calculée sur des caractéristiques quelconques.	Est une mesure complètement globale, elle ne prend pas en compte l'aspect spatial des régions ni leur forme.
Random Walk relaxation times (Eq. 1.8)	Détermine l'homogénéité en faisant le rapport des temps de relaxation d'un processus markovien sur un graphe tenant compte de la probabilité d'avoir une frontière entre deux pixels et un graphe modélisant la géométrie.	Permet de déterminer si le résultat est sur-segmenté ou sous-segmenté.	Son implémentation peut s'avérer complexe et coûteuse.
Visible color difference (Eq. 1.1.2)	Mesure l'homogénéité et l'hétérogénéité en termes de distance dans l'espace de couleurs CIELab.	Modélise assez bien la perception humaine des couleurs.	Limité à des images couleur.
Entropie pondérée (Eq. 1.10)	Prend en compte l'entropie mutuelle de chaque segment avec ses voisins.	Quantifie mieux l'homogénéité des régions que l'entropie classique.	N'utilise que les scores (min et max) de deux segments pour l'évaluation globale de la segmentation.

1.1.3 Combinaison de métriques

De nombreuses approches visant à combiner différentes métriques afin d'obtenir des résultats plus robustes ont été proposées dans la littérature. Par exemple, [Lin et al., 2013] présente un cadre d'évaluation basé sur des méthodes d'apprentissage automatique pour combiner les résultats de différentes métriques indépendantes pouvant être supervisées ou non. L'apprentissage est réalisé sur un jeu de données contenant des segmentations manuelles. Les segmentations ont préalablement été présentées (deux par deux) à des individus devant choisir la meilleure d'entre

elles. Le problème revient donc à apprendre un modèle pour classifier des segmentations en deux classes (bonne et mauvaise). Ce modèle était ensuite comparé avec le jugement humain pour évaluer la précision des résultats. Deux expériences ont été menées.

- La première pour déterminer le taux de complémentarité des différentes métriques employées. Elle consistait à utiliser chaque métrique de manière indépendante et à les comparer entre elles.
- La seconde pour évaluer différentes segmentations avec le cadre proposé. Les résultats montrent que la métrique combinée donne des résultats plus pertinents que ceux des métriques calculées individuellement.

Similairement, [Vojodi et al., 2013] propose une méthode basée sur la programmation génétique afin de combiner des métriques indépendantes. Les individus sont des arbres où les nœuds correspondent à des opérations arithmétiques et les feuilles sont des métriques individuelles mesurant l'uniformité ou l'hétérogénéité des régions. L'opérateur de reproduction se limite à copier les individus ayant la meilleure valeur de *fitness*. L'opérateur de croisement échange deux sous-arbres entre deux individus afin d'en générer un nouveau. Finalement, l'opérateur de mutation échange aléatoirement un sous-arbre d'un individu avec un sous-arbre de ses parents choisi de manière aléatoire. Ainsi, l'idée est d'explorer les différentes façons de combiner les métriques individuelles pour générer ainsi une nouvelle métrique composite plus efficace.

1.2 Évaluation de la segmentation en télédétection

Les images de télédétection ont des caractéristiques particulières. À très haute résolution spatiale notamment, la variabilité et la complexité des objets identifiables sont remarquables. La figure 1.3 illustre à la même échelle plusieurs exemples d'objets identifiables dans une même image THRS. Dans la figure 1.3a, on observe un cimetière composé de blocs réguliers, hétérogènes en termes de couleur mais très texturés. Dans la figure 1.3b, on observe une zone industrielle contenant des bâtiments de taille considérable, de forme rectangulaire et plutôt homogènes en termes de couleur. Dans la figure 1.3c, on observe un quartier résidentiel composé principalement de bâtiments de petite taille, de formes variables et des couleurs différentes : gris, marron et rouge brique principalement. Sur ces deux dernières images, on constate également la présence de quelques éléments de végétation isolés ainsi que des routes qui sont des objets très allongés. On y distingue même des voitures comme des petits objets de couleurs différentes que l'on pourrait presque confondre avec du bruit. La figure 1.3d montre une zone de forêt très dense. Il est impossible d'y identifier chaque arbre de manière indépendante, mais l'ensemble constitue une très grande zone ayant une couleur et une texture particulières. La figure 1.3e représente un plan d'eau, très homogène, de taille importante et avec des contours plutôt irréguliers. Enfin, la figure 1.3f présente un terrain sportif aux contours ronds. Aux alentours, on observe un grand nombre d'objets de petite taille tels que des arbres, des bâtiments ou des routes. Tous ces objets proviennent de la même image. Il est alors évident que la segmentation (et l'évaluation) de celle-ci est une tâche extrêmement difficile.

Par ailleurs, la taille des images peut atteindre plusieurs gigapixels. De plus, l'acquisition de celles-ci entraîne du bruit et donc de l'incertitude et de l'imprécision sur la totalité des objets. De ce fait, les objets d'intérêt n'ont pas de contours bien définis. Enfin, la notion de qualité varie d'un être humain à un autre. Tous ces facteurs rendent très difficile, voire impossible, une évaluation visuelle des résultats de manière systématique et rigoureuse. Ainsi, il est utopique de vouloir créer manuellement un jeu de données de référence fiable et complet pour appliquer des méthodes d'évaluation supervisées. Une possibilité pour contourner ce problème pourrait



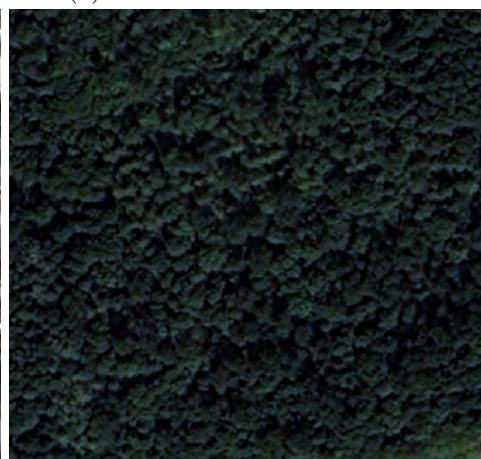
(a) Vue sur un cimetière.



(b) Vue sur une zone industrielle.



(c) Vue sur un quartier résidentiel.



(d) Vue sur une zone de forêt dense.



(e) Vue sur un plan d'eau.



(f) Vue sur un terrain sportif.

FIGURE 1.3 – Variabilité des objets identifiables dans les images de télédétection à THRS.

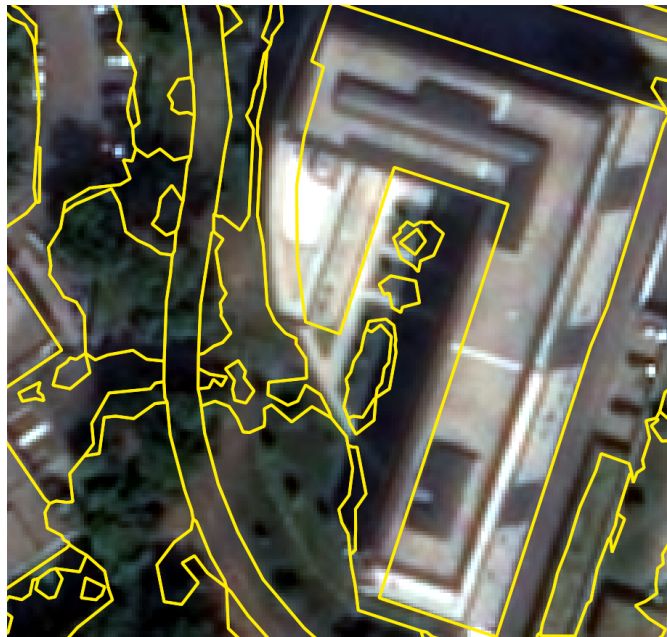


FIGURE 1.4 – Superposition d’une segmentation issue d’une base de données géographiques avec l’image correspondante.

consister à utiliser des bases de données existantes telles que OpenStreetMaps³ ou la BD TOPO⁴, comme données de référence. Néanmoins, celles-ci contiennent des nombreuses imprécisions et ne sont pas forcément à jour sur les régions d’intérêt, comme le montre la figure 1.4. On constate en particulier que les arbres qui dépassent sur la route n’apparaissent pas sur les contours de ces données ; un tel segment de route sera impossible à obtenir avec une approche de segmentation basée sur les propriétés radiométriques. On observe également que les contours du bâtiment ne sont pas correctement alignés avec les frontières de la segmentation.

Par ailleurs, les méthodes non-supervisées classiques ne semblent pas adaptées aux images en télédétection. En effet, elles se basent uniquement sur des critères d’homogénéité. Or, les objets présents dans les images THRS s’avèrent très hétérogènes.

Dans cette section, nous nous intéressons aux approches d’évaluation de la segmentation spécifiquement conçues en télédétection. Nous étudions les travaux qui existent sur la matière dans l’objectif de nous diriger vers la caractérisation des propriétés qui déterminent la bonne qualité d’une segmentation d’images de télédétection.

1.2.1 Évaluation supervisée en télédétection

[Corcoran et al., 2010] présente les caractéristiques que toute métrique supervisée devrait avoir pour être utilisable dans le cas de la télédétection :

- elle ne doit pas permettre des cas aberrants où une segmentation irréaliste obtient un score de qualité élevé ;

3. <http://www.openstreetmap.org/>

4. <http://professionnels.ign.fr/bdtopo>

- elle ne doit pas supposer que le nombre et la taille des objets sont les mêmes entre la segmentation à évaluer et la vérité-terrain ;
- elle devrait être tolérante aux raffinements uniquement dans les zones ambiguës pour les êtres humains ;
- elle doit donner des résultats comparables sur différentes segmentations d'une ou plusieurs images.

Il existe dans la littérature plusieurs cadres de travail pour évaluer de manière supervisée la qualité des algorithmes de segmentation et de leurs résultats dans le contexte de la télédétection. Dans [Zhan et al., 2005], les auteurs effectuent l'évaluation sur le résultat d'un extracteur d'objets d'une classe donnée, plutôt que sur une segmentation complète de l'image. L'évaluation est orientée par rapport à deux axes principalement :

- les erreurs de classification, mesurés en termes d'exactitude et d'exhaustivité ;
- les erreurs géométriques, mesurés en termes de positionnement et de mesure spatiale.

Pour évaluer les erreurs de classification, des approches classiques telles que les matrices de confusion et les concepts de précision et rappel, ainsi que le coefficient de Kappa sont employés. L'évaluation de la classification peut se faire au niveau des pixels : si un pixel classifié comme bâtiment se trouve dans un polygone de référence de cette classe, celui-ci est considéré comme correctement classifié. L'évaluation peut également se faire au niveau des objets, auquel cas il est nécessaire d'abord de déterminer la qualité des objets selon plusieurs critères tels que la position, la forme, et la taille. Le premier problème à résoudre est celui de mettre en correspondance les objets classifiés avec les polygones de référence, puisque généralement les objets ne correspondent pas exactement à la vérité-terrain et plusieurs cas de chevauchement peuvent se présenter. Les auteurs considèrent par exemple, qu'un bâtiment doit avoir une taille d'au moins 10m² et qu'un objet doit chevaucher au moins 50% de l'aire du polygone de référence.

Une métrique pour évaluer la qualité en termes de similarité de taille est donnée par :

$$Sim_size(O_i) = \frac{\min(Size_C(O_i), Size_R(O_i))}{\max(Size_C(O_i), Size_R(O_i))} \quad (1.11)$$

Une autre métrique pour évaluer la qualité en termes de position des objets est définie comme suit :

$$Q_Loc(O_i) = \sqrt{(X_C(O_i) - X_R(O_i))^2 + (Y_C(O_i) - Y_R(O_i))^2} \quad (1.12)$$

Elle correspond à la distance euclidienne entre les barycentres de l'objet classifié et celui de référence. La moyenne et l'écart type de ces métriques sur la totalité des objets sont considérés également comme mesures de qualité globale.

Une autre approche permettant d'ajouter une zone de tolérance aux contours des segments est proposée dans [Weidner, 2008]. Dans un premier temps, un ensemble de métriques couramment utilisées est présenté (Tableau 1.3). Soit S un ensemble de segments et R un ensemble de polygones de référence. Pour chaque polygone R_i , on construit les segments correspondants selon :

$$S_i = \bigcup_j S_j \quad \forall S_j \quad \text{avec} \quad \frac{|S_j \cap R_i|}{|S_j|} \geq 0.5 \quad (1.13)$$

Afin de pouvoir gérer l'imprécision des contours, un taux de qualité pondérée est défini par :

TABLEAU 1.3 – Métriques supervisées définies dans [Weidner, 2008].

Nom	Formule
Taux de détection	$\rho_d = \frac{ S \cap R }{ R }$
Taux de faux positifs	$\rho_{fp} = \frac{ S \setminus R }{ R }$
Taux de faux négatifs	$\rho_{fn} = \frac{ R \setminus S }{ R }$
Facteur de branchement	$\rho_b = \frac{ S \setminus R }{ S \cap R }$
Facteur d'échec	$\rho_m = \frac{ R \setminus S }{ S \cap R }$
Dissemblance de forme	$\rho_b = \frac{ (S \setminus R) \cup (R \setminus S) }{ R }$
Taux de qualité	$\rho_q = \frac{ S \cap R }{ S \cup R } = 1 - \frac{ (S \setminus R) \cup (R \setminus S) }{ R \cup S }$

$$\rho_{qw} = 1 - \frac{A}{|S \cap R| + A} \quad (1.14)$$

$$A = \sum_{x \in (S \setminus R)} w(d(x, R)) + \sum_{x \in (R \setminus S)} w(d(x, S))$$

où $d(x, \mathcal{A})$ mesure la distance la plus courte d'un point x à un ensemble de points \mathcal{A} , et w est une fonction de poids qui permet de tolérer plus ou moins les imperfections des segments selon leur distance au polygone de référence.

L'étude effectuée par [Möller et al., 2013] est très similaire. Les auteurs étudient la précision des objets classifiés (segments) d'un point de vue géométrique selon deux axes : recouvrement et positionnement par rapport aux polygones de référence. Ces métriques sont calculées selon les segments et selon les polygones de référence. Plusieurs configurations de ces métriques sont alors combinées par une moyenne géométrique. Ils effectuent l'évaluation de manière locale pour chaque objet de référence, mais également de manière globale en utilisant les moyennes et les médianes des différentes métriques. De plus, ils définissent un seuil sur la taille des objets participant à l'évaluation globale, ce qui permet d'avoir une estimation plus ou moins fine de la qualité en fonction des besoins.

De manière similaire, [Witharana et al., 2014] développe des métriques plus sophistiquées pour effectuer la mise en correspondance des segments avec des polygones de référence. Les auteurs définissent la sous- et sur-segmentation ainsi que la segmentation correcte en termes de relations plusieurs-à-un, un-à-plusieurs et un-à-un entre les segments et les polygones de référence. Ils considèrent 3 types de segment :

- objet candidat : un segment quelconque ;
- objet correspondant : un objet candidat chevauchant plus d'une certaine aire d'un objet de référence ;
- objet satisfaisant : un objet candidat entièrement contenu dans un objet de référence.

[Clinton et al., 2010] présente une méthode pour combiner plusieurs métriques supervisées. Contrairement à d'autres approches supervisées qui comparent deux partitions de l'image entière, les auteurs font l'évaluation uniquement par rapport à des objets d'intérêt délimités manuellement. Les auteurs présentent plusieurs métriques de base ainsi que plusieurs manières de les combiner.

Les résultats montrent que selon la métrique utilisée pour un même ensemble de segmentations, la « meilleure » segmentation peut varier drastiquement. Cela est dû probablement au fait

que chaque métrique mesure un aspect différent de la segmentation à la fois ; c'est pourquoi les auteurs recommandent d'implémenter un grand nombre de métriques, et de trouver des stratégies pour les combiner efficacement.

[Räsänen et al., 2013] présente une étude dont l'objectif est de vérifier à l'aide de métriques de qualité supervisée, si la meilleure segmentation selon un ou plusieurs critères conduit à une meilleure classification. Autrement dit, il s'agit de montrer l'existence d'une corrélation entre la qualité de la segmentation et celle de la classification. En effet, la classification finale peut être vue comme un moyen de déterminer la qualité de la segmentation, puisque les deux processus sont liés et que la meilleure segmentation doit être celle qui permet de mieux classer les différents objets.

D'après les auteurs, il y a plusieurs aspects qui affectent la qualité de la segmentation :

- l'algorithme employé ;
- le paramétrage de l'algorithme ;
- les différents poids attribués aux canaux (bandes radiométriques) de l'image ;
- la (re)classification des canaux ;
- les différentes transformations pouvant être appliquées aux canaux.

Un ensemble de segmentations est choisi d'après une évaluation subjective, et un autre en fonction des métriques supervisées. Les résultats montrent que selon le canal utilisé, différents types d'objet sont mieux segmentés, il en est de même pour plusieurs paramétrages. Après avoir effectué la classification des objets, les auteurs déduisent que les segmentations plus fines (tendant vers la sur-segmentation) conduisent à une classification plus précise. La précision de la classification a varié entre 60%, pour une segmentation bien notée par les métriques de qualité supervisée et 72%, pour la meilleure segmentation notée de manière subjective. Ils concluent qu'il n'existe pas de méthode unique permettant de déterminer la segmentation qui produira la meilleure classification. Il est donc nécessaire de bien définir à l'avance les propriétés qui vont déterminer la bonne qualité d'une segmentation dans un contexte applicatif donné. Dans cette étude, par exemple, les auteurs cherchaient à avoir des segments :

- significatifs mais simples ;
- ayant des frontières alignées avec les polygones de référence ;
- les plus grands possibles, tout en respectant les deux premières conditions.

Le tableau 1.4 résume les différentes métriques supervisées étudiées dans cette section. Il existe des solutions permettant d'évaluer des aspects différents de la segmentation. Ainsi, pour chaque application, il est important de définir les caractéristiques des objets que l'on cherche à segmenter et de trouver des métriques permettant de les quantifier au mieux.

TABLEAU 1.4 – Tableau de synthèse des métriques supervisées pour la télédétection.

Nom	Description	Avantages	Limites
Similarité de taille (Eq. 1.11)	Compare la taille de l'objet extrait avec l'objet de référence.	Affecte un score entre 0 et 1 si l'objet extrait est plus petit ou plus grand que celui de référence.	Ne s'applique qu'à un seul objet.

Nom	Description	Avantages	Limites
Similarité de localisation (Tableau 1.3)	Mesure l'erreur de localisation d'un objet avec la distance euclidienne des barycentres.		Ne s'applique qu'à un seul objet.
Taux de détection (Tableau 1.3)	Compte combien d'objets sont intersectés de plus de 50% par des segments.	Par définition borné entre 0 et 1.	Pas assez informatif : il faut le combiner avec d'autres indices.
Taux de faux positifs (Tableau 1.3)	Mesure le nombre de points qui n'intersectent pas d'objet de référence.	Par définition borné entre 0 et 1.	Pas assez informatif : il faut le combiner avec d'autres indices.
Taux de faux négatifs (Tableau 1.3)	Mesure le nombre de points appartenant aux objets de référence qui ne sont pas intersectés par un segment.	Par définition borné entre 0 et 1.	Pas assez informatif : il faut le combiner avec d'autres indices.
Facteur de branchement (Tableau 1.3)	Compte le nombre de points mal identifiés par rapport au nombre de points bien identifiés.	Par définition borné entre 0 et 1.	pas assez informatif : il faut le combiner avec d'autres indices.
Facteur d'échec (Tableau 1.3)	Compte le nombre de points de référence qui n'ont pas été identifiés par rapport au nombre de points de référence qui ont été identifiés.	Par définition borné entre 0 et 1.	Pas assez informatif : il faut le combiner avec d'autres indices.
Dissemblance de forme (Tableau 1.3)	Mesure la similarité de forme entre l'objet classifié et celui de référence.	Par définition bornée entre 0 et 1.	Pas assez informatif : il faut le combiner avec d'autres indices.
Taux de qualité (Tableau 1.3)	Compte le nombre de points qui intersectent effectivement un objet de référence.	Par définition bornée entre 0 et 1.	Pas assez informatif : il faut le combiner avec d'autres indices.

Nom	Description	Avantages	Limites
Taux de qualité imprécis (Eq. 1.14)	Ajoute une pondération en fonction de la distance du segment à l'objet de référence.	Permet de diminuer l'influence des imprécisions de l'image sur le résultat.	Il faut trouver la bonne fonction de poids et la bonne distance afin de ne pas biaiser les résultats.

1.2.2 Évaluation non-supervisée en télédétection

Les méthodes non-supervisées ont été étudiées dans une moindre mesure. Elles devraient posséder les deux caractéristiques suivantes [Corcoran et al., 2010] :

- travailler sur un ensemble de caractéristiques modélisant correctement la perception humaine pour identifier des objets ;
- prendre en compte le domaine spatial de l'image, puisque le système visuel humain segmente une image en évaluant les contrastes aux frontières des objets et ne tient pas compte des objets non adjacents.

Les auteurs évaluent deux métriques non-supervisées et les comparent avec une méthode supervisée afin d'estimer leur efficacité. La première métrique est un indice mesurant la validité d'un clustering. Elle travaille donc dans l'espace des données et ne tient pas compte du domaine spatial. La deuxième métrique proposée par les auteurs répond aux deux critères évoqués précédemment. Elle est définie par :

$$\begin{aligned}
 SU &= \frac{\textit{separation}}{\textit{cohesion}} \\
 \textit{cohesion} &= \sum_{i=1}^k \frac{\textit{objectSize}(i)}{\textit{totalArea}} \|\textit{Variance}(\textit{Features}(i))\| \\
 \textit{separation} &= \sum_{i=1}^k \frac{\textit{objectSize}(i)}{\textit{totalArea}} \sum_{j=1}^k \textit{neigh}(i, j) \frac{\textit{objectSize}(i)}{\textit{neighboursSize}(i)} \|\textit{Mean}(i) - \textit{Mean}(j)\| \\
 \textit{neigh}(i, j) &= \begin{cases} 1 & \text{si } i \text{ et } j \text{ sont des segments adjacents} \\ 0 & \text{sinon} \end{cases}
 \end{aligned} \tag{1.15}$$

où le terme de cohésion quantifie la qualité des segments en fonction de leur taille et de la variance des pixels. Le terme de séparation quantifie la qualité inter-région afin de tenir compte de l'aspect spatial. Ce terme est calculé en fonction de chaque segment et de son voisinage.

Une autre métrique non-supervisée est présentée dans [Zhang et al., 2012]. Elle est définie comme suit :

$$\begin{aligned}
 Z &= T + \lambda D \\
 T &= \frac{1}{10S} \sqrt{R} \sum_{i=1}^R \left(\frac{E_i}{1 + \log A_i} \right) \\
 D(I) &= \frac{\left(\sum_{i=1}^b \sum_{j=1}^R \frac{(m_{ij} - mm_i)}{R} \right)}{\sqrt{R}}
 \end{aligned} \tag{1.16}$$

où le terme T quantifie l'homogénéité des régions, S est la taille de l'image, R est le nombre de régions, E_i la somme des distances euclidiennes de chaque pixel avec la valeur moyenne de la

région i , A_i est l'aire de la région i et m_{ij} la valeur moyenne de la bande i dans la région j . D quantifie l'hétérogénéité entre les régions avec mm_i , la valeur moyenne de toutes les régions sur la bande i et b le nombre de bandes spectrales.

Le premier terme augmente lorsque le résultat est sur-segmenté, alors que le second augmente lorsque le résultat est sous-segmenté. Le poids λ permet donc de favoriser plus ou moins la sous-segmentation de l'image. La valeur de la segmentation optimale doit minimiser Z .

Les auteurs ont montré que la métrique se comporte de manière similaire à une métrique supervisée, ce qui est un bon indice sur la pertinence de la métrique. Cependant, elle ne tient compte que de l'aspect spectral de l'image, sans considérer la sémantique ni l'aspect spatial lié au segments.

[Johnson and Xie, 2011] présente une métrique de qualité non-supervisée et son application de manière locale pour déterminer des segments qui seraient sur- ou sous-segmentés, afin de les fusionner ou les segmenter plus finement. Elle comporte un terme quantifiant l'homogénéité intra-segment ($wVar$) et un terme quantifiant l'hétérogénéité inter-segment (MI) définis par :

$$wVar = \frac{\sum_{i=1}^n a_i \times v_i}{\sum_{i=1}^n a_i} \quad (1.17)$$

$$MI = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 \left(\sum_{i \neq j} \sum w_{ij} \right)}$$

Plus la variance est élevée, plus une région devrait être hétérogène. On cherche donc à obtenir un score faible pour la variance globale. Le terme d'hétérogénéité MI correspond à l'indice de Moran où \bar{y} est la valeur moyenne de l'image et w_{ij} est égal à 1 si et seulement si les régions i et j sont adjacentes ; sinon il est égal à 0. L'indice de Moran est peu élevé lorsque les régions sont hétérogènes.

Les deux termes sont normalisés entre 0 et 1 par $(X - X_{min}) / (X_{max} - X_{min})$ et finalement additionnés pour déterminer un score global qui doit être faible lorsque la segmentation est de bonne qualité.

Les métriques spécifiquement conçues pour des applications en télédétection étudiées dans ce document sont résumées dans le tableau 1.5. Ces métriques réalisent une évaluation de manière globale, ne sont pas bornées et sont difficiles à interpréter.

TABLEAU 1.5 – Tableau de synthèse des métriques non-supervisées pour la télédétection.

Nom	Description	Avantages	Limites
Spatial measure SU (Eq. 1.15)	Prend en compte l'aspect spatial pour mesurer l'homogénéité intra-région et l'hétérogénéité inter-région.	Mesure le contraste inter-région de manière locale (avec les voisins).	Tend à favoriser légèrement les sur-segmentations.

Nom	Description	Avantages	Limites
Z-measure (Eq. 1.16)	Quantifie la qualité en maximisant deux termes complémentaires limitant la sur- et la sous-segmentation.	Donne des résultats cohérents avec des métriques supervisées.	Ne tient pas compte des aspects sémantiques ou géométriques des segments.
Variance + Moran's Index (Eq. 1.17)	L'homogénéité est quantifiée par la variance et l'hétérogénéité par l'indice de Moran.	Donne des résultats consistants avec des métriques supervisées.	Le score est calculé sur chaque bande et la moyenne des scores est donnée comme résultat ; cela peut entraîner une perte de précision si les bandes ne sont pas également informatives.

1.3 UOA : une métrique quantifiant la sur- et la sous-segmentation

Comme nous avons vu dans les sections précédentes, les approches subjectives représentent une tâche non seulement coûteuse pour l'expert mais aussi entachée d'incertitude. Les approches supervisées requièrent des données de référence qui sont difficiles à obtenir en télédétection. Enfin, les approches non-supervisées existantes se limitent à une évaluation globale et difficile à interpréter.

Les hypothèses faites par les approches OBIA nous mènent à penser que la meilleure façon de s'attaquer à l'évaluation globale de la segmentation consiste alors à considérer la qualité locale de chaque segment. Dans cette section, nous proposons un nouvel indice non-supervisé permettant de quantifier de manière locale (par segment) le taux de sur- et sous-segmentation selon un critère d'homogénéité donné. Ainsi, nous définissons une approche d'évaluation locale permettant de rester dans un paradigme orienté objet.

Nous proposons également deux variantes permettant de quantifier la qualité de la segmentation de manière globale en agrégeant les scores de qualité locale.

1.3.1 Évaluation locale

Soit $S = \{R_i \mid 0 \leq i < M\}$ une segmentation composée de M segments. Pour chaque segment R_i de S , on définit $\mathcal{N}(R_i)$ comme l'ensemble des segments adjacents à R_i . Soit $H(R_i)$ une fonction à valeurs dans $[0, 1]$ et caractérisant l'homogénéité d'un segment R_i , où 0 correspond à une homogénéité parfaite et 1 correspond à une hétérogénéité totale. Soit δ un seuil sur les valeurs de $H(R_i)$. Nous définissons la fonction d'évaluation locale ϕ_δ comme suit :

$$\phi_\delta(R_i) = \begin{cases} -1 & \text{si } H(R_i) > \delta \\ 1 & \text{si } H(R_i) \leq \delta \text{ et } \exists R_j \in \mathcal{N}(R_i) \mid H(R_i \cup R_j) \leq \delta \\ 0 & \text{sinon} \end{cases} \quad (1.18)$$

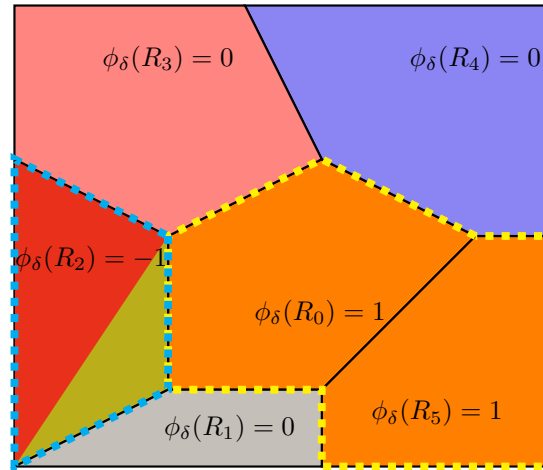


FIGURE 1.5 – Exemple illustrant l'approche d'évaluation locale. Les lignes pointillées en jaune montrent un exemple de sur-segmentation. Les lignes pointillées en bleu montrent un exemple de sous-segmentation.

Intuitivement, nous considérons qu'un segment R_i est de bonne qualité s'il est bien isolé par rapport à son voisinage en fonction de δ . On retrouve ainsi trois cas de figure possibles :

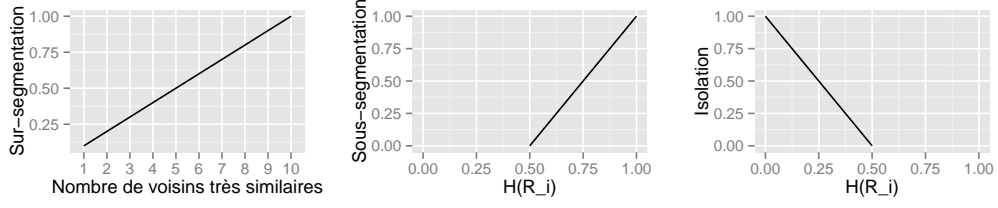
- R_i est trop hétérogène (i.e., $H(R_i) > \delta$) ; il est donc considéré comme sous-segmenté. Dans ce cas, le segment est pénalisé avec un poids négatif ;
- R_i est assez homogène (i.e., $H(R_i) \leq \delta$), mais il existe au moins un segment dans son voisinage tel que leur union serait elle aussi assez homogène (i.e., $\exists R_j \in \mathcal{N}(R_i), H(R_i \cup R_j) \leq \delta$) ; il est donc considéré comme sur-segmenté. Dans ce cas, le segment est pénalisé avec un poids positif ;
- R_i est assez homogène et il est bien isolé par rapport à son voisinage, c'est-à-dire : $\nexists R_j \in \mathcal{N}(R_i), H(R_i \cup R_j) \leq \delta$. Dans ce cas, aucune pénalité n'est appliquée.

La figure 1.5 illustre ce principe d'évaluation locale. Considérons un indice d'homogénéité H tel que $H(R_i) = 0$ si et seulement si tous les pixels de R_i ont la même valeur, et $H(R_i) = 1$ sinon. Définissons $\delta = 0.5$. Le segment R_2 n'est pas homogène car il contient des pixels de deux couleurs distinctes ; on a $H(R_2) = 1 > \delta$. R_2 est donc marqué comme sous-segmenté. R_0 et R_5 sont tous les deux considérés comme sur-segmentés. En effet, $H(R_0) = 0 < \delta$, mais également $H(R_0 \cup R_5) = 0 < \delta$. Le même raisonnement peut être appliqué réciproquement pour R_5 . En pratique, ces deux segments pourraient être fusionnés pour former un seul segment bien isolé. En suivant le même raisonnement, nous observons que les segments R_1 , R_3 , et R_4 sont bien isolés par rapport à leurs voisins.

La définition précédente de la fonction d'évaluation locale $\phi_\delta(R_i)$ considère la qualité des segments de manière rigide : un segment est soit sur-segmenté, soit sous-segmenté, soit bien isolé par rapport à ses voisins. Il est donc intéressant de mieux quantifier la qualité locale. Deux aspects, entre autres, peuvent alors être pris en compte :

- l'écart entre le taux d'homogénéité $H(R_i)$ et le seuil δ ;
- le voisinage du segment.

Ainsi, pour quantifier la sur-segmentation de manière plus fine, nous considérons le nombre de voisins de R_i qui lui sont très similaires en fonction de H et de δ :



(a) Taux de sur-segmentation avec $|neigh(R_i)| = 10$. (b) Taux de sous-segmentation avec $\delta = 0.5$. (c) Taux d'isolation du segment avec $\delta = 0.5$.

FIGURE 1.6 – Fonctions pour quantifier de manière fine la qualité locale.

$$over = \frac{N_v}{|\mathcal{N}(R_i)|} \quad (1.19)$$

$$N_v = |\{R_j \in \mathcal{N}(R_i) \mid H(R_i \cup R_j) \leq \delta\}|$$

où N_v est le nombre de segments voisins considérés comme trop similaires à R_i par rapport à δ . Dans la figure 1.6a, nous observons le comportement du taux de sur-segmentation pour un segment ayant 10 voisins. Nous remarquons que plus il y a des segments similaires dans le voisinage de R_i plus il va être pénalisé comme sur-segmentation.

De même, pour quantifier plus finement la sous-segmentation, nous considérons la distance entre la valeur de $H(R_i)$ et δ :

$$under = \frac{H(R_i) - \delta}{1 - \delta} \quad (1.20)$$

Dans la figure 1.6b, nous observons le comportement du taux de sous-segmentation avec $\delta = 0.5$. Nous remarquons que plus le segment est hétérogène, plus $H(R_i)$ est loin de δ et plus le segment est pénalisé comme sous-segmentation.

Enfin, pour mieux quantifier le taux d'isolation du segment, nous considérons le rapport entre $H(R_i)$ et δ :

$$isolation = 1 - \frac{H(R_i)}{\delta} \quad (1.21)$$

Dans la figure 1.6c, nous observons le comportement du taux d'isolation avec $\delta = 0.5$. Cette fonction se comporte de manière analogue à la fonction précédente. Nous remarquons que plus le segment est homogène, plus $H(R_i)$ est loin de δ et plus le segment est considéré comme bien isolé.

La fonction d'évaluation locale fine est alors définie par :

$$\phi_{\delta_{fine}}(R) = \begin{cases} under(R) & si \quad \phi_{\delta}(R) = -1 \\ over(R) & si \quad \phi_{\delta}(R) = 1 \\ 1 - isolation(R) & sinon \end{cases} \quad (1.22)$$

1.3.2 Évaluation globale

La qualité globale d'une segmentation peut-être vue comme une combinaison de la qualité de chacun des segments qui la composent. Nous proposons deux manières d'agrèger les différentes

mesures de qualité locale afin d'obtenir une estimation de qualité globale ; nous appelons ces métriques Under- and Over-segmentation Aware (UOA).

Nous définissons une première métrique globale, nommée UOA_{Σ} , comme la somme pondérée des scores de qualité locale, définie comme suit :

$$UOA_{\Sigma} = \sum_i \omega(R_i) \cdot \phi_{\delta}(R_i) \quad (1.23)$$

où ω est une fonction poids telle que pour tout R_i , on a $\omega(R_i) \geq 0$ et $\sum_i \omega(R_i) = 1$. Dans nos différentes expériences, nous considérons $\omega(R_i) = \frac{N_i}{N}$ avec N_i le nombre de pixels dans le segment R_i et N le nombre total de pixels de l'image. Autrement dit, chaque segment est pondéré proportionnellement à sa taille dans l'image. Ce choix permet de prendre en compte la diversité d'objets ayant des tailles différentes. En effet, les objets bien isolés ne sont pas pénalisés, quelle que soit leur taille. Les zones sur-segmentées sont formées par des regroupements de segments trop petits : l'apport des pénalisations de ces segments à la métrique est donc proportionnel à la taille de l'objet géographique sous-jacent. Similairement, les zones sous-segmentées contiennent, par hypothèse, plusieurs objets au sein d'un seul segment : l'apport de la pénalisation est proportionnel à la taille des objets sous-jacents.

Cette métrique fournit un indice direct sur le taux de sur- ou sous-segmentation dans l'image. Cependant, elle peut donner une estimation peu précise de la qualité lorsque les deux types d'erreurs sont présents dans des proportions similaires dans l'image, puisque les pénalisations auront tendance à s'annuler mutuellement. Ceci nous a mené à proposer une seconde variante. La deuxième métrique globale que nous avons définie quantifie dans un premier temps le taux de sous- et sur-segmentation de manière indépendante, et les considère ensuite comme un vecteur bi-dimensionnel. Le score final de cette métrique est alors la norme L_2 de ce vecteur. Ainsi, elle permet de pénaliser en même temps les deux types d'erreur. Elle est définie comme suit :

$$UOA_{L_2} = \sqrt{\Psi^2 + \Theta^2} \quad (1.24)$$

$$\Psi = \sum_{R_i | \phi_{\delta}(R_i) = -1} \omega(R_i)$$

$$\Theta = \sum_{R_i | \phi_{\delta}(R_i) = 1} \omega(R_i)$$

où Ψ et Θ représentent les taux de sous- et sur-segmentation respectivement.

1.3.3 Propriétés

Nous présentons maintenant quelques propriétés des métriques proposées :

- Les deux variantes UOA_{Σ} et UOA_{L_2} sont bornées, ce qui est une propriété très utile, et souvent manquante dans les métriques non-supervisées existantes. En effet, UOA_{Σ} varie de -1 (sous-segmentation forte) à 1 (sur-segmentation forte), et trouve sa valeur optimale en 0 . UOA_{L_2} varie de 0 (absence d'erreur) à 1 (tous les segments sont erronés) ;
- UOA_{Σ} est très expressive lorsque l'on évalue des segmentations très sur-segmentées (une valeur positive élevée) ou très sous-segmentées (une valeur négative élevée), mais elle peut donner des résultats inattendus si l'image a une proportion similaire de régions sur-segmentées et sous-segmentées. En revanche, UOA_{L_2} permet de trouver la segmentation avec le moins d'erreurs possibles, mais elle ne permet pas de savoir directement si le résultat est plutôt sur- ou sous-segmenté ;

- le choix du méta-paramètre H requiert un certain niveau d'expertise. En effet, il est crucial de choisir un indice d'homogénéité pertinent pour l'application en question. De plus, le choix de δ est également important puisqu'il dépend du comportement de H et doit être adapté à l'échelle de segmentation souhaité (cf. Section 1.4) ;
- la complexité algorithmique est dans le meilleur des cas $\mathcal{O}(M)$ avec M le nombre de segments dans la segmentation, lorsque la segmentation est totalement sous-segmentée. En moyenne, la complexité peut être approximée par $\mathcal{O}(MK)$ avec K le nombre moyen de voisins par segment. Néanmoins, K est généralement très petit par rapport à M et borné par une faible valeur ; ainsi nous pouvons considérer que la complexité est linéaire par rapport à M .

1.4 Étude expérimentale de UOA

Dans cette section nous validons notre proposition au moyen de plusieurs expériences. Dans un premier temps, nous présentons plusieurs indices d'homogénéité et étudions leur comportement de manière générale. Dans un deuxième temps, nous étudions le comportement des métriques proposées par rapport à plusieurs indices d'homogénéité, et à leur sensibilité au paramètre δ . Dans un troisième temps, nous menons une étude comparative avec deux métriques non-supervisées de l'état de l'art. Finalement, nous discutons de l'intérêt de disposer d'un indice sur la qualité locale à chaque segment.

1.4.1 Analyse des indices d'homogénéité

Par définition, UOA est une métrique générique qui autorise de choisir librement le critère d'homogénéité employé en fonction de l'application. En effet, l'indice d'homogénéité H est un méta-paramètre de UOA et il est possible d'utiliser une large gamme d'indices existants dans la littérature tels que l'angle spectral [Yang et al., 2014] ou des mesures d'homogénéité texturale basées sur les matrices de co-occurrence [Kekre et al., 2010], voire de nouveaux indices conçus explicitement pour une application donnée. Étant donné un indice d'homogénéité H , nous rappelons que $H(R_i)$ est égal à 0 lorsque le segment R_i est complètement homogène et une valeur positive non nécessairement bornée lorsque R_i est hétérogène. Intuitivement, un bon critère d'homogénéité devrait avoir un score faible pour les petits segments et des valeurs élevées pour les segments de grande taille. Plus particulièrement, le score de l'indice H en considérant l'image entière, devrait être la valeur maximale possible (il ne peut y avoir de segment plus hétérogène que l'image elle-même).

Nous avons étudié 5 indices d'homogénéité différents. Dans cette section, nous allons les introduire et les analyser de manière individuelle. Nous avons calculé chaque indice pour 174 151 segments provenant de 100 segmentations de qualité variée.

Dans un premier temps, nous observons le comportement de chaque indice H par rapport à la taille des segments. Ensuite, nous normalisons les scores obtenus dans l'intervalle $[0, 1]$ et nous observons le nombre de segments tels que $H(R_i) \leq \delta$ avec $0 < \delta < 1$. Cette étude nous permettra d'avoir une première intuition sur la manière de choisir le seuil δ nécessaire au calcul de UOA .

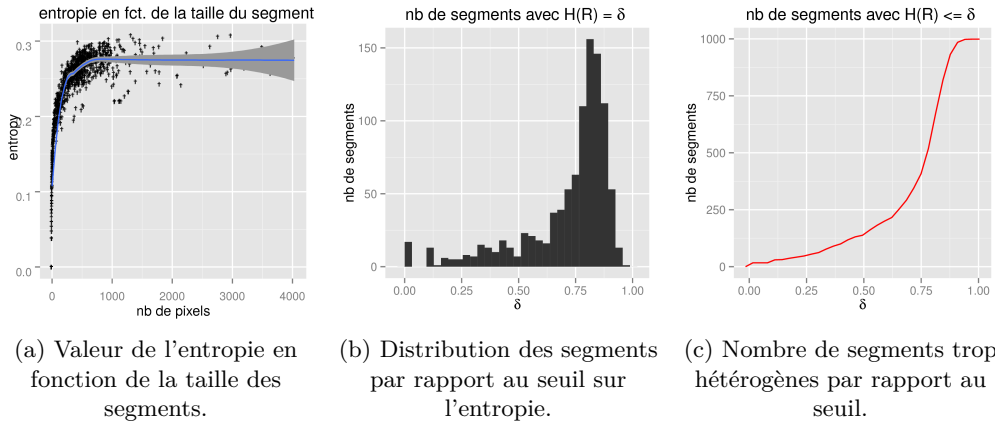


FIGURE 1.7 – Analyse de l'entropie.

1.4.1.1 Indice basé sur l'entropie (H_e)

Étant donnée une propriété P pouvant être mesurée sur l'ensemble des pixels d'une image prenant les valeurs t , l'entropie d'un segment est définie par :

$$H_e(R_i) = - \sum_t \frac{N_i(t)}{S_i} \log \frac{N_i(t)}{S_i} \quad (1.25)$$

où S_i est le nombre de pixels dans le segment R_i , $N_i(t)$ est le nombre de pixels ayant la valeur t dans le segment R_i et M_t est le nombre de valeurs possibles pour t dans le segment R_i . Le rapport $\frac{N_i(t)}{S_i}$ est la probabilité qu'un pixel ait la valeur t dans le segment i . Ainsi, lorsque le segment est complètement homogène $H_e(R_i) = 0$, puisque $M_t = 1$ et donc $N_i(t) = S_i$.

L'entropie des segments est calculée à partir des réponses radiométriques sur chaque bande de manière indépendante, puis la moyenne des entropies des bandes est calculée.

La figure 1.7a montre les valeurs d'entropie en fonction de la taille des segments (mesurée en nombre de pixels). On observe que l'entropie augmente très rapidement avec la taille des segments quand celle-ci est assez faible (< 1000 pixels), puis croît de manière plus lente lorsque la taille des segments devient importante.

La figure 1.7b montre la distribution des segments par rapport à l'entropie normalisée. On observe un déséquilibre marqué de cette distribution vers les valeurs comprises entre 0.8 et 0.9. La figure 1.7c montre la distribution cumulative des segments par rapport à l'entropie normalisée. On remarque que l'entropie est plus sensible au seuil dans l'intervalle $[0.6, 0.9]$: le nombre de segments dont l'entropie normalisée est inférieure à δ augmente très rapidement dans cet intervalle.

1.4.1.2 Indice basé sur le contraste des couleurs (H_c)

Le contraste sur une image en niveau de gris quantifie les différences entre les zones claires et les zones foncées de l'image. Il est défini par :

$$H_c = \frac{I_{max}(R_i) - I_{min}(R_i)}{I_{max}(R_i) + I_{min}(R_i)} \quad (1.26)$$

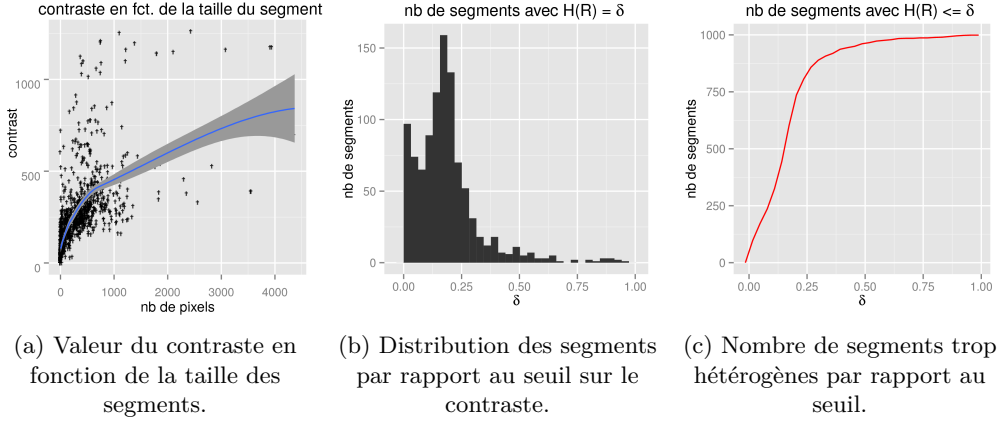


FIGURE 1.8 – Analyse du contraste.

avec $I_{max}(R_i)$ et $I_{min}(R_i)$ les intensités (réponses radiométriques dans notre cas) maximale et minimale dans le segment. Cet indice prend ses valeurs entre 0 et 1. En effet, lorsque le segment est complètement homogène, $I_{max}(R_i) = I_{min}(R_i)$, donc $C(R_i) = 0$. Au contraire, dans le pire des cas, $I_{min}(R_i) = 0$ et $I_{max}(R_i) = I_{max}(I)$ donc $I_{max}(I) - I_{min}(R_i) = I_{max}(R_i) + I_{min}(R_i)$ ainsi $C(R_i) = 1$.

La figure 1.8a montre les valeurs du contraste en fonction de la taille des segments (mesurée en nombre de pixels). On observe que le contraste augmente presque linéairement avec la taille des segments. Lorsque la taille des segments est très importante on observe l'apparition de points dont le contraste est remarquablement bas par rapport au reste.

La figure 1.8b montre la distribution des segments par rapport au contraste normalisé. On observe un déséquilibre marqué de cette distribution vers les valeurs comprises entre 0.0 et 0.25. La figure 1.8c montre la distribution cumulative des segments par rapport au contraste normalisé. Similairement, on remarque que le contraste est très sensible dans l'intervalle $[0.0, 0.25]$.

1.4.1.3 Indice basé sur CIELab (H_{CLAB})

L'espace de couleurs CIELab a la propriété de bien modéliser la perception des couleurs du système visuel humain. En effet, la distance euclidienne mesurée dans cet espace de données correspond bien à la différence perçue entre deux couleurs, ce qui n'est pas toujours le cas dans RGB. L'espace CIELab possède 3 composantes :

- L : correspond à la luminance et varie entre 0 et 100 ;
- a* : correspond à la quantité de vert (-127) ou de rouge (+128) dans l'image ;
- b* : correspond à la quantité de bleu (-127) ou de jaune (+128) dans l'image.

Nous définissons le contraste dans l'espace CIELab comme la distance euclidienne (Δ_{E^*}) entre les couleurs les plus distantes dans le segment, soit :

$$H_{CLAB}(R_i) = \Delta_{E^*}(max_{lab}(R_i), min_{lab}(R_i))$$

$$\Delta_{E^*}(c_1, c_2) = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \quad (1.27)$$

Lorsque le segment est complètement homogène, on a $C_{lab}(R_i) = 0$ et dans le pire des cas, $C_{lab}(R_i)$ est égal à 374, 232. Néanmoins dans la pratique la valeur maximale de cet indice est aux alentours de 52.

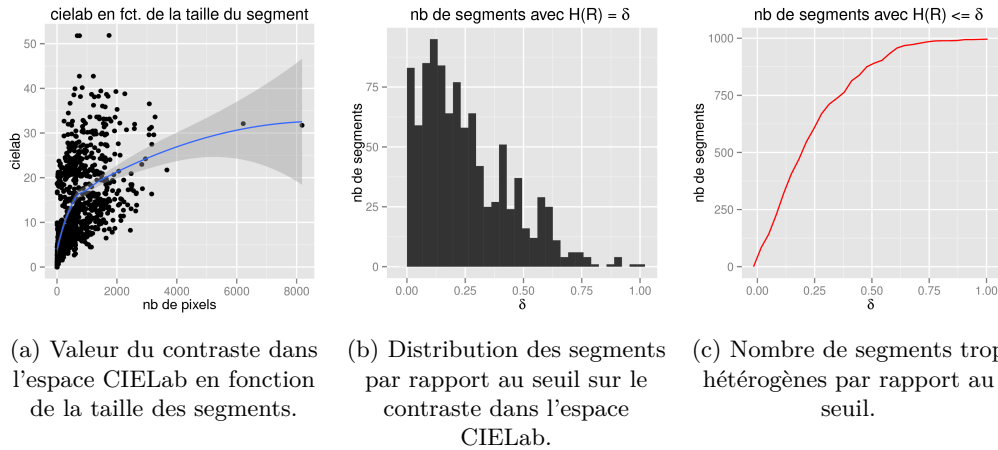


FIGURE 1.9 – Analyse du contraste dans l'espace CIELab.

Cette indice n'est pas directement applicable aux images de télédétection. En effet, l'espace CIELab a été conçu pour des images couleur RGB « classiques ». Or, une image Pleiades par exemple, possède 4 bandes spectrales qui sont encodées en 64 bits. Afin d'appliquer cette mesure, il faut notamment conserver uniquement 3 bandes spectrales (rouge, vert, bleu en l'occurrence) et les quantifier en 256 niveaux chacune des bandes. Finalement, on calcule la transformation de l'espace RGB vers CIELab pour effectuer le calcul de cet indice.

La figure 1.9a montre les valeurs du contraste dans l'espace CIELab en fonction de la taille des segments (mesurée en nombre de pixels). On observe que l'indice augmente rapidement avec la taille des segments lorsque celle-ci est inférieure à 1 000 pixels, puis continue d'augmenter plus lentement lorsque la taille des segments devient plus importante.

La figure 1.9b montre la distribution des segments par rapport au contraste dans l'espace CIELab normalisé. On remarque que la distribution est déséquilibrée vers les valeurs comprises entre 0.0 et 0.5. La figure 1.8c montre la distribution cumulative des segments par rapport au contraste dans l'espace CIELab normalisé. On constate que l'indice est sensible dans l'intervalle [0.0,0.5].

1.4.1.4 Indice de cohésion (H_{co})

La cohésion a été définie dans [Corcoran et al., 2010], et correspond à la norme du vecteur des variances des caractéristiques du segment :

$$H_{co}(R_i) = \frac{N_i}{N} \|V(R_i)\|$$

$$\|v\| = \sqrt{\sum_{i=0}^k v(i)^2}$$
(1.28)

où N_i est la taille du segment, N est la taille de l'image et $V(R_i)$ est un vecteur contenant la variance des pixels de chaque bande dans le segment. Lorsque le segment est complètement homogène, la cohésion est égale à 0. Lorsque le segment est très hétérogène, sa valeur est très haute mais il est difficile d'établir une borne supérieure puisque cela dépend de l'image.

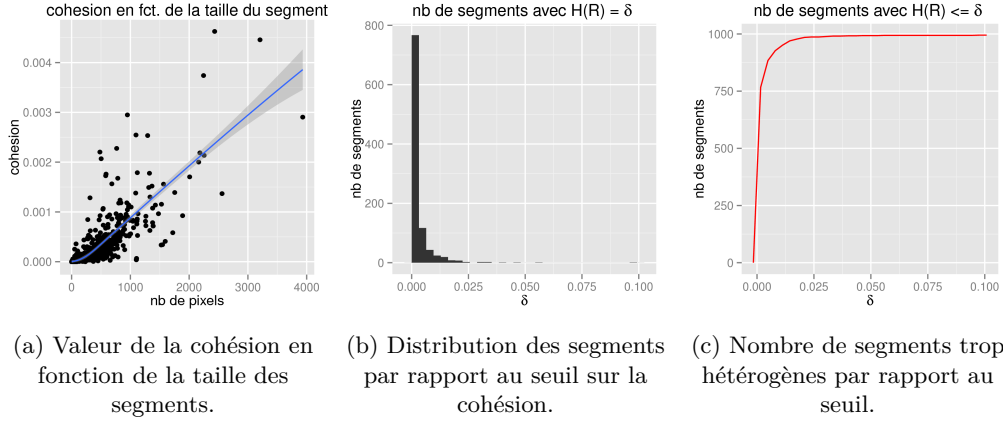


FIGURE 1.10 – Analyse de la cohésion.

La figure 1.10a montre les valeurs de la cohésion en fonction de la taille des segments (mesurée en nombre de pixels). On observe que l'indice augmente linéairement avec la taille des segments.

La figure 1.10b montre la distribution des segments par rapport à la cohésion normalisée. On remarque un déséquilibre très marqué vers les valeurs comprises entre 0.0 et 0.01. En effet, les valeurs de la cohésion sont très proches de 0 et ont une variance très faible. La figure 1.10c montre la distribution cumulative des segments par rapport à la cohésion normalisée. On observe que cet indice est trop sensible dans l'intervalle $[0.0, 0.025]$. En effet, la presque totalité des segments ont une cohésion inférieure à 0.025.

1.4.1.5 Indice basé sur la variance H_v

L'indice basée sur la variance est la moyenne des variances des pixels dans chaque bande du segment. Il est défini par :

$$H_v(R_i) = \frac{1}{b} \sum_{j=0}^b \sigma(R_{i_b})^2 \quad (1.29)$$

$$\sigma(R_{i_b})^2 = \frac{1}{N_i} \sum_{p \in R_i} (R_{i_b}(p) - \text{mean}(R_{i_b}))^2$$

où R_{i_b} est la bande radiométrique b et N_i est la taille du segment. Lorsque le segment est complètement homogène, la variance est égale à 0, et lorsque le segment est très hétérogène sa valeur est très élevée ; cependant il est difficile d'établir une borne supérieure puisque cela dépend de l'image.

La figure 1.11a montre les valeurs de la variance en fonction de la taille des segments (mesurée en nombre de pixels). On observe que lorsque la taille des segments est inférieure à 2 000 pixels, l'indice est très variable ; on perçoit tout de même une tendance croissante.

La figure 1.11b montre la distribution des segments par rapport à la variance normalisée. On remarque un déséquilibre marqué vers les valeurs comprises entre 0.0 et 0.2. La figure 1.10c montre la distribution cumulative des segments par rapport à la variance normalisée. On observe que cet indice est trop sensible dans l'intervalle $[0.0, 0.2]$: la majorité des segments présente une variance inférieure à 0.2.

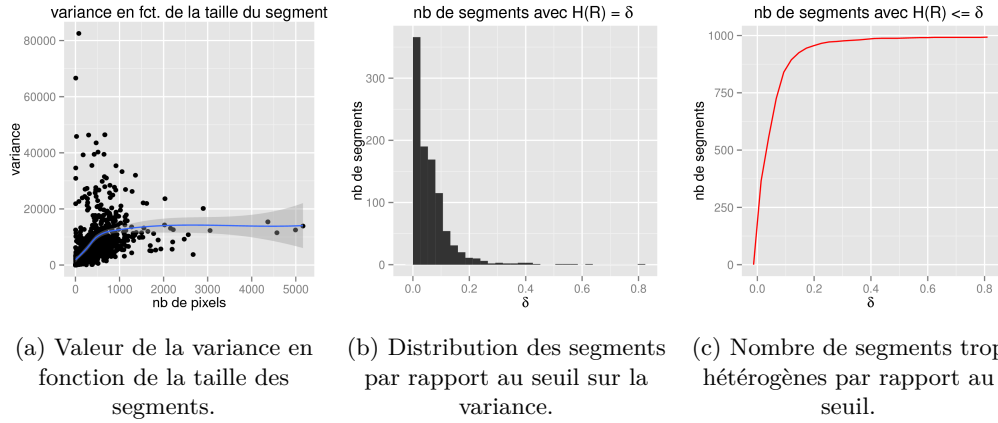


FIGURE 1.11 – Analyse de la variance.

1.4.1.6 Indice combiné H_μ

Les distributions des segments par rapport aux indices introduits précédemment sont toutes asymétriques et se comportent de manières différentes. Afin de réduire le biais induit par chacun de ces indices, il est possible de calculer leur moyenne et de considérer cette valeur comme un nouvel indice d'homogénéité. Dans un premier temps, les valeurs de chaque indice sont normalisées comme suit :

$$\begin{aligned}
 \min &= \min_{S_i \in S} H(S_i) \\
 \max &= \max_{S_i \in S} H(S_i) \\
 \text{normalize}(S_i, H) &= \frac{H(S_i) - H(\min)}{H(\max) - H(\min)}
 \end{aligned} \tag{1.30}$$

où S est une segmentation, S_i le segment évalué et H un indice d'homogénéité.

Ensuite, la moyenne des indices est calculée et l'erreur-type est additionnée afin de pénaliser davantage les segments hétérogènes. L'indice combiné est donc donné par :

$$\begin{aligned}
 H_\mu(S_i) &= \mu(S_i) + \frac{\sigma(S_i)}{\sqrt{n}} \\
 n &= \text{card}(\mathbb{H}) \\
 \mu(S_i) &= \frac{1}{n} \sum_{H_j \in \mathbb{H}} \text{normalize}(S_i, H_j) \\
 \sigma(S_i) &= \sqrt{\frac{1}{n} \sum_{H_j \in \mathbb{H}} (\text{normalize}(S_i, H_j) - \mu)^2}
 \end{aligned} \tag{1.31}$$

où \mathbb{H} est l'ensemble des indices d'homogénéité à employer.

La figure 1.12a montre les valeurs de l'indice combiné en fonction de la taille des segments (mesurée en nombre de pixels). On observe que les valeurs de l'indice augmentent rapidement avec la taille des segments lorsque celle-ci est inférieure à 1000 pixels, puis continue d'augmenter plus lentement quand la taille devient de plus en plus grande.

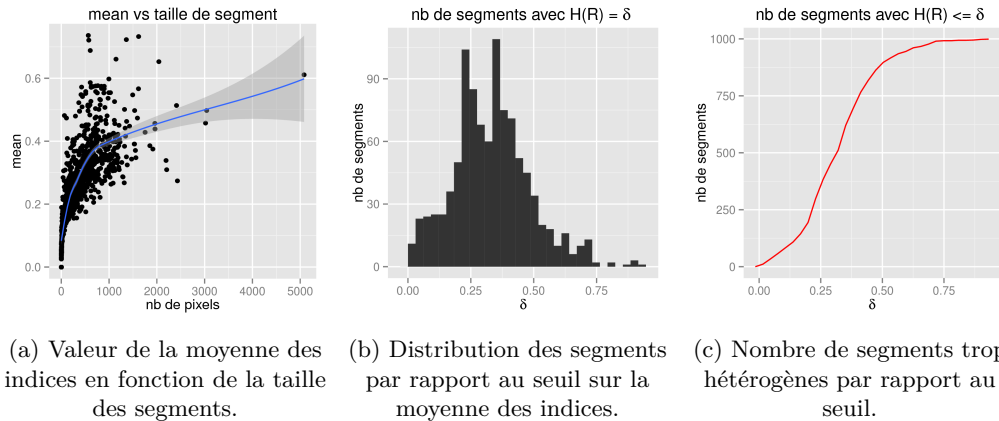


FIGURE 1.12 – Analyse de la moyenne des indices.

TABLEAU 1.6 – Corrélation entre les indices d'homogénéité et la taille des segments.

Indice	Pearson	Kendall	Spearman
H_e	0.560	0.780	0.926
H_c	0.562	0.635	0.819
$H_{C_{LAB}}$	0.559	0.552	0.732
H_{cO}	0.810	0.768	0.922
H_v	0.350	0.529	0.717
H_μ	0.648	0.685	0.859

La figure 1.12b montre la distribution des segments par rapport à l'indice combiné. Cette fois-ci, on observe une distribution plus proche d'une gaussienne avec une moyenne de 0.4 et un écart type de 0.15. La figure 1.12c montre la distribution cumulative des segments par rapport à l'indice combiné. On observe que le cumul augmente de manière presque linéaire entre 0.0 et 0.5, puis continue de croître plus lentement entre 0.5 et 0.8.

Comme nous le verrons dans la section suivante, cet indice nous semble le plus pertinent pour choisir δ de manière intuitive. En effet, dans une segmentation de bonne qualité, il est peu probable d'avoir des segments très homogènes correspondant à des segments très petits ou à des grandes zones lisses, ainsi que des segments très hétérogènes correspondant à de très grands segments.

Afin de vérifier les observations que nous avons faites à propos du comportement des indices d'homogénéité par rapport à la taille des segments, nous avons calculé 3 mesures de corrélation (Pearson, Kendall et Spearman) entre ces deux variables pour chaque indice. Le tableau 1.6 confirme qu'il existe bien une corrélation entre les variables. En particulier, le coefficient de Spearman qui est adapté aux distributions asymétriques donne des scores de corrélation assez importants. Les deux autres indices trouvent également une corrélation positive bien que les valeurs soient plus faibles.

FIGURE 1.13 – Image employée pour l'étude de *UOA*.

1.4.2 Analyse de sensibilité

Le score des métriques UOA_{Σ} et UOA_{L_2} dépend directement de l'indice d'homogénéité H et du seuil δ choisi. Nous avons effectué une série d'expériences pour mettre en évidence l'influence de ces deux paramètres sur UOA. Pour cela nous avons généré 100 segmentations Mean-shift en variant les paramètres, afin d'obtenir un ensemble de segmentations de qualité variable allant de très sur-segmentées à très sous-segmentées d'après une vérification visuelle subjective. L'image employée pour cette expérience est un extrait de 1024×1024 pixels d'une image Pléiades de la zone urbaine de Strasbourg contenant des objets de types divers (voir figure 1.13). Elle contient notamment, des bâtiments de tailles variées, des routes et quelques zones de végétation.

1.4.2.1 UOA et le paramètre H

Nous étudions le comportement de *UOA* par rapport aux 6 indices d'homogénéité présentés dans la section 1.4.1. Nous avons calculé chaque indice H pour tous les segments et nous avons observé la variation des taux de sur- et sous-segmentation en fonction de δ . Les figures 1.14 et 1.15 illustrent cette analyse de sensibilité pour les 6 indices d'homogénéité choisis : les graphes représentent δ en abscisse et la sur-/sous-segmentation en ordonnée. Chaque graphe montre les mesures faites pour 5 segmentations choisies au hasard parmi les 100 segmentations générées.

Nous remarquons que chaque indice possède un comportement différent, ce qui corrobore les analyses effectuées précédemment. En effet, chaque indice est fortement sensible aux variations de δ dans des intervalles différents. Le tableau 1.7 résume ces observations. Pour chaque indice, nous présentons l'intervalle (estimé empiriquement) sur lequel l'indice est le plus sensible.

Par ailleurs, la figure 1.15 montre le taux de segments jugés comme bien isolés. La valeur maximale correspond, comme attendu, au point de croisement entre la sur- et la sous-segmentation. Remarquons que le meilleur taux de segments bien isolés obtenu sur la totalité des segmentations est de 0.6 environ. En moyenne celui-ci se trouve autour de 0.4. Cela montre que même dans le meilleurs de cas, il reste encore des erreurs de sur- ou sous-segmentation. Au demeurant, nous pouvons noter que le comportement de *UOA* varie plus ou moins en fonction de l'indice

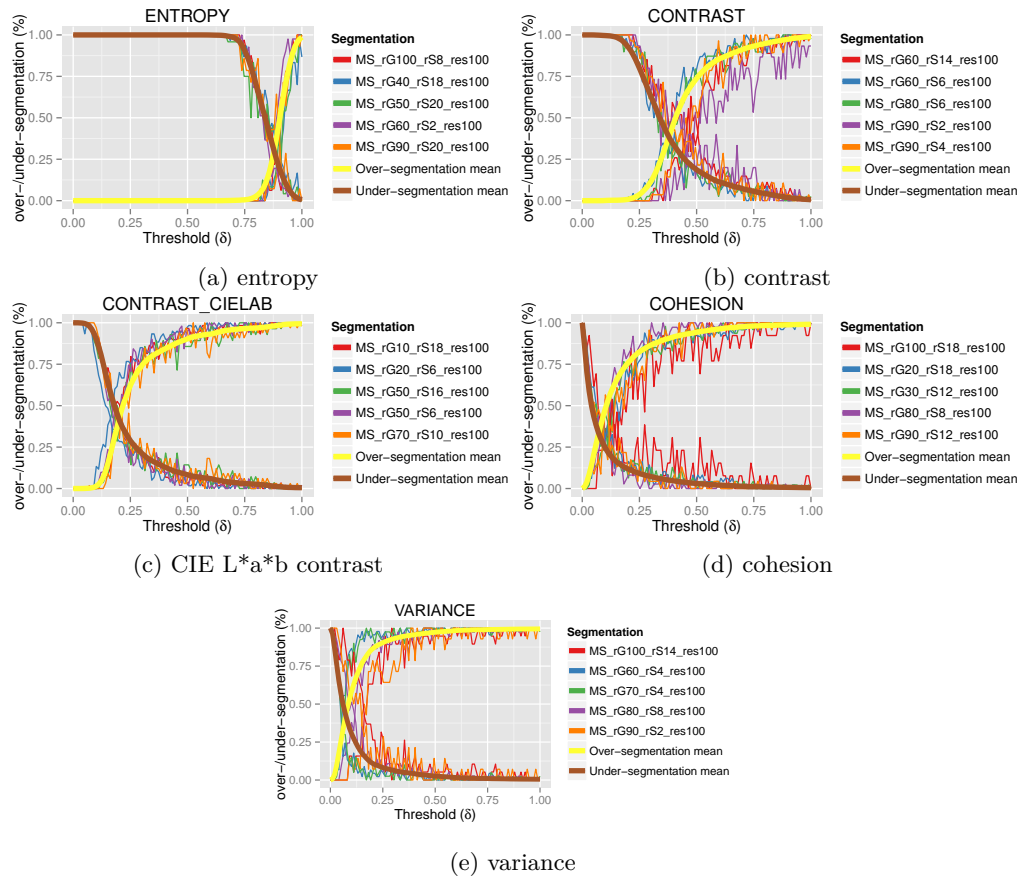


FIGURE 1.14 – Analyse de sensibilité de Ψ et Θ en fonction de δ pour les 5 indices d'homogénéité testés. Les segmentations sont labellisées selon les paramètres employés sur l'algorithme Mean-shift : rang (rG), rayon spatial (rS) taille minimum (res). La figure illustre également la moyenne calculée sur les 100 segmentations.

TABLEAU 1.7 – Sensibilité des indices en fonction de δ .

Indice	Intervalle
Entropie (Eq. 1.25)	[0.7, 0.9]
Contraste (Eq. 1.26)	[0.2, 0.5]
Contraste CIE Lab (Eq. 1.27)	[0.1, 0.3]
Cohésion (Eq. 1.28)	[0.0, 0.2]
Variance (Eq. 1.29)	[0.0, 0.2]
Combiné (Eq. 1.31)	[0.2, 0.7]

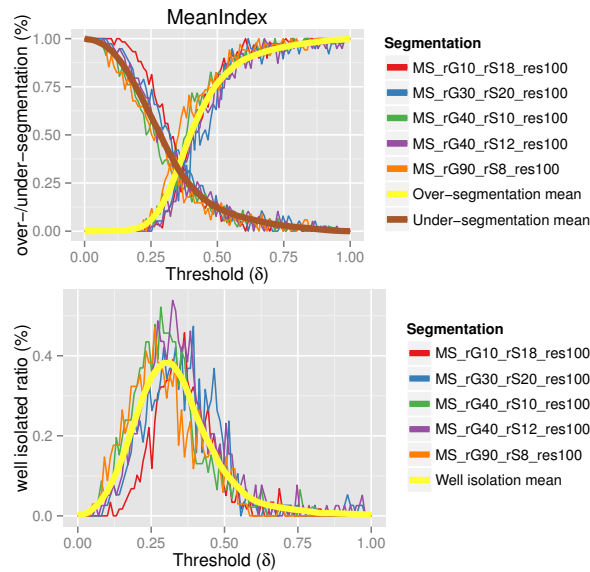


FIGURE 1.15 – Analyse de sensibilité de Ψ , Θ , et le taux de segments bien isolées en fonction de δ pour la moyenne des 5 indices testés.

H choisi. Enfin, le seuil δ doit être déterminé en fonction de l'indice d'homogénéité, puisque sa valeur optimale dépend de celui-ci.

1.4.2.2 UOA et le paramètre δ

Pour cette expérience, nous utilisons l'indice d'homogénéité combiné (Eq. 1.31). En effet, cet indice a un comportement plutôt stable et permet de mieux étudier le comportement de UOA par rapport à δ en évitant le biais introduit par le choix de l'indice d'homogénéité. Nous calculons UOA_{Σ} et UOA_{L_2} pour δ allant de 0 à 1 par pas de 0.01, et nous obtenons ainsi 10 000 observations. Nous conservons également les taux de sur-segmentation (Θ) et sous-segmentation (Ψ), ainsi que le taux de segments jugés bien isolés, qui est donné par :

$$UOA_{ok} = 1 - (\Theta + \Psi) \quad (1.32)$$

La figure 1.16 présente l'évolution de la sous-segmentation en fonction de la sur-segmentation pour 10 segmentations (toutes confondues) choisies au hasard. Chaque point correspond à l'évaluation d'une segmentation en employant une valeur de δ donnée. Celle-ci est donnée par la couleur : rouge pour δ proche de 0, jaune pour δ proche de 0.5 et bleu pour δ proche de 1. Les points mis en valeur correspondent aux évaluations optimales trouvées en utilisant les métriques UOA_{L_2} , UOA_{ok} , et UOA_{Σ} (UOA_{norm} , UOA_{ok} et UOA_{σ} respectivement sur la figure 1.16) de chaque segmentation. Les valeurs précises de ces métriques ainsi que la valeur de δ utilisée sont données dans le tableau 1.8. Rappelons que les valeurs optimales doivent minimiser la valeur absolue de UOA_{Σ} ainsi que UOA_{L_2} et maximiser UOA_{ok} . Nous remarquons les faits suivants :

- UOA_{Σ} a une valeur optimale de δ entre 0.30 et 0.42 ; en moyenne les taux de sur- et sous-segmentation sont de 0.42 ;
- UOA_{L_2} a des valeurs optimales de δ entre 0.30 et 0.43 ;
- UOA_{Σ} et UOA_{L_2} semblent avoir des comportements légèrement différents ;

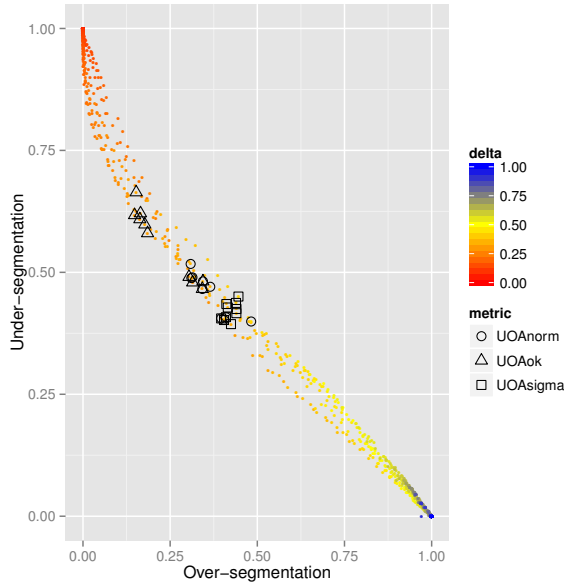


FIGURE 1.16 – Sous- vs. sur-segmentation avec δ allant de 0 à 1, pour 10 segmentations sélectionnées aléatoirement.

- UOA_{L_2} possède en général une estimation du taux de segments bien isolés supérieur à celui de UOA_{Σ} . Cela peut s'expliquer par le fait que UOA_{Σ} écrase l'information en faisant la moyenne de la sous- et la sur-segmentation ;
- UOA_{ok} a des valeurs optimales de δ entre 0.28 et 0.35. Le taux de segments bien isolés est supérieur à ceux trouvés par les autres métriques bien que le taux de sous-segmentation soit plus important.

Par ailleurs, nous observons une forte corrélation entre δ , le taux de sur-segmentation et le taux de sous-segmentation. En effet, lorsque δ augmente, le taux de sur-segmentation détecté par UOA augmente et le taux de sous-segmentation diminue.

Globalement, nous observons que nos métriques présentent un comportement cohérent et permettent effectivement de déterminer des valeurs optimales de δ qui résultent en un bon compromis entre les deux types d'erreur.

1.4.3 Étude comparative avec des méthodes existantes

Nous comparons UOA_{Σ} et UOA_{L_2} avec deux métriques présentées dans la Section 1.2.2 : la métrique Z [Zhang et al., 2012] et la métrique SU [Corcoran et al., 2010]. Pour cette expérience nous avons fixé la valeur de $\delta = 0.35$ pour le calcul de UOA_{Σ} et UOA_{L_2} . Cette valeur est l'une des valeurs optimales trouvées lors de l'expérience présentée dans la section 1.4.2.2 (voir tableau 1.8). Nous avons évalué et classé les 100 segmentations Mean-shift.

Les figures 1.17a, 1.17c, 1.17e et 1.18a, montrent les valeurs des métriques testées obtenues pour toutes les segmentations. L'axe des x représente les 100 segmentations triées par leur nombre de segments, qui varie de 979 à 4 555. L'axe des y représente les valeurs de chaque métrique. Il est à noter que les métriques Z et SU ne sont pas bornées. Ainsi elles ont été normalisées entre 0 et 1 afin de permettre une comparaison objective de celles-ci. De plus, les valeurs résultantes ont été modifiées pour faire en sorte que 0 signifie basse qualité et 1 signifie haute qualité.

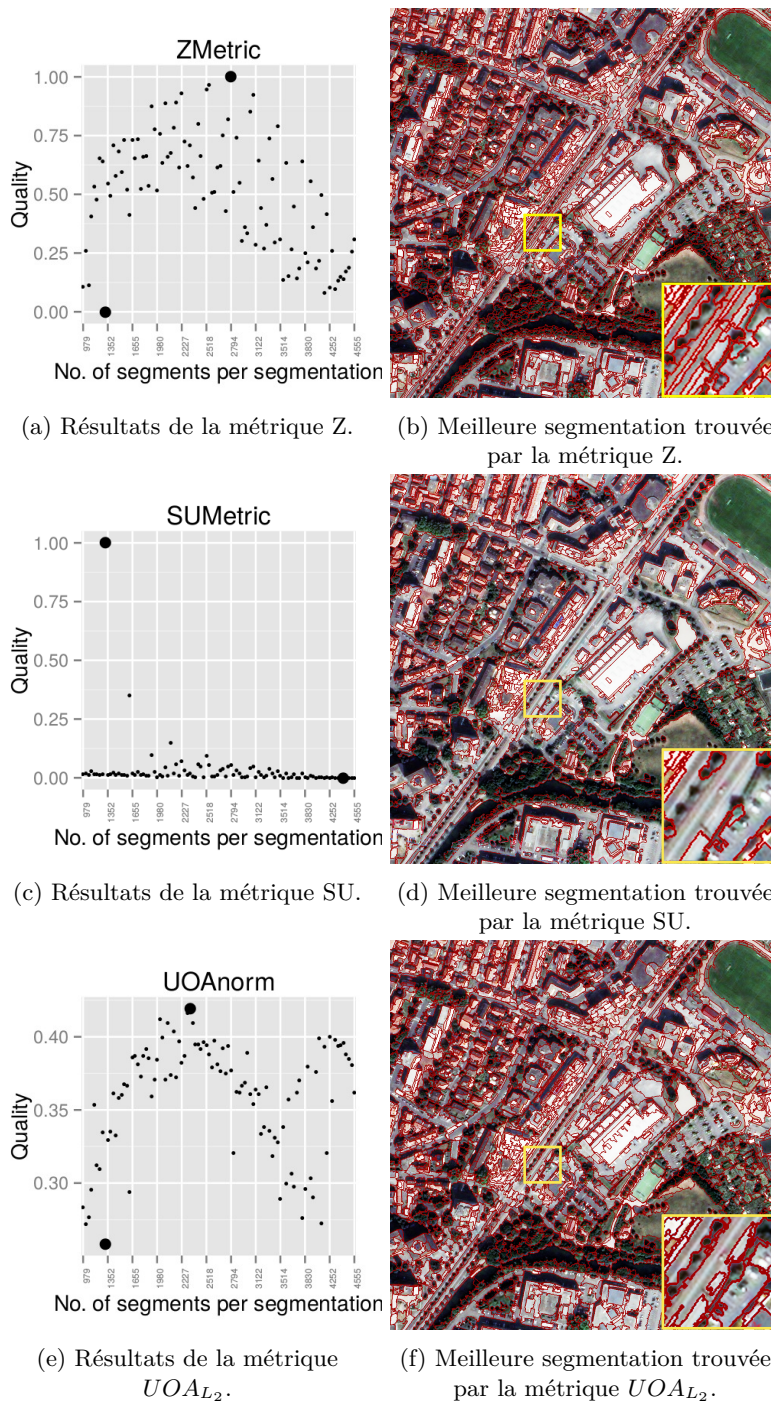


FIGURE 1.17 – Meilleures segmentations classées par les métriques Z, SU et $UOAL_2$. À gauche : Les scores de chaque métrique en fonction du nombre de segments ; plus le score est élevé, meilleure est la qualité. Les gros points correspondent aux scores maximaux et minimaux. À droite : l'image correspondant à la meilleure segmentation trouvée par chaque métrique.

TABLEAU 1.8 – Résultats des métriques agrégées sur 10 segmentations choisies aléatoirement.

metric	aggregate	δ	UOA_{Σ}	Θ	Ψ	UOA_{L_2}	UOA_{ok}
UOA_{Σ}	Min.	0.300	-0.023	0.396	0.393	0.567	0.103
	Median	0.365	0.001	0.419	0.420	0.597	0.155
	Mean	0.360	0.002	0.422	0.420	0.596	0.157
	Max.	0.420	0.030	0.445	0.450	0.633	0.198
UOA_{L_2}	Min.	0.300	-0.208	0.308	0.399	0.567	0.118
	Median	0.345	-0.114	0.354	0.468	0.580	0.186
	Mean	0.348	-0.085	0.367	0.452	0.587	0.179
	Max.	0.430	0.083	0.482	0.517	0.626	0.200
UOA_{ok}	Min.	0.280	-0.511	0.148	0.465	0.573	0.174
	Median	0.310	-0.407	0.181	0.589	0.616	0.210
	Mean	0.314	-0.331	0.229	0.560	0.614	0.209
	Max.	0.350	-0.123	0.344	0.664	0.681	0.234

En nous basant sur les résultats de cette expérience, nous pouvons faire les remarques suivantes :

- La métrique SU présente quelques valeurs aberrantes (Figure 1.17c). Cela pourrait s'expliquer par le fait que l'ensemble de propriétés employé pour calculer la métrique n'est pas explicitement donné dans l'article d'origine. Nous avons employé uniquement les valeurs des bandes radiométriques pour faire ce calcul. Une inspection visuelle du meilleur résultat obtenu révèle qu'il s'agit effectivement d'un cas de sous-segmentation. Nous concluons que cette métrique n'est pas utilisable de manière générale pour tout type de problème en télédétection.
- Les métriques UOA_{L_2} et Z trouvent des segmentations optimales ayant un nombre de segments proche : 2 291 et 2 781 segments respectivement (Figures 1.17e et 1.17a). Néanmoins, la métrique Z semble moins robuste que UOA_{L_2} . En effet, elle donne des valeurs très différentes à des segmentations qui sont proches en termes de nombre de segments (e.g., 0.45 pour une segmentation avec 2 694 segments et 0.80 pour une segmentation avec 2 736 segments). Une inspection visuelle des meilleures segmentations trouvées par ces deux métriques montre que les deux sont effectivement des solutions correctes, mais la segmentation trouvée par UOA_{L_2} possède des frontières plus régulières (par exemple le stade en haut et à droite de l'image, ou la route dans la zone zoomée). En effet les bords irréguliers qui se trouvent dans la segmentation trouvée par la métrique Z sont composés par plusieurs petits segments qui ne sont pas pertinents ; ces petits segments font augmenter le taux de sur-segmentation de UOA_{L_2} , ce qui résulte en une pénalisation de cette segmentation et donc une meilleure estimation de sa qualité.

La figure 1.18a, montre les résultats de l'évaluation avec UOA_{Σ} . Remarquons que pour cette métrique nous avons conservé les valeurs telles quelles. En effet, l'un des principaux avantages de UOA_{Σ} est la capacité d'indiquer des résultats fortement sur-segmentés (valeurs proches de 1) et fortement sous-segmentés (valeurs proches de -1). Le graphe montre comme UOA_{Σ} augmente linéairement lorsque le nombre de segments augmente. Les figures 1.18b et 1.18c montrent les résultats les plus sur- et sous-segmentés. Une inspection visuelle permet de corroborer que UOA_{Σ} permet effectivement d'identifier les deux types de résultats erronés.

Pour valider davantage notre proposition, nous avons répété l'expérience avec une image plus complexe de $4\,000 \times 3\,000$ pixels. Les meilleures segmentations trouvées par les différentes

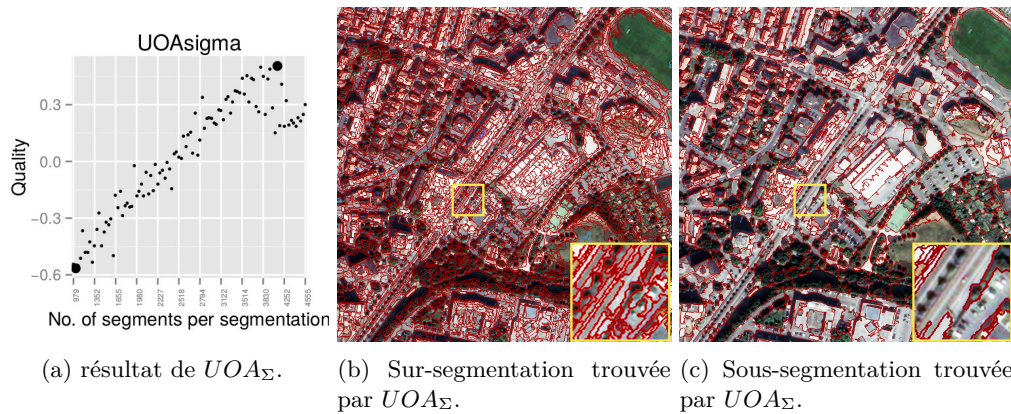


FIGURE 1.18 – Segmentations trouvées par UOA_{Σ} . À gauche : Les scores de UOA_{Σ} en fonction du nombre de segments. Un score négatif signifie une forte présence de sous-segmentation, et un score positif signifie une forte présence de sur-segmentation (voir Eq. 1.23). Au milieu : le résultat le plus sur-segmenté trouvé par UOA_{Σ} . À droite le résultat le plus sous-segmenté trouvé par UOA_{Σ} .

métriques sont montrées dans les figures 1.20, 1.21 et 1.22. Les résultats sont cohérents avec ceux de l'expérience précédente. La meilleure segmentation trouvée par SU est globalement sous-segmentée. Néanmoins des petits objets tels que les bâtiments sont bien segmentés. Le résultat trouvé par la métrique Z est bon, mais contient quelques zones sous-segmentées. La meilleure segmentation trouvée par $UOAL_2$ est globalement de bonne qualité. Les petites structures (arbres, bâtiments, etc.) et les grandes structures (rivières, routes, champs, etc.) sont bien segmentées. Nous ne constatons pas de sous-segmentation évidente, mais on y trouve quelques zones qui restent sur-segmentées, notamment des bâtiments visibles dans les figures 1.22b et 1.22c. Néanmoins, ces erreurs représentent un moindre problème puisqu'elle sont correctement détectées par l'approche d'évaluation locale et une correction facile à mettre en œuvre est envisageable.

1.4.4 Intérêt de l'évaluation locale

Maintenant, nous montrons comment l'information concernant la qualité locale de chaque segment peut être employée dans un processus de post-traitement pour corriger davantage la segmentation. L'image montrée dans la figure 1.19 correspond à un extrait de 256×256 pixels contenant principalement des bâtiments résidentiels. La figure 1.19b montre une segmentation sélectionnée de manière optimale grâce à la métrique $UOAL_2$. Si nous nous focalisons sur les maisons, nous observons que plusieurs d'entre elles sont sur-segmentées et que ceci est correctement indiqué par la fonction d'évaluation locale ϕ_{δ} (segments en jaune). Ces segments pourraient tout à fait être fusionnés entre eux afin d'obtenir une meilleure segmentation des maisons par exemple. Similairement, quelques segments sont correctement identifiés comme sous-segmentés (segments en rouge) et pourraient être corrigés en les modifiant pour faire en sorte que leur taille soit réduite.

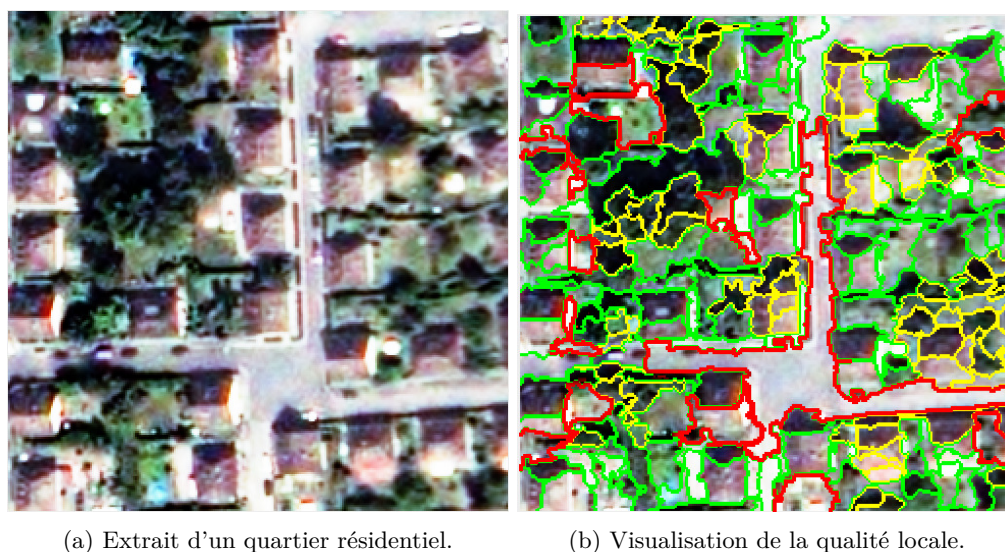


FIGURE 1.19 – Illustration des scores de qualité locale. Les segments en jaune correspondent à la sur-segmentation, les segments rouges correspondent à la sous-segmentation et les segments verts sont bien isolés par rapport à leur voisinage.

1.5 Bilan scientifique

Nous avons étudié différentes approches permettant d'évaluer la qualité en segmentation d'images. Les approches subjectives sont difficiles et coûteuses à mettre en place. Les approches supervisées constituent un bon moyen de quantifier la similarité entre deux segmentations. Lorsque l'on dispose d'une vérité-terrain, ces approches fournissent un indice de qualité. Néanmoins, une telle vérité-terrain n'est pas toujours disponible et cela demande beaucoup de travail de les générer manuellement. Finalement, les approches non-supervisées permettent d'avoir un indice de qualité directement à partir de l'image segmentée, mais il est très difficile de modéliser correctement le jugement humain. De plus, les images de télédétection possèdent des propriétés particulières dont il faut tenir compte afin de caractériser correctement la qualité des segmentations de ce type d'images.

Pour répondre à cette problématique nous avons proposé une méthode innovante pour évaluer la segmentation d'images de télédétection de manière non-supervisée. Contrairement aux approches de l'état de l'art, notre méthodologie se base sur une évaluation locale de chaque segment ; cela nous permet de rester dans le paradigme orienté objet pour effectuer des évaluations plus appropriées. Les deux métriques proposées sont définies dans un intervalle borné, sont faciles à interpréter et permettent de trouver des segmentations de bonne qualité parmi un grand ensemble de segmentations différentes.

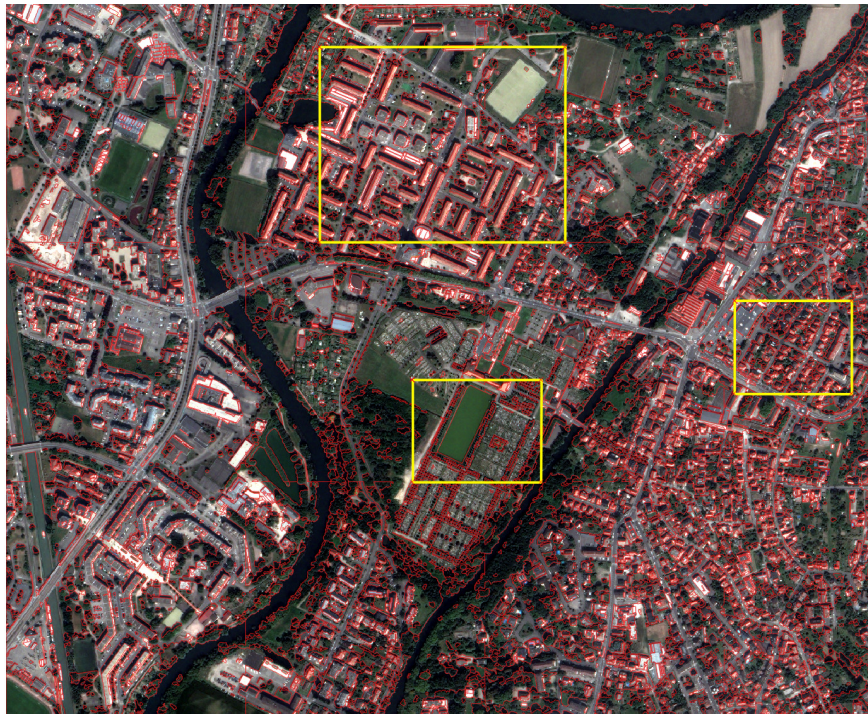
Par ailleurs nous avons vu que l'estimation locale de la qualité de chaque segment peut s'avérer une information supplémentaire et utile, permettant de mettre en place un processus de remise en cause de la segmentation.

Les travaux présentés dans ce chapitre ont donné lieu à une publication dans une conférence francophone et à une publication dans une revue internationale :

- [Troya Galvis, 2015] Atelier Fouille de Données Complexes (FDC 2015), rencontres franco-phones sur l'Extraction et la Gestion des Connaissances (EGC 2015) ;

-
- [Troya-Galvis et al., 2015] Revue internationale *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Un module pour l'évaluation de la segmentation de manière non-supervisée a été développé et intégré dans la librairie de segmentation Java Segmentation Library (JSL) développée et maintenue par l'équipe Sciences des Données et Connaissances (SDC). Un outil indépendant pour l'évaluation et la visualisation de la qualité locale de la segmentation a été également développé (voir annexe D).



(a) Zone complexe.



(b) Zone urbaine.

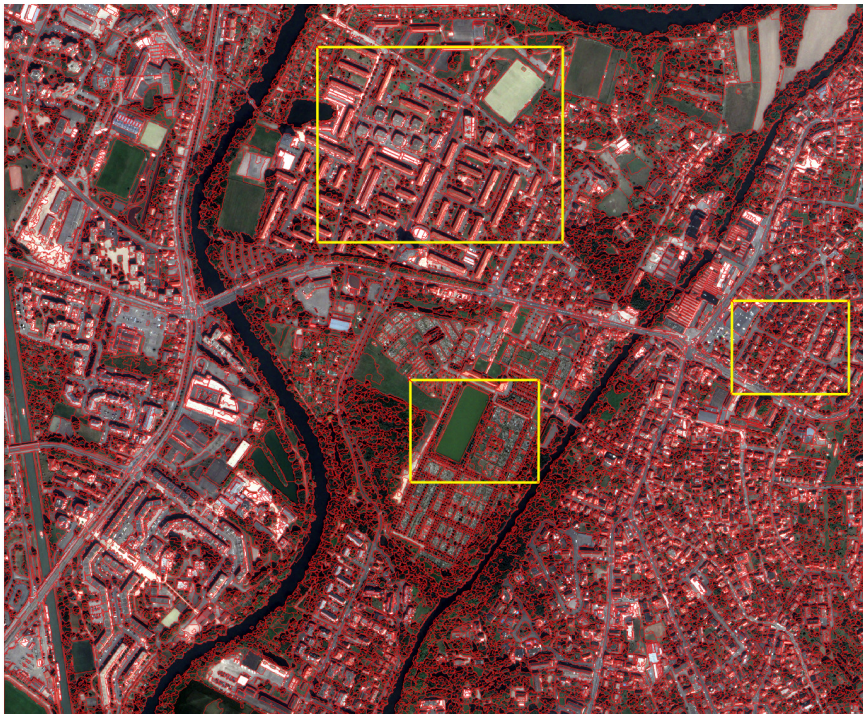


(c) Zone industrielle.

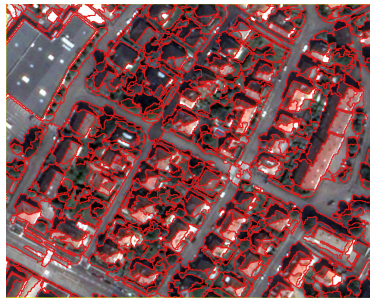


(d) Zone de végétation.

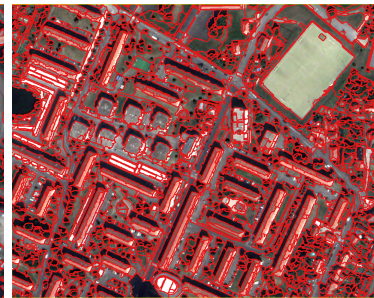
FIGURE 1.20 – Meilleur résultat trouvé par la métrique SU sur un extrait de 3000×4000 pixels contenant une variété d'objets complexes.



(a) Zone complexe.



(b) Zone urbaine.

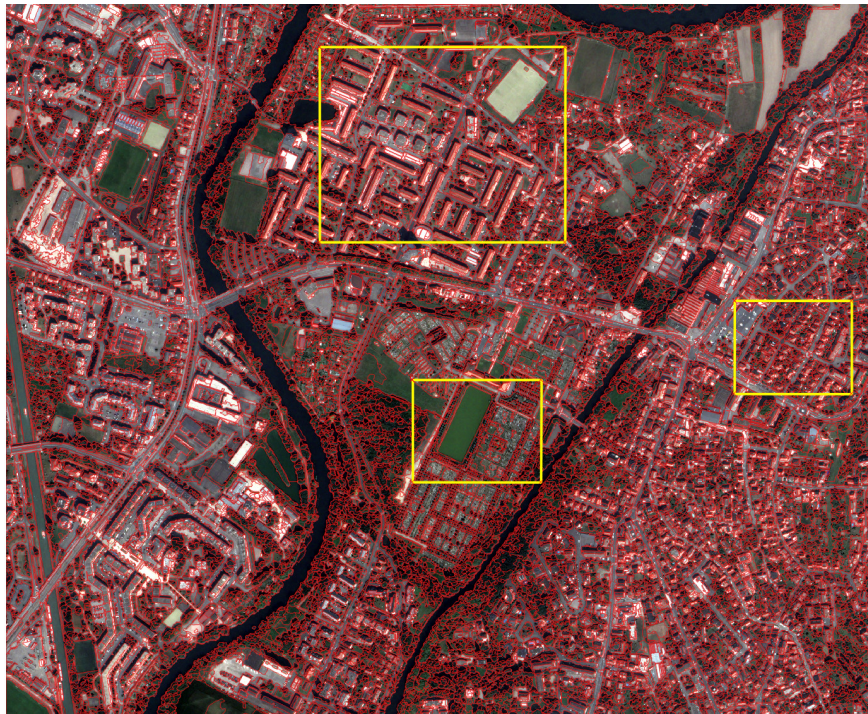


(c) Zone industrielle.

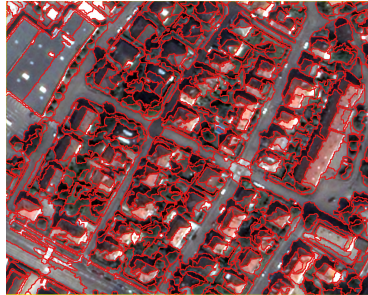


(d) Zone de végétation.

FIGURE 1.21 – Meilleur résultat trouvé par la métrique Z sur un extrait de 3000×4000 pixels contenant une variété d'objets complexes.



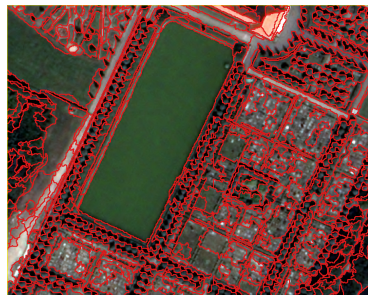
(a) Zone complexe.



(b) Zone urbaine.



(c) Zone industrielle.



(d) Zone de végétation.

FIGURE 1.22 – Meilleur résultat trouvé par la métrique UOA_{L_2} sur un extrait de 3000×4000 pixels contenant une variété d'objets complexes.

Création d'un jeu de données de référence de qualité

Les données de référence jouent un rôle important dans le processus d'analyse d'images. En effet, elles sont souvent employées comme jeu de données d'entraînement pour l'apprentissage de modèles de classification supervisée, mais également pour l'évaluation objective des résultats. De fait, il est nécessaire de disposer de données de référence les plus précises possible, d'une part, pour assurer la bonne qualité des modèles de classification appris et, d'autre part, pour permettre une évaluation pertinente des résultats. Un jeu de données de référence de bonne qualité se doit d'avoir essentiellement deux propriétés :

1. il doit fournir une labellisation la plus complète et précise possible des objets d'intérêt dans l'image ;
2. les contours des segments de référence doivent s'aligner correctement avec les objets.

Il est assez courant d'employer des données issues de bases de données géographiques comme données de référence. Néanmoins, celles-ci peuvent être obsolètes ou ne pas correspondre à la réalité au moment de l'acquisition de l'image. Un exemple est donné dans la figure 2.1. Il s'agit d'un extrait des données de référence fournies par l'expert géographe¹ dans le cadre du projet COCLICO. Ces données respectent la première propriété, c'est-à-dire qu'elles fournissent une labellisation *a priori* cohérente avec l'image. Néanmoins, la deuxième propriété n'est pas satisfaite. En effet, lorsque l'on superpose les contours des segments de référence avec l'image, on réalise que ceux-ci ne correspondent que partiellement aux contours des objets. Ils présentent un décalage non régulier dû à l'angle d'acquisition des images ayant servi à la construction de ces données et aux corrections géométriques effectuées sur ces images. Malheureusement, il ne semble pas avoir de transformation évidente permettant de recalculer correctement les données.

En vue d'améliorer les données de référence à notre disposition, nous proposons de faire appel au mécanisme de *crowd-sourcing* pour les évaluer, quantifier leur niveau d'incertitude, et surtout, créer un jeu de données fiable pouvant être employé pour l'apprentissage de modèles ou pour la validation des résultats. Dans la suite de ce chapitre, nous présentons en détail les données de référence initiales ainsi que la procédure d'amélioration mise en place. Ensuite, nous présentons le concept de *crowd-sourcing* et nous en motivons l'intérêt pour les applications en télédétection. Finalement, nous présentons l'expérience menée et nous analysons les résultats.

1. Anne Puissant, LIVE.

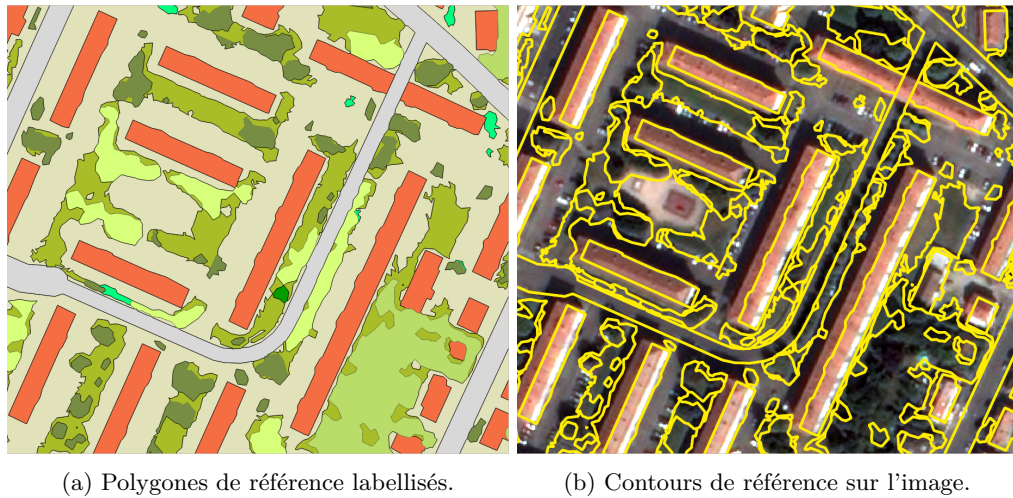


FIGURE 2.1 – Exemple de données de référence issues de bases de données géographiques.

2.1 Présentation des données

La figure 2.2 illustre nos données de référence sur la zone de la communauté urbaine de Strasbourg. Elles ont été préparées par le SERTIT² suivant une méthodologie faisant intervenir plusieurs sources d'information. Il s'agit d'une classification et sa cartographie faites à partir de plusieurs images satellites des mois d'avril et août 2012, complétées par des informations externes telles que les bases de données Topo (contours des bâtiments), Parcellaire (limites communales) et Carto (îlots géographiques bâtis, homogènes ou végétalisés) de l'IGN³. L'ensemble de ces données a été corrigé et affiné par photo-interprétation assistée par ordinateur. Le produit final de ces données consiste en 79 796 polygones regroupés en 16 classes thématiques (Tableau 2.1). Il est à noter que ces classes sont fortement déséquilibrées, comme le montre la figure 2.11.

À première vue, ces données semblent de très bonne qualité et tout-à-fait adaptées et utilisables comme données de référence en télédétection. En effet, sur la figure 2.1a, on voit que celles-ci sont complètes et que les polygones correspondant aux bâtiments et aux routes sont très réguliers et bien délimités. Néanmoins, une inspection plus détaillée (voir figure 2.1b) révèle que les polygones ne s'alignent pas correctement avec les objets dans l'image, et ce notamment, au niveau des bâtiments; les polygones ne couvrent qu'une portion de ceux-ci. Une autre imperfection est observable sur le bâtiment en haut à droite de l'image qui est découpé en deux par un segment de route d'après les données de référence. Par ailleurs, ce désalignement n'est pas régulier sur toute l'image. Par conséquent, ces données ne paraissent pas directement utilisables en l'état, que ce soit pour apprendre des modèles de classification ou pour l'évaluation de ceux-ci; cela entraînerait des résultats biaisés et dont l'erreur et l'incertitude sont difficiles à estimer.

En vue d'améliorer ces données de référence et ainsi obtenir des données plus adaptées à notre problématique, nous avons défini une procédure de raffinement d'une sur-segmentation qui se sert de labels proposés pour corriger les erreurs de la sur-segmentation et pour (ré-)attribuer un label aux segments. Cette procédure est décrite formellement par l'algorithme 1. Elle consiste à superposer les segments de la sur-segmentation avec la carte de labels et à attribuer à chaque segment de la sur-segmentation le label majoritairement présent. Finalement, tous les segments

2. <http://sertit.u-strasbg.fr/>

3. <http://www.ign.fr/>



FIGURE 2.2 – Données de référence sur l'image de Strasbourg fournies par l'expert.

adjacents ayant le même label sont fusionnés entre eux. Cette procédure permet effectivement d'obtenir des données de référence améliorées, comme le montre la figure 2.3. En effet, les frontières des segments de la sur-segmentation ont tendance à s'aligner avec les objets géographiques. De fait, l'attribution du label du segment de référence majoritaire permet d'obtenir des segments correctement labellisés. Enfin, la procédure de fusion en post-traitement permet de réduire les erreurs de sur-segmentation. La sur-segmentation employée pour effectuer ce traitement est composée de 187 057 segments et les données obtenues après traitement sont composées de 26 872 segments. Le nombre de segments a donc été réduit de façon très importante. Une inspection visuelle suggère que la segmentation a été corrigée de manière considérable et que les labels attribués sont cohérents. Néanmoins, il est nécessaire de trouver un moyen d'évaluer la qualité de ces données afin de pouvoir les exploiter efficacement.

Algorithme 1 : Affiner segmentation

Data : Sur-segmentation S_o
Data : Segmentation de référence labellisée S_r
begin
 foreach $s_i \in S_o$ **do**
 trouver l'ensemble $S_I = \{s_j \mid s_i \cap s_j \neq \emptyset\}$
 foreach $s_j \in S_I$ **do** calculer $AI_j = \frac{Area(s_i \cap s_j)}{Area(s_i)}$;
 $label(s_i) \leftarrow label(\arg \max_{s_j} AI_j)$
 foreach $s_i \in S_o$ **do** $s_i \leftarrow s_i \cup s_j \mid label(s_i) = label(s_j)$;

2.2 Crowd-sourcing pour évaluer la qualité des données de référence

Le *crowd-sourcing* consiste à diviser un problème complexe en sous-tâches facilement réalisables par une personne ayant un minimum d'entraînement ou d'expertise sur un domaine. Ces tâches sont alors soumises à un grand nombre de personnes, couramment appelées contributeurs ou encore sourceurs (la « foule »), qui vont les traiter jusqu'à ce que le problème initial soit résolu. En effet, il a été démontré qu'une évaluation faite en agrégeant les avis d'un groupe d'individus différents est souvent plus pertinente que celle d'un seul individu qu'il fasse partie du groupe ou non [Surowiecki, 2005]. Les problèmes susceptibles d'être soumis au *crowd-sourcing* sont ceux difficilement réalisables de manière automatique par une machine et/ou faisant souvent intervenir un jugement humain, généralement non remplaçable par un algorithme. Des exemples de problèmes classiquement résolus à l'aide du *crowd-sourcing* sont : la modération de contenus, la catégorisation d'images, collection de données, etc.

La tâche d'évaluer la qualité ou la pertinence de ces données de référence semble donc être adaptée au *crowd-sourcing*. En effet, la seule façon d'estimer la pertinence de ces données est d'en faire un traitement manuel. Bien que l'évaluation des données de télédétection demande habituellement une certaine expertise concernant la zone étudiée, ainsi que du contexte applicatif, en guidant la foule par le biais d'outils adaptés, et en collectant un grand nombre d'avis, nous pouvons espérer collecter assez d'avis, même non-experts, pour évaluer ces données, en déduire leur qualité et les rendre exploitables en télédétection.

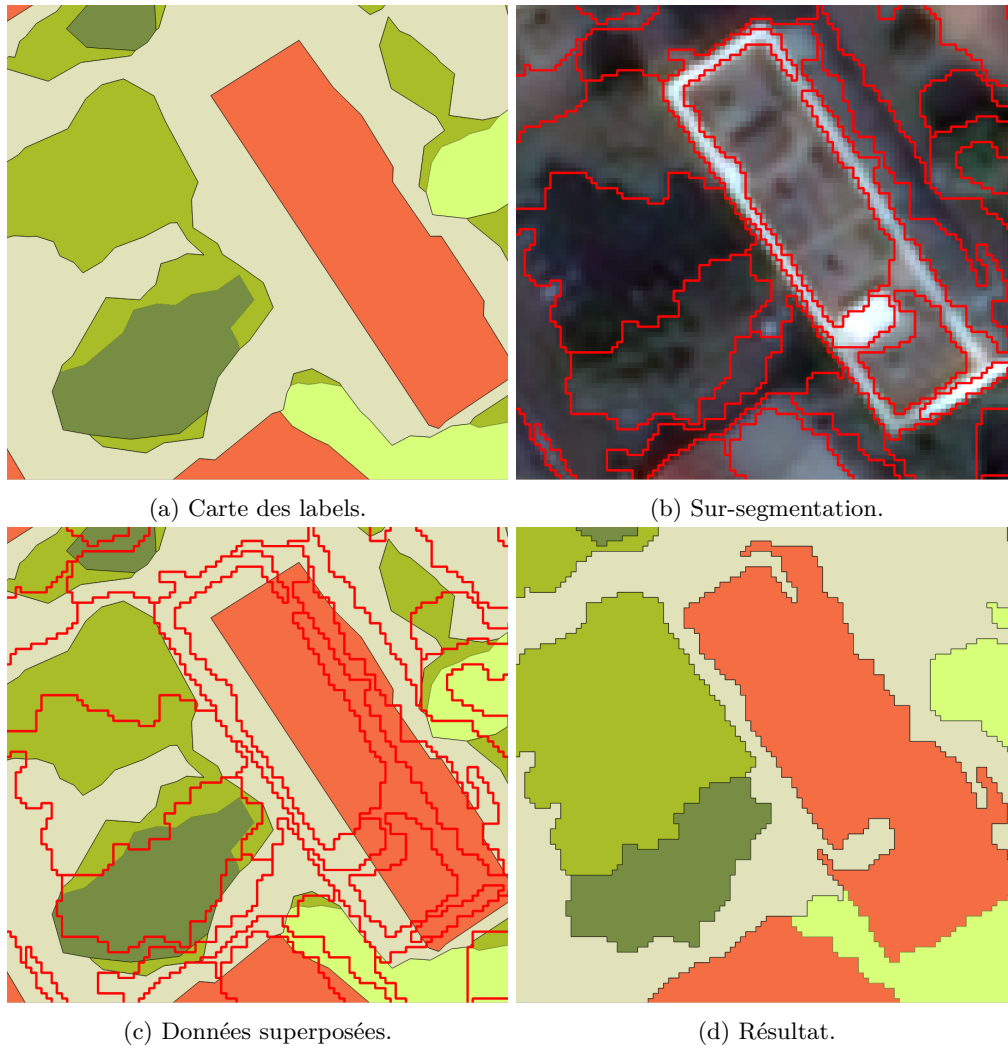


FIGURE 2.3 – Illustration de la procédure d'amélioration des données de référence.

Pour effectuer cette tâche de *crowd-sourcing* nous avons décidé d'utiliser la plateforme CrowdFlower⁴. Sur cette plateforme, les problèmes sont divisés en sous-problèmes appelés *jobs* qui sont constitués d'un ensemble de tâches unitaires (*unit task*). Chaque tâche unitaire consiste en un certain nombre de traitements à effectuer sur une donnée. Ces tâches unitaires devraient être, en principe, facilement réalisables par un individu non-expert du domaine.

Afin d'assurer la qualité des données générées, CrowdFlower met en place deux stratégies.

- Il est possible de déterminer le nombre de fois que l'on souhaite traiter une même donnée par des contributeurs différents. En effet, plus le nombre d'avis est important, plus leur agrégation sera proche de la réalité.
- Il est possible de définir des tâches unitaires de test, aussi appelées questions de contrôle. Le créateur du projet doit répondre manuellement à ces questions. Elles permettent de vérifier si les réponses des contributeurs correspondent aux critères du projet. Ainsi, il est possible de repérer si la description de la tâche n'est pas suffisamment claire, si un contributeur se trompe de manière anecdotique ou s'il est en train de répondre de manière incohérente ou aléatoire. Par ailleurs, elles permettent d'associer à chaque contributeur un score de confiance qui permet d'éviter, au moyen d'un seuil, les moins performants d'entre eux.

La définition d'un projet de *crowd-sourcing* dans CrowdFlower passe par trois étapes. La première consiste à présenter la tâche à effectuer et à détailler les instructions à suivre afin de compléter la tâche avec succès. La deuxième consiste à définir la tâche en tant que telle ; il s'agit d'une part de télécharger sur la plateforme l'ensemble des données à traiter et d'autre part, de définir le questionnaire ou l'activité à réaliser par les contributeurs sur ces données. La plateforme propose nativement différents formats possibles pour collecter les réponses aux questions :

- un champs de texte simple (possibilité de définir une expression régulière pour valider les réponses possibles) ;
- un champs de texte à plusieurs lignes ;
- une question oui/non par une case à cocher ;
- une question à choix multiples avec plusieurs réponses possibles (*checkbox*) ;
- une question à choix multiples avec une réponse unique (*exclusive checkbox*) ;
- une question à choix multiples avec une réponse unique (liste déroulante).

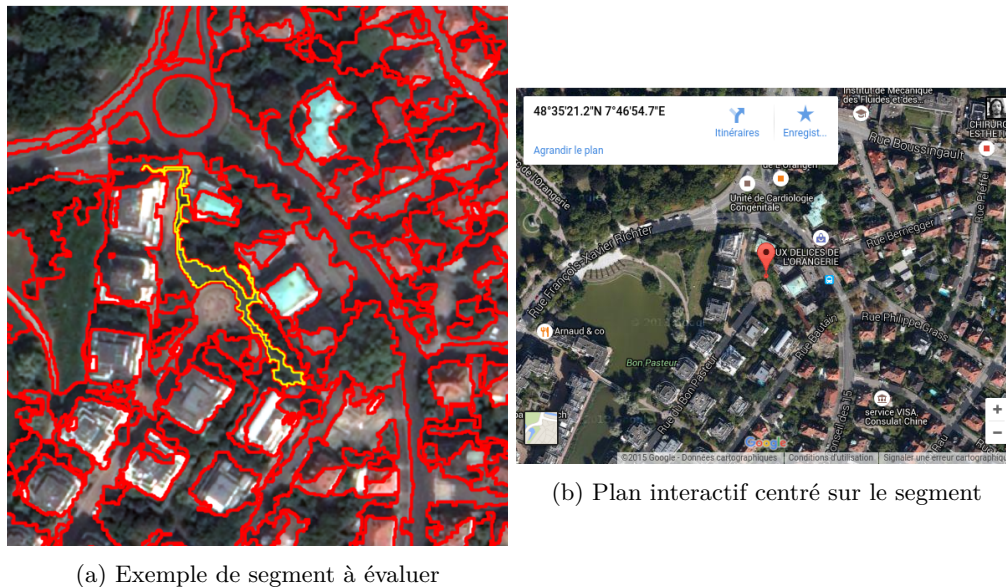
Finalement, la troisième étape consiste à répondre manuellement aux questions de contrôle qui serviront à assurer la bonne qualité des réponses collectées.

Une fois la tâche entièrement définie, elle est rendue publique et les contributeurs peuvent commencer à réaliser le travail. La plateforme met à disposition des outils pour surveiller en temps réel l'évolution du *job* et la qualité des réponses collectées. Les contributeurs ont la possibilité de contester une question de contrôle lorsqu'ils considèrent que celle-ci est fausse ou incohérente. Ainsi, il est possible de mettre en pause le processus à n'importe quel moment pour corriger les questions de contrôle et améliorer ainsi la qualité des réponses.

Nous avons défini un projet de *crowd-sourcing* pour évaluer les données de référence améliorées en termes de segmentation et de labellisation. Pour ce faire, nous avons sélectionné 10 000 segments de manière aléatoire parmi les 26 872 soit 37.21% des segments. Comme illustré par la figure 2.4, pour chaque segment, nous présentons aux participants une capture d'écran contenant les segments voisins en rouge et le segment à évaluer en jaune. Nous profitons du fait que les données sont géo-référencées pour proposer au sourceur une carte interactive Google Maps centrée sur le barycentre du segment à évaluer, pour faciliter la tâche d'interprétation au participant.

Nous avons posé les trois questions suivantes :

4. <https://crowdfower.com>



(a) Exemple de segment à évaluer

(b) Plan interactif centré sur le segment

FIGURE 2.4 – Exemple de segment soumis à évaluation lors du crowd-sourcing.

1. Le segment semble-t-il sur-segmenté (trop petit), sous-segmenté (trop large) ou de la bonne taille par rapport à l'objet géographique (choix unique parmi les 3) ;
2. Quelle note donneriez-vous à la qualité de ce segment (choix unique parmi les entiers de 1 à 10, avec 1 la moindre qualité et 10 la meilleure qualité) ;
3. Quelle est la classe de l'objet géographique sous-jacent ? (choix unique parmi : végétation haute, végétation basse, parcelle agricole, bâtiment, sol nu, eau artificielle, eau naturelle, routes, ombre, trop sur-segmenté et trop sous-segmenté).

Il est à noter que les classes proposées aux participants ne sont pas exactement celles présentes dans les données de référence d'origine. En effet, ces dernières sont certainement trop spécifiques et demanderaient une expertise trop importante aux participants. Par conséquent, nous avons préféré les simplifier en suivant les correspondances décrites par le tableau 2.1. Nous observons notamment qu'il y a des classes présentes dans les données de référence qui n'ont pas d'équivalent dans les classes proposées aux sourceurs et vice-versa.

2.2.1 Analyse statistique des résultats

Dans un premier temps, nous présentons une analyse purement statistique des données issues du *crowd-sourcing*, sans aucun traitement supplémentaire et sans y intégrer aucune connaissance thématique.

Les 3 questions décrites ont été soumises à 165 contributeurs d'origines différentes (les États-Unis et plusieurs pays européens). Pour chaque segment des 10 000 choisis, nous avons obtenu 3 avis minimum, soit 30 000 évaluations au total. Le coût de cette expérience était de 3 883€, soit 12 centimes environ par réponse. Néanmoins, après une comparaison avec d'autres projets, nous avons réalisé que la tâche a été largement surpayée et le même résultat aurait pu être obtenu facilement avec la moitié de ce budget.

La figure 2.5 montre la distribution spatiale des segments évalués par la foule lors de l'expérience de crowd-sourcing. Nous observons que les segments proviennent de divers endroits dans

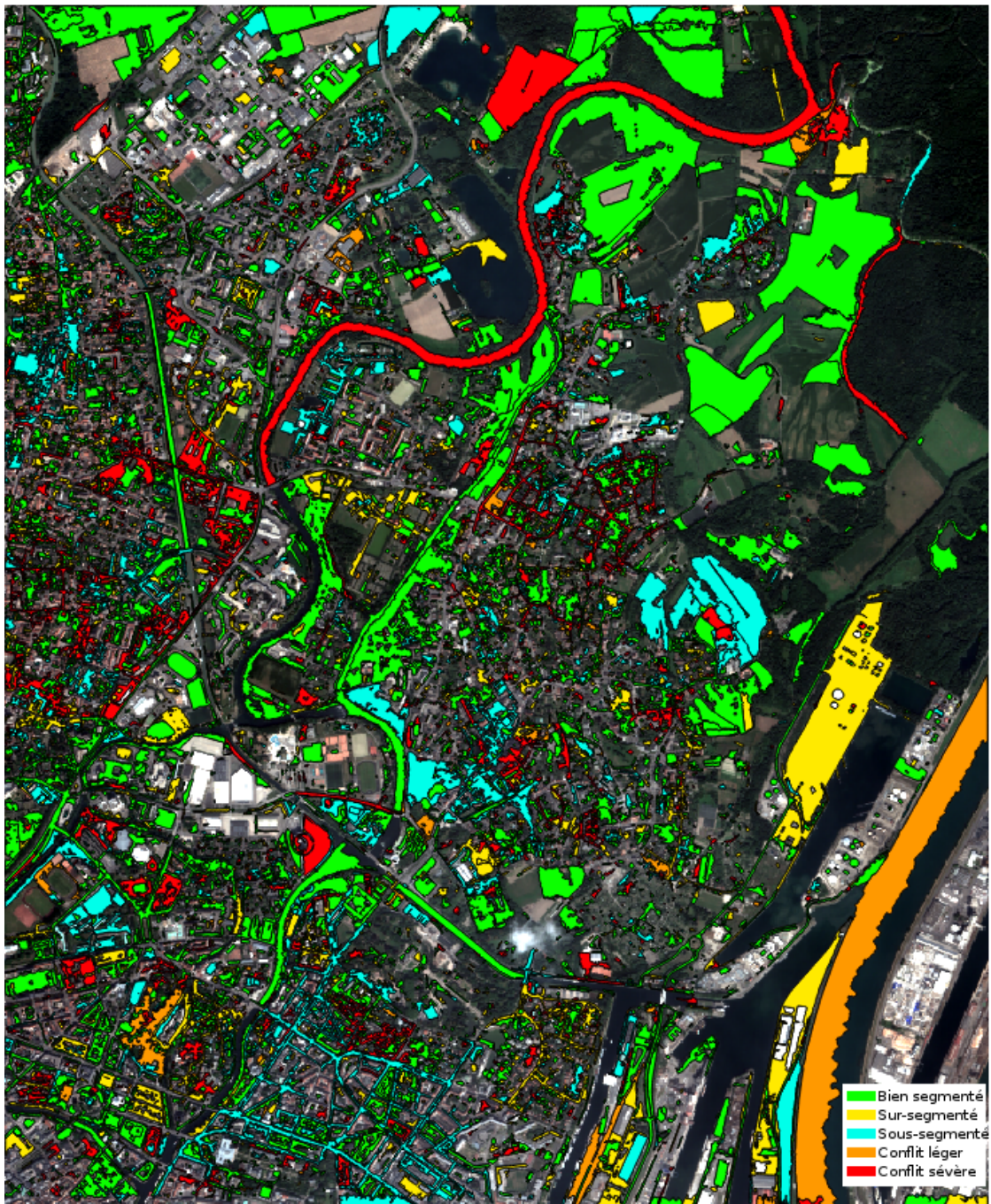


FIGURE 2.5 – Distribution spatiale des 10 000 segments évalués par la foule sur la qualité de la segmentation.

TABLEAU 2.1 – Équivalences des classes proposées aux contributeurs et dans les données de référence.

Label proposé aux sourceurs	Label présent dans les données de référence
Végétation haute	Arbre isolé, Bois, Bosquet, Forêt, Haie
Végétation basse	Bande enherbée, Végétation herbacée
Parcelle agricole	Culture supposée de printemps
Bâtiments	Bâtiments
Sol nu	Culture supposée d'hiver
Eau artificielle	Eau artificielle
Eau naturelle	Eau naturelle
Routes	Voirie
Ombre	Pas de classe
Trop sur-segmenté	Pas de classe
Trop sous-segmenté	Pas de classe
Pas de classe	Surface non-végétalisée
Pas de classe	Voie ferrée
Pas de classe	Toiture végétalisée

l'image et qu'ils couvrent une partie représentative de celle-ci. Les couleurs correspondent au jugement attribué par la foule : les segments jaunes ont été jugés comme sur-segmentés par la majorité des contributeurs ; les segments turquoises ont été jugés comme sous-segmentés par la majorité des contributeurs ; les segments verts ont été jugés comme bien segmentés par la majorité des contributeurs ; les segments rouges ont le même nombre de votes pour sur-, sous- et bonne segmentation ; et les segments oranges ont le même nombre de votes majoritaires sur deux aspects.

Les images dans les figures 2.6a–2.6e donnent un aperçu détaillé des résultats. La figure 2.6a illustre un exemple de segments jugés majoritairement comme correctement segmentés ; ce sont des segments de petite taille mais qui délimitent bien les objets correspondants. La figure 2.6b montre un exemple de segments jugés comme sur-segmentés ; ce sont des segments de très petite taille correspondant seulement à une partie des objets sous-jacents ; on observe notamment plusieurs exemples de demi-toits. La figure 2.6c montre un exemple de segment jugé comme sous-segmenté. Il s'agit en fait d'un unique segment de très grande taille (le segment ne tient pas dans sa totalité sur l'extrait d'image) qui regroupe plusieurs types d'objets tels que des routes, du sol nu et des bâtiments. La figure 2.6d montre un exemple de segment ayant un conflit relatif (avis partagés sur deux réponses). On constate qu'il s'agit d'un segment difficile à évaluer car il regroupe majoritairement de la végétation herbacée mais aussi une portion de route. Finalement, la figure 2.6e illustre un exemple de segments en conflit total (tous les avis différents). Il s'agit de segments assez larges qui correspondent bien aux contours des objets mais qui couvrent souvent plusieurs objets de même type.

Le tableau 2.2 synthétise les résultats des données agrégées sur l'évaluation de la segmentation. Il y a un total de 5 329 segments jugés comme bien segmentés, ce qui représente une surface de 4 130 388 m^2 , 2 719 segments jugés comme sur-segmentés, représentant une surface de 1 183 049 m^2 et 681 segments jugés comme sous-segmentés, correspondant à une surface de 1 297 565 m^2 . Bien que le nombre de segments sous-segmentés soit moindre, nous constatons que la surface qu'ils représentent est plus importante que celle des segments sur-segmentés puisqu'il s'agit la plupart du temps de gros segments. Il y a également 1 046 segments en conflit total et 225 segments avec un conflit léger.

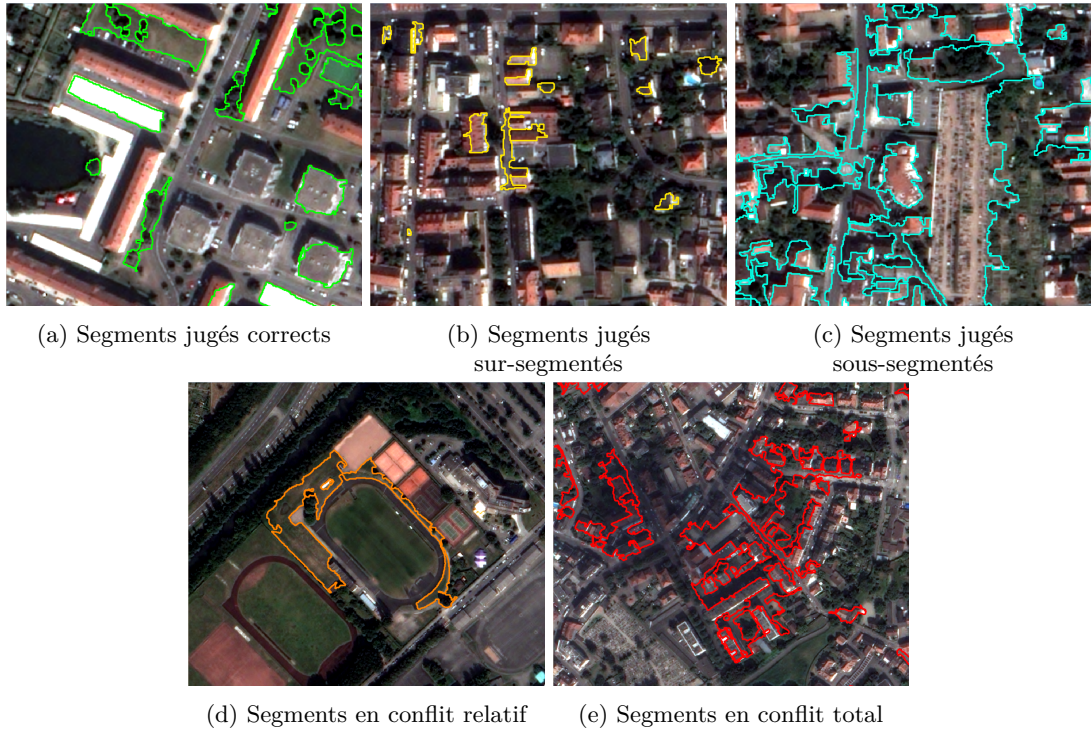


FIGURE 2.6 – Exemples de segments évalués par les sourceurs.

TABLEAU 2.2 – Nombre, aire et périmètre des segments évalués selon leur qualité de segmentation.

Évaluation	Nombre	Aire (m^2)	Périmètre (m)
correct	5 329	4 130 388	1 125 707
sur	2 719	1 183 049	411 230
sous	681	1 297 565	382 479
conflit sévère	1 046	1 153 447	350 943
conflits léger	225	539 738	66 097
total	10 000	8 304 188	2 336 456

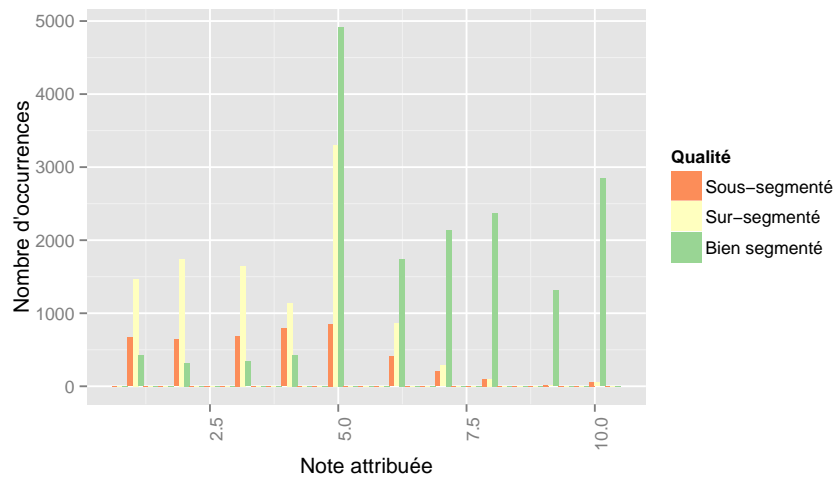


FIGURE 2.7 – Distribution de la moyenne des notes par segment par rapport à leur qualité.

La figure 2.7 montre la distribution des notes attribuées aux segments par rapport à l'évaluation qualitative de ceux-ci. Nous observons une corrélation évidente des notes avec le jugement qualitatif. En effet, les notes au-dessus de 5 correspondent majoritairement à des segments jugés comme bien segmentés, et les notes en dessous correspondent à des segments jugés comme sur- ou sous-segmentés. Il y a tout de même un certain nombre d'évaluations aberrantes où des segments jugés comme bien segmentés se sont vu attribuer une note en dessous de 5 et vice-versa. Un autre point remarquable est le fait que la valeur dominante des notes est de 5. Ceci révèle un manque de confiance de la part de la foule au moment d'évaluer les segments dont il faudra tenir compte lors de l'exploitation de ces données dans une application quelconque.

La figure 2.8 montre la distribution des labels présents dans les données de référence par rapport à la qualité des segments jugée par la foule. Nous observons que les jugements sont distribués similairement sur toutes les classes : la majorité des segments sont jugés bien segmentés. Néanmoins, un grand nombre de segments sont jugés sur-segmentés. Finalement, une minorité de segments sont jugés sous-segmentés. Il n'y a pas de classe qui semble mieux ou moins bien segmentée que les autres. Nous voyons que les classes correspondant aux parcelles agricoles, aux routes, à l'eau et à la classe inconnue sont moins présentes, mais ce déséquilibre est déjà présent dans les données de référence d'origine.

La figure 2.9 montre la distribution des labels donnés par la foule par rapport à la qualité des segments jugée par la foule. Cette fois-ci, nous observons une distribution plus irrégulière. Ainsi, les classes végétation haute, végétation basse et ombre sont jugées majoritairement comme bien segmentées. En revanche, les classes bâtiments et routes sont jugées majoritairement comme sur-segmentées. Nous observons également la présence des segments pour lesquels la foule n'a pas donné de label, car le segment semblait être trop sur-segmenté ou trop sous-segmenté. Néanmoins, les évaluations correspondantes dans ces deux cas semblent être contradictoires, puisqu'elles comportent des jugements des trois types, alors que l'on aurait attendu ne trouver que le jugement correspondant au label en question, c'est-à-dire sur-segmenté ou sous-segmenté uniquement.

La figure 2.10 montre la distribution des labels attribués par la foule par rapport aux labels d'origine dans les données de référence. Nous observons ici un fait surprenant : il existe une grande confusion entre les labels attribués par la foule et ceux d'origine, il n'y a pas de tendance

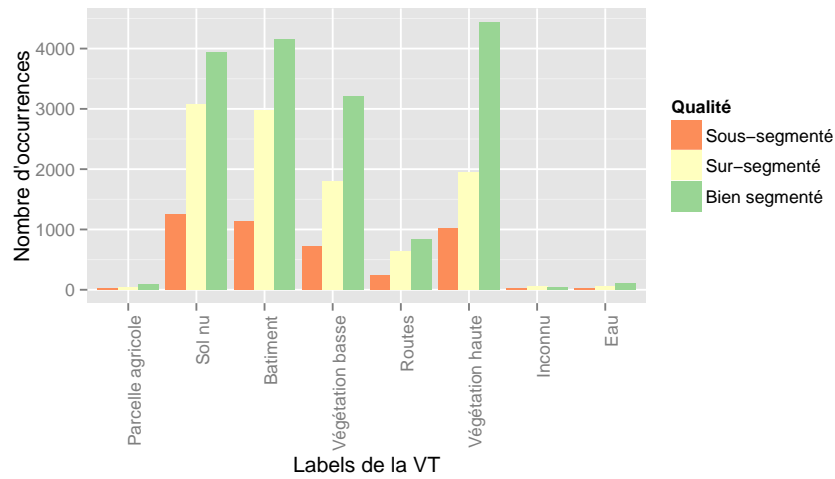


FIGURE 2.8 – Distribution des labels des segments dans les données de référence par rapport à la qualité des segments.

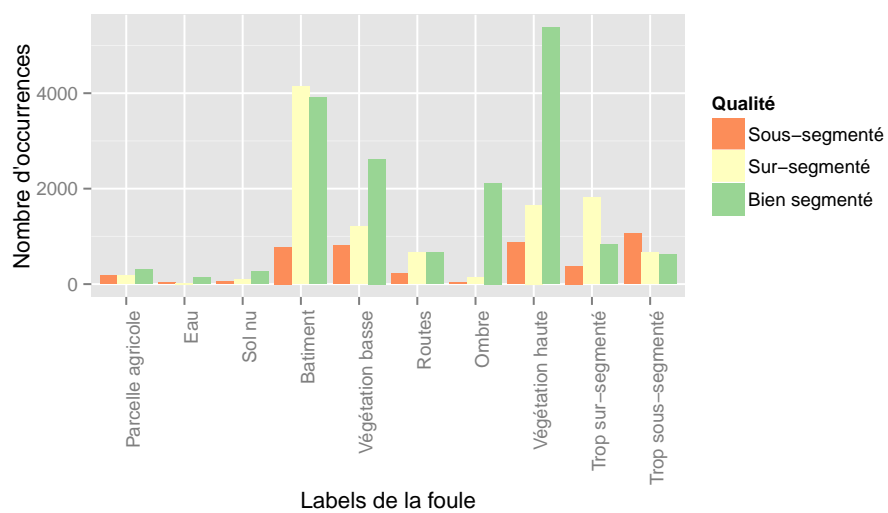


FIGURE 2.9 – Distribution des labels de la foule par rapport à la qualité des segments.

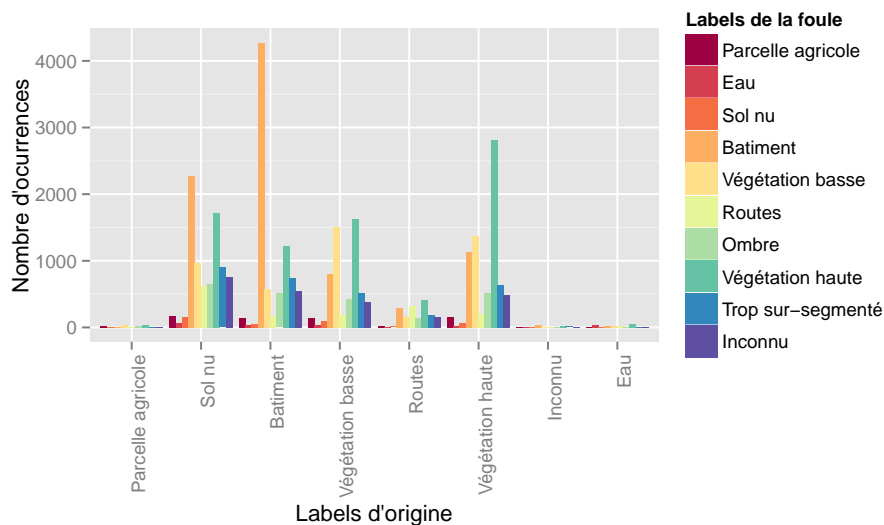


FIGURE 2.10 – Distribution des labels attribués par la foule par rapport aux labels des données de référence.

évidente qui puisse expliquer ce fait, et une étude thématique plus poussée s'avère nécessaire pour expliquer ce phénomène et pour déterminer la pertinence de ces données labellisées.

2.2.2 Analyse thématique des données

La première analyse des résultats du *crowd-sourcing* a révélé que les données de référence contenaient encore une quantité non négligeable d'imprécisions et d'incertitudes associées. Afin de rendre ces données exploitables, il est nécessaire de réaliser une analyse d'un point de vue thématique, permettant ainsi de créer une base de données de référence fiable, avec une estimation objective de la qualité de chaque exemple présent dans la base. Dans cette section, nous présentons la méthodologie mise en place pour la création d'une telle base de données.

En vue de créer la base de données d'exemples, nous avons eu recours à un expert géographe⁵ pour analyser les données du *crowd-sourcing* d'un point de vue plus thématique. Cette nouvelle analyse a permis de mettre en évidence plusieurs observations :

Le déséquilibre des classes. Les classes présentes dans les données de référence ne sont pas distribuées uniformément, comme le montre la figure 2.11. En effet, on observe une prédominance des classes de végétation et de bâtis. Rappelons que toutes les classes de végétation sont regroupées uniquement en deux classes lors du *crowd-sourcing*. De plus, des classes comme l'eau ou la voirie sont beaucoup moins représentées. Cela pose problème pour obtenir un jeu de données conséquent, comme nous le verrons dans la suite de cette section.

Les incohérences d'évaluation au niveau de la segmentation. Plusieurs incohérences par rapport à la segmentation de plusieurs segments ont été détectées. Notamment, 5 072 contributeurs ont qualifié des segments comme étant sur- ou sous-segmentés, mais 1 434 d'entre eux ont attribué une note de segmentation élevée à ces mêmes segments. De plus, 353 segments ont été jugés comme sur-segmentés mais le label qui leur a été attribué est « trop grand pour décider ». Enfin, 567 segments ont été jugés comme sous-segmentés mais le label qui leur a été attribué

5. Arthur Bourachot, étudiant en Master Observation de la Terre et Géomatique.

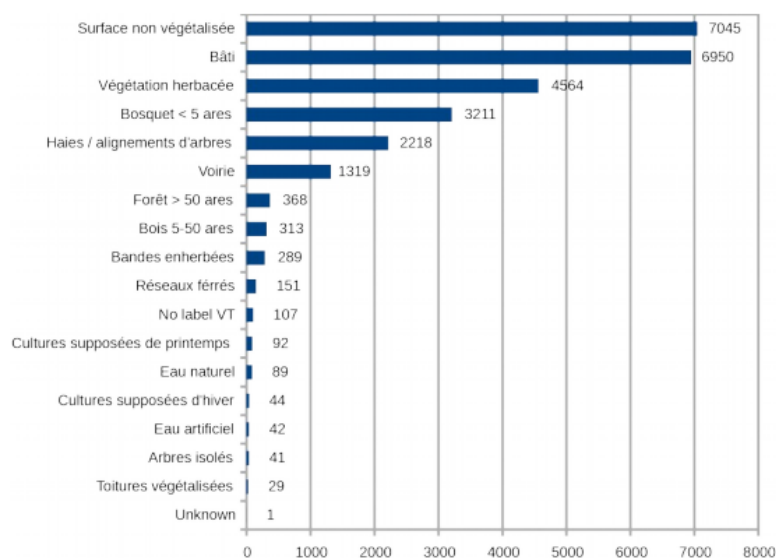


FIGURE 2.11 – Distribution des classes dans les données de référence d'origine.

est « trop petit pour décider ». Cela laisse penser que les instructions du *crowd-sourcing* étaient peut-être insuffisamment claires.

Les incohérences au niveau des labels attribués. Une bonne partie des segments évalués n'ont pas été labellisés de manière unanime par les contributeurs, et dans certains cas, les labels attribués étaient contradictoires. Cela suggère que le choix de centrer la carte Google map sur le barycentre des segments n'était probablement pas très pertinent. En effet, un nombre élevé de segments ont une forme irrégulière et, de ce fait, le barycentre de ceux-ci ne se trouve pas nécessairement à l'intérieur du segment. Cela a probablement introduit un biais sur les réponses des contributeurs.

2.2.3 Création d'une base de données fiable

Afin de rendre les données exploitables dans un cadre d'analyse d'images, il est nécessaire de les agréger et de les nettoyer pour faire en sorte de ne conserver que les données de qualité. Pour cela, nous avons mis en place la méthodologie présentée dans la figure 2.12. Le traitement commence avec la base de données annotée par des contributeurs ; il se compose ensuite par la succession des étapes suivantes :

1. **Analyse statistique et nettoyage de données incohérentes.** Cette étape d'analyse statistique et empirique des données permet d'éliminer les éléments impossibles à traiter car ils possèdent des évaluations trop contradictoires.
2. **Agrégation des données.** Il est impératif de combiner les avis des différents contributeurs sur un même segment, afin de lui attribuer un label définitif. Le label définitif est celui qui a été suggéré le plus fréquemment par les contributeurs. Néanmoins, il est nécessaire de pouvoir quantifier le degré de confiance que l'on peut attribuer à ce label. Pour ce faire, nous avons défini un indicateur de consensus comme suit :

$$\text{consensus} = \frac{O_f}{T} \quad (2.1)$$

TABLEAU 2.3 – Correspondance des labels proposés lors de la deuxième phase de *crowd-sourcing*.

Label d'origine	Label 1ere phase	Label 2eme phase
Arbres isolés	Végétation haute	Zone forestière
Bois		
Bosquet		
Forêt		
Haie/alignements d'arbres		
Bandes enherbées	Végétation basse	Zone herbacée
Végétation herbacée		
Toitures végétalisées		
Culture supposée de printemps	Parcelle agricole	Zone cultivée
Culture supposée d'hiver		
Bâtiments	Bâtiments	Zone bâtie
Surface non végétalisée	Sol nu	Zone artificielle ou minéralisée
Eau naturelle	Eau naturelle	Zone d'eau
Eau artificielle	Eau artificielle	
Voirie	Routes	Réseaux
Réseaux ferrés		
Pas présente	Inconnu	Inconnu
Pas présente	Ombre	Ombre

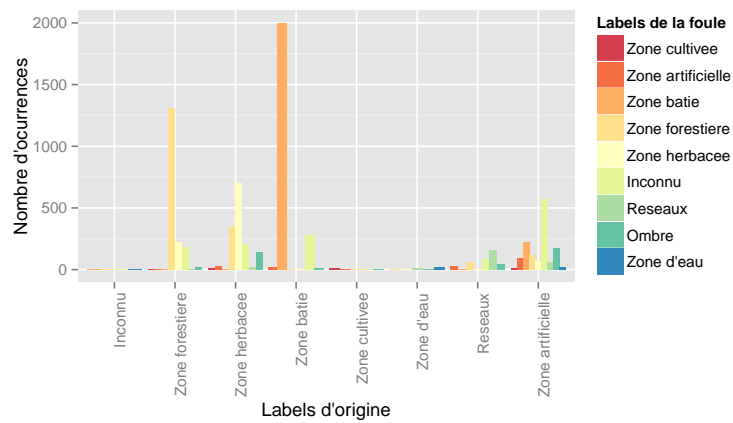
1. la description de la tâche a été simplifiée et améliorée grâce à l'ajout de plusieurs exemples explicites sur la qualité de la segmentation ;
2. les labels proposés aux contributeurs ont été simplifiés et regroupés dans des classes plus génériques et plus couramment utilisées en télédétection. Les nouveaux labels sont présentés dans le tableau 2.3 ;
3. la carte Google map est toujours centrée sur le barycentre des objets, mais celui-ci n'est pas indiqué explicitement sur la carte.

Plusieurs questions du questionnaire ont été retravaillées pour les rendre plus faciles à comprendre et plus pertinentes. Le nouveau questionnaire se présente comme suit :

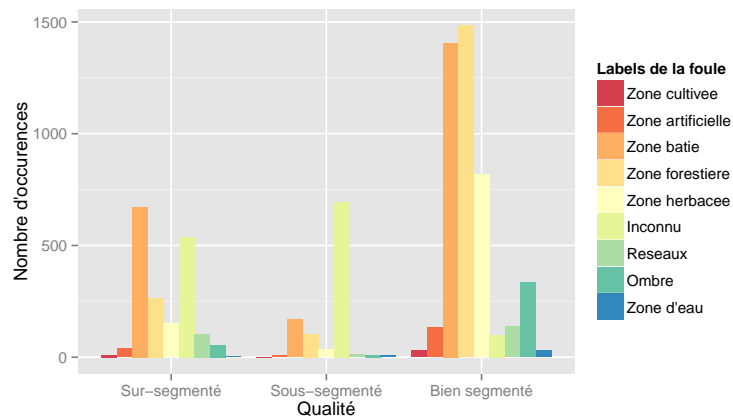
- Choisissez le label qui décrit le mieux la zone comprise dans le segment jaune. Sélectionnez « Je ne sais pas » si aucun label ne semble correspondre au segment.
- Le segment a été trop difficile à labelliser, pourquoi ? (cette question n'est posée que si le contributeur a répondu « Je ne sais pas » à la question précédente.
- Le segment jaune semble être : trop petit par rapport à l'objet sous-jacent, trop large par rapport à l'objet sous-jacent, ou de la bonne taille par rapport à l'objet sous-jacent ?
- Le segment était-il difficile à labelliser ?
- Êtes-vous habitué aux concepts de la segmentation en analyse d'images ?

Pour cette deuxième phase de *crowd-sourcing*, nous avons évalué deux sous-ensembles de segments différents pour étudier plus précisément l'influence de la présentation des données sur les résultats.

Le premier sous-ensemble consiste en 2 858 segments qui correspondent aux segments ayant un indice de consensus faible et relatif (cf. Eq. 2.1) après avoir filtré les segments visiblement trop larges et trop petits pour correspondre à un objet géographique intéressant. La figure 2.13 montre les résultats pour l'ensemble de segments presque consensuels. On observe que la confusion entre

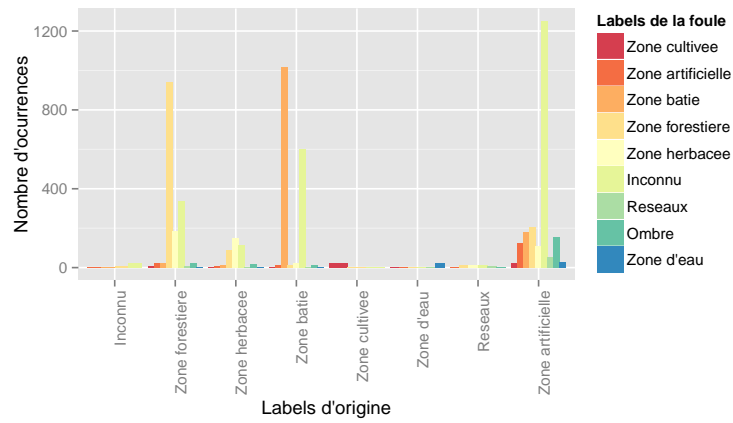


(a)

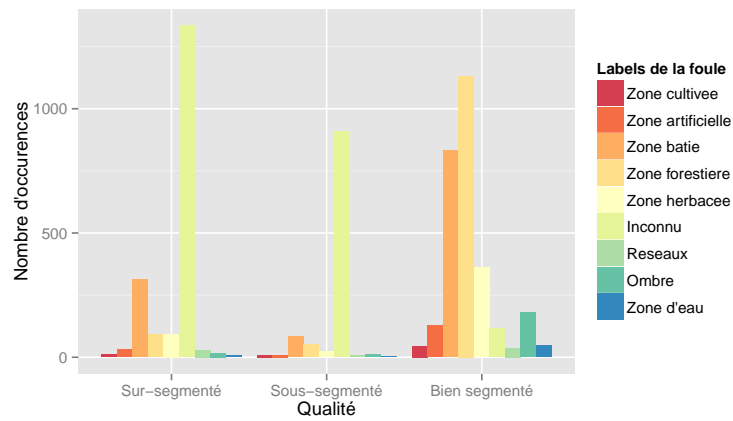


(b)

FIGURE 2.13 – Résultats de la deuxième phase avec les segments à consensus faible et relatif.



(a)



(b)

FIGURE 2.14 – Résultats de la deuxième phase avec les segments en désaccord et désaccord total.

ces deux labels est nettement moins importante (Figure 2.13a) que lors de la première phase de *crowd-sourcing*, où il y avait beaucoup de bruit et de labels contradictoires. En effet, la classe « Zone bâtie » ne présente quasiment pas de confusion ; on trouve uniquement quelques segments que la foule a labellisé comme « Inconnu ». Ensuite, on voit que les labels des classes « Zone forestière » et « Zone herbacée » sont majoritairement concordants avec ceux d'origine, bien qu'il existe des confusions légères entre ces deux classes, ce qui est compréhensible vu leur nature proche et la complexité de certaines zones de l'image. Pour les classes moins représentées, cet effet n'est pas clairement observable sur le graphique mais les labels sont le plus souvent concordants avec ceux d'origine. La seule classe où il reste des confusions considérables est « Zone artificielle ». En effet, le label majoritaire attribué par la foule est « Inconnu », mais on remarque également la présence de tous les autres labels. Ceci n'est pas surprenant puisque la classe « Zone artificielle » correspond à des objets minéralisés divers qui ne sont pas des bâtiments ou de réseaux routiers. On y trouve donc des objets très différents tels qu'un cimetière ou un terrain de jeux. Par ailleurs, on observe que la majorité des segments ont été jugés comme bien segmentés (Figure 2.13b). On remarque également que parmi les segments sous-segmentés le label majoritaire est « Inconnu », ce qui est normal puisque, par définition, un segment sous-segmenté contient plusieurs objets différents et il est donc impossible de lui attribuer un label unique. Finalement, on remarque qu'une bonne partie des bâtiments sont sur-segmentés.

Le deuxième sous-ensemble consiste en 2 525 segments qui correspondent à des segments ne présentant pas de consensus (désaccord et désaccord total). Ce sont donc des segments qui ont été particulièrement difficiles à évaluer lors de la première phase de *crowd-sourcing*. La figure 2.14 montre les résultats pour l'ensemble des segments non consensuels. On constate une fois de plus que la confusion des labels est moindre que lors de la première phase (Figure 2.14a). Néanmoins, on remarque la présence d'un nombre très important de segments avec le label « Inconnu ». De plus, on constate que la plupart des segments labellisés ainsi sont soit sur-segmentés, soit sous-segmentés (Figure 2.14b), ce qui explique en partie pourquoi ils ont posé problème lors de la première phase.

Nous avons vu que les résultats de cette deuxième phase ont été beaucoup plus consensuels et cohérents que ceux de la première phase. Nous pouvons en déduire que la manière de présenter les données et la tâche à effectuer est cruciale au bon déroulement d'une expérience de *crowd-sourcing*. Il est nécessaire de choisir de bons exemples, qui soient pertinents et qui montrent aux contributeur le résultat attendu. Il est également important de faire une description claire et concise de la tâche et de bien spécifier l'objectif de celle-ci, en évitant d'utiliser un vocabulaire du domaine trop spécialisé, afin d'éviter de perdre ou confondre les contributeurs non experts. Finalement, il faut que les questions soient explicites et qu'elles ne reposent pas sur un avis subjectif du contributeur qui est impossible à vérifier et qui entraîne des résultats très bruités et peu exploitables.

Enfin, nous avons effectué les traitements décrits dans la section 2.2.3 sur les données de la deuxième phase. Au final, nous disposons d'une base d'exemples comportant 2 781 segments de bonne qualité, dont 1 130 sont des zones bâties, 974 sont des zones forestières, 421 sont des zones herbacées, 93 sont des routes, 76 correspondent à de l'ombre, 51 sont des zones artificialisées, 19 sont des zones d'eau et 17 sont des parcelles agricoles.

2.3 Bilan scientifique

Le *crowd-sourcing* s'avère une solution intéressante lorsque l'on doit traiter un grand volume de données dont le traitement nécessite l'intervention d'un être humain. Néanmoins, il est important de définir clairement l'objectif recherché. En effet, le manque d'expertise des contri-

buteurs, des contributeurs parasites, ou l'incompréhension des instructions, peut résulter en des données incohérentes et/ou peu fiables. Nous avons employé le *crowd-sourcing* pour l'évaluation de données de référence en télédétection. Nous avons procédé en deux phases : la première a servi à identifier et filtrer les segments de mauvaise qualité ainsi que les segments dont l'évaluation est particulièrement difficile ; la deuxième, s'est focalisée sur ces segments difficiles et a permis d'obtenir une évaluation plus pertinente de ceux-ci. Nous avons ainsi généré une base de données d'exemples dont la qualité est élevée et qui peut servir tant à l'apprentissage automatique de modèles de prédiction qu'à l'évaluation des résultats des méthodes d'analyse d'images en télédétection.

Ces travaux ont donné lieu à deux publications scientifiques en collaboration avec une équipe du LIPN :

- [Grozavu et al., 2015] conférence internationale Machine Vision Applications (MVA2015) ;
- [Sublime et al., 2015] conférence internationale Machine Vision Applications (MVA2015).

Une base de données d'exemples corrigée a été produite et sera rendue publique prochainement.

Deuxième partie

**Approche collaborative
multi-paradigme pour l'extraction
mono-classe**

CoSC : Collaborative Segmentation and Classification

Dans ce chapitre, nous proposons un cadre formel et générique pour l'extraction d'une classe thématique. Il se base sur une méthode collaborative de segmentation et de classification. Dans un premier temps, les différentes étapes qui constituent le processus sont présentées. Dans un second temps, plusieurs implémentations du processus sont proposées et comparées. Finalement, une étude expérimentale consistant à extraire des zones de végétation dans une image de l'agglomération strasbourgeoise est présentée.

3.1 Approches mono-classe pour l'analyse d'image de télédétection

L'extraction, dans une image, d'une seule classe thématique consiste à labelliser uniquement les segments correspondant à des objets de cette classe. Les approches mono-classes permettent de focaliser les efforts sur la conception d'algorithmes exploitant des caractéristiques particulières de la classe recherchée. [Musci et al., 2013] affirme que la meilleure stratégie pour travailler avec les images à très haute résolution spatiale consiste à employer des segmentations séparées pour chaque classe, étant donnée la grande variabilité des objets présents. Ainsi, chacune d'entre elles peut être paramétrée spécifiquement pour mieux retrouver les objets de la classe, menant ainsi à des meilleures performances. Néanmoins, les approches classiques effectuent la segmentation de l'image de manière indépendante à la classification. Or, il a été montré que la segmentation a un impact sur le résultat de la classification [Gao et al., 2011], mais elle n'est jamais remise en cause en fonction de ces résultats. Par conséquent, si la segmentation initiale n'est pas adaptée aux critères employés par la méthode de classification, les résultats de cette dernière ont une forte chance d'être sous-optimaux.

Plusieurs méthodologies existent pour extraire une classe d'objets, comme des bâtiments [Liasis and Stavrou, 2016] ou des routes [Unsalan and Sirmacek, 2012, Lacoste et al., 2010]. Cependant, ces approches sont très spécifiques et très dépendantes des données, du type d'objet et du contexte applicatif.

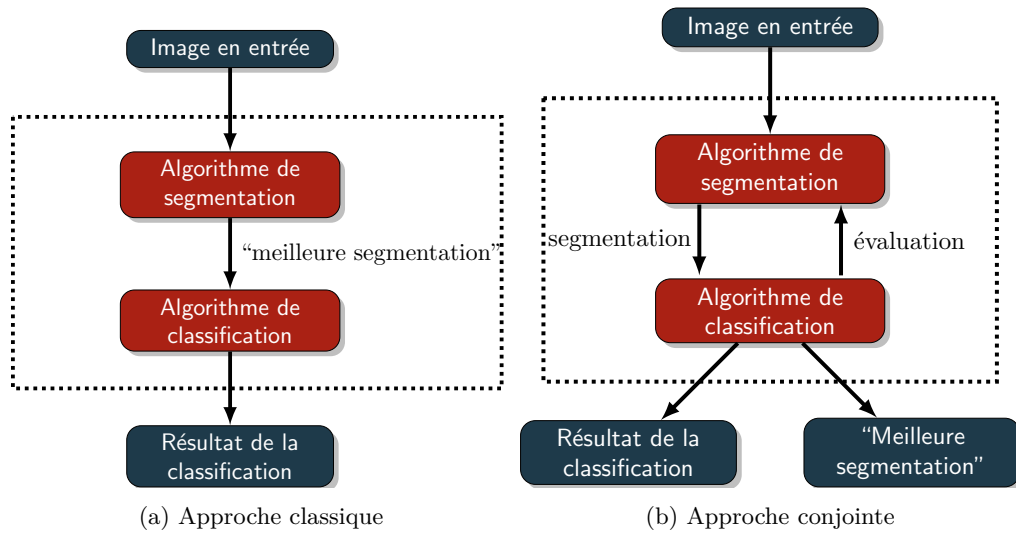


FIGURE 3.1 – Comparaison des approches d’analyse d’images [Farmer and Jain, 2005].

Quelques approches plus générales pour l’extraction mono-classe ont ainsi été proposées. Par exemple dans [Belarte et al., 2013], les auteurs proposent une méthode pour l’apprentissage de règles de classification floues à partir d’exemples d’une classe fournis par un expert.

Comme nous l’avons déjà évoqué, la segmentation de l’image joue un rôle crucial dans le paradigme d’analyse d’images basée objet (OBIA). Il est nécessaire de chercher la segmentation qui permet de classifier au mieux l’image. Une idée de plus en plus étudiée consiste à combiner les deux techniques (classification et segmentation), afin d’améliorer les deux résultats simultanément. [Farmer and Jain, 2005] présente un cadre de segmentation guidée par la classification. La particularité de cette approche est d’effectuer la segmentation de manière conjointe avec la classification, comme le montre la figure 3.1b. Les approches classiques effectuent la segmentation puis la classification de manière séquentielle, en supposant que la segmentation obtenue, utilisée pour la classification, est la meilleure. En revanche, l’approche conjointe, effectue des aller-retours entre la segmentation et la classification, en utilisant la qualité de la classification pour mesurer la qualité de la segmentation et ainsi guider le processus. Néanmoins, dans cette approche, la position de l’objet est approximativement connue. De plus, elle suppose qu’il existe un sous-ensemble de segments initiaux pour reconstruire l’objet d’intérêt. Cette idée a également été explorée par [Sellaouti et al., 2012]. Les auteurs emploient un critère basé sur la classification pour sélectionner les graines pour amorcer un processus de segmentation par croissance de régions hiérarchique et itératif. Cette approche se base sur un modèle de classification basé sur des critères pouvant être très spécifiques au jeu de données et donc difficile à reproduire ou généraliser.

Par ailleurs, ces méthodes supposent que la segmentation initiale est sur-segmentée et ne tiennent compte que de la classification pour guider la segmentation. Nous proposons un cadre générique Collaboratif de Segmentation et de Classification (CoSC) pour l’analyse mono-classe d’images de télédétection. Le processus est guidé par la qualité de la segmentation et de la classification. De plus, la segmentation initiale peut être *a priori* quelconque.

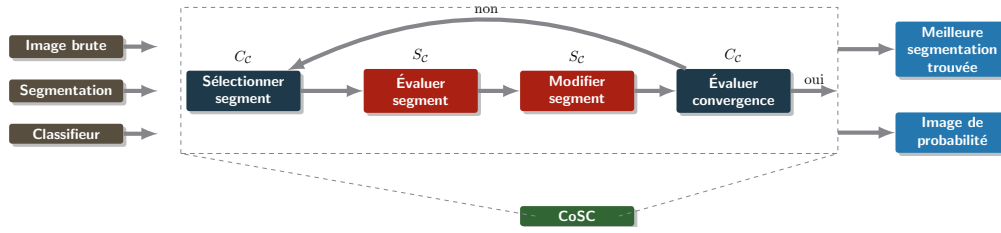


FIGURE 3.2 – Schéma conceptuel de CoSC.

3.2 Proposition

Les paradigmes de segmentation et de classification ont des objectifs assez différents, mais ils sont étroitement liés dans le contexte de la télédétection. Notre intuition nous dicte que des interactions entre ces deux approches devraient permettre une amélioration conjointe de leurs résultats, puisque la segmentation idéale est celle qui permet de mieux classer l'image et une classification parfaite produit une bijection entre les segments et les objets géographiques représentés dans l'image.

Ainsi, notre cadre collaboratif CoSC est spécialisé dans l'extraction d'une seule classe thématique C_k choisie par l'expert. Dans la suite, nous supposons disposer :

- d'une segmentation initiale $\mathcal{S} = \{R_i \mid 0 < i < N_R\}$ où N_R est le nombre de segments et chaque segment R_i est défini par $R_i = \{(x_k^i, y_k^i) \mid 0 < k < M_i\}$ avec M_i le nombre de pixels dans R_i ;
- d'un classifieur un-contre-tous \mathcal{C}_{C_k} entraîné pour discriminer la classe C_k .

Le processus que nous proposons, illustré dans la figure 3.2 consiste alors à sélectionner un segment selon un critère donné et à le modifier localement en fonction de sa qualité pour tenter d'améliorer sa classification. Ces deux étapes sont répétées jusqu'à convergence.

3.2.1 Définitions

Nous posons maintenant quelques définitions nécessaires pour décrire formellement le cadre proposé.

Définition 1 *Un segmenteur spécialisé SC_k est un agent (processus) de segmentation capable d'évaluer (par rapport aux caractéristiques de la classe C_k) et de modifier localement un segment R_i donné.*

Définition 2 *Un extracteur de classe CC_k est un agent de classification capable de déterminer à partir de \mathcal{C}_{C_k} , la probabilité $P_{C_k}(R)$ qu'un segment $R \in \mathcal{S}$ appartienne à la classe C_k .*

Définition 3 *Un agent de collaboration SC_{C_k} est un agent capable de gérer la collaboration entre un segmenteur spécialisé SC et un extracteur de classe CC_k .*

Définition 4 *Soient deux seuils $T_{\neq} < T_{\in}$ compris dans $[0, 1]$. Un segment $R_i \in C_k$ (resp. $R_i \notin C_k$) ssi $P_{C_k}(R_i) \geq T_{\in}$ (resp. $P_{C_k}(R_i) \leq T_{\neq}$)¹.*

Définition 5 *Un segment R tel que $T_{\neq} < P_{C_k}(R_i) < T_{\in}$ est dit ambigu.*

1. Ces deux seuils définissent donc une zone de rejet [Chow, 1970] sur \mathcal{C}_{C_k} .

3.2.2 Le processus CoSC

Le processus collaboratif, réalisé par SC_{C_k} , consiste à échanger des informations entre un segmenteur spécialisé S_{C_k} et un extracteur de classe C_{C_k} de sorte que :

1. S_{C_k} propose une segmentation initiale \mathcal{S} quelconque.
2. Une étape de modifications locales² est alternée avec une étape d'évaluation jusqu'à ce que la méthode converge :
 - (a) Dans l'étape de modification locale t , C_{C_k} classe les segments de la segmentation \mathcal{S}_t selon leur probabilité d'appartenance $P_C(R_i)$ et sélectionne un segment s_c suivant un critère donné et le communique au segmenteur. Le segmenteur S_C évalue s_c de manière locale pour déterminer si s_c est sur-, sous-, ou bien segmenté ; il modifie alors s_c en conséquence pour essayer d'améliorer sa qualité.
 - (b) Dans l'étape d'évaluation, la segmentation modifiée \mathcal{S}_{t+1} est transmise à l'extracteur C_{C_k} qui la classe et l'évalue de façon non-supervisée.
 - (c) Si le processus a convergé alors FIN (la meilleure segmentation/classification a été trouvée) sinon, l'étape 2 est ré-itérée.

Dans les sous-sections suivantes nous présentons en détail ces différentes étapes.

3.2.2.1 Modifications locales

À chaque itération, un segment s_c est choisi par l'extracteur C_{C_k} pour être candidat à modification. Il est crucial de sélectionner un "bon" segment candidat, autrement dit, un segment dont la modification pourrait réellement améliorer la classification. Nous verrons dans la section 3.3 que l'on peut mettre en œuvre plusieurs stratégies comme, par exemple, sélectionner un segment de manière aléatoire ou sélectionner un segment avec une probabilité très proche de T_C ou de $T_{\mathcal{C}}$.

On définit un opérateur de modification locale \mathcal{O} (fusion de segments, suppression de segment, etc.) comme une fonction $\mathcal{O} : \mathfrak{D}_i \rightarrow \mathfrak{D}_i$ où $\mathfrak{D}_i = R_i \cup \mathfrak{N}_{R_i}$ et \mathfrak{N}_{R_i} est l'ensemble de points appartenant aux segments dans le voisinage de R_i (segments adjacents).

Soient O_L , U_L , et W_L , trois listes de tels opérateurs \mathcal{O} . Lors de l'étape de modification, le segment s_c est évalué. Selon qu'il soit sur- (resp. sous-, resp. bien) segmenté, un opérateur choisi aléatoirement parmi les opérateurs \mathcal{O} dans la liste O_L , (resp. U_L , resp. W_L) est appliqué (cf. Algo. 2 et 3).

Algorithme 2 : ApplyModifications

```

Data : Segment :  $s_c$ , List of  $\mathcal{O} : L$ 
Result : Segment :  $R_m$ 
begin
   $L.shuffle()$ 
  while  $L.iterator.has\_next()$  do
     $R_m \leftarrow modify(s_c, L.iterator.next())$ 
    if  $R_m \neq s_c$  then
      | return  $R_m$ 
  return  $s_c$ 

```

2. Lors de la première itération, l'étape 2.a n'est pas effectuée : $S_1 = S_0$.

Algorithme 3 : LocalModification

```

Data : Segment :  $s_c$ 
Result : Segment :  $R_m$ 
begin
  if oversegmentation( $s_c$ ) then
    |  $R_m \leftarrow \text{ApplyModifications}(s_c, O_L)$ 
  else if undersegmentation( $s_c$ ) then
    |  $R_m \leftarrow \text{ApplyModifications}(s_c, U_L)$ 
  else
    |  $R_m \leftarrow \text{ApplyModifications}(s_c, W_L)$ 
  return  $R_m$ 

```

À titre d'exemple, dans l'étude présentée dans la section 3.4, O_L est réduite à un opérateur de fusion \mathcal{F} basé sur la distance euclidienne dans l'espace $L * a * b$ [Chen and Wang, 2004]; U_L est réduite à un opérateur d'érosion morphologique \mathcal{E} ; et W_L contient trois opérateurs : \mathcal{F} , \mathcal{E} ainsi qu'un opérateur de dilatation \mathcal{D} . Dans ce cas, si s_c est :

- sur-segmenté, \mathcal{F} est appliqué;
- sous-segmenté, \mathcal{E} est appliqué;
- bien segmenté, \mathcal{F} , \mathcal{E} et \mathcal{D} sont appliqués dans un ordre aléatoire, le premier qui modifie effectivement le segment est conservé (si la topologie du segment est changée par un opérateur, la modification n'est pas prise en compte).

3.2.2.2 Convergence

La qualité de la solution courante est évaluée de façon non-supervisée, c'est-à-dire, sans l'utilisation de connaissances externes (données de référence) qui sont souvent manquantes, incertaines ou incomplètes en télédétection. Cette fonction d'évaluation peut être définie de plusieurs façons et prendre en compte, par exemple, des critères basés sur la classification, sur la segmentation, ou sur les deux, comme nous le montrons dans la Section 3.3.

Pour éviter une convergence prématurée, le processus autorise un nombre D de modifications dégradantes, en espérant sortir des minima locaux avec :

$$D = \frac{\mathcal{U} \times N_R}{3} \quad (3.1)$$

où \mathcal{U} est le pourcentage de segments ambigus. Quand le processus atteint D étapes consécutives sans amélioration, le processus effectue un retour en arrière vers la meilleure solution trouvée. La convergence est atteinte lorsque plus aucune amélioration n'est apportée à la meilleure solution courante. Il est à noter que D diminue proportionnellement au nombre de segments ambigus. C'est en quelque sorte une procédure d'optimisation par recuit simulé simplifiée et adaptée aux particularités des données de télédétection.

3.2.3 Données en sortie

À l'issue du processus, le résultat se compose de la segmentation modifiée et d'une image de probabilité où chaque pixel représente la probabilité d'appartenance de celui-ci à la classe \mathcal{C}_k .

En résumé, CoSC est par construction une méthode générique pouvant être appliquée facilement pour extraire des objets de différents types (une classe à la fois). Le processus repose sur

des critères de qualité afin de choisir le segment candidat et pour appliquer les modifications locales à celui-ci.

3.3 Implémentation du processus CoSC

Afin d'étudier le comportement de CoSC, nous en avons implémenté plusieurs variantes. Ainsi, nous avons étudié l'influence des différents paramètres sur le comportement de la méthode, ainsi que sur les résultats de celle-ci. Pour cela, nous avons testé toutes les combinaisons possibles des différents paramètres sur quatre classes thématiques : Eau, Végétation, Bâtiments et Routes. Les résultats sont présentés en Section 3.3.2.

3.3.1 Les différents paramètres

Une instance entièrement définie de CoSC dépend de quatre paramètres :

- le modèle de classification, noté P_c ;
- la stratégie de modification, notée P_m ;
- la stratégie d'évaluation de la solution courante (fonction objectif), notée P_e ;
- la stratégie de sélection du segment candidat, notée P_s .

Dans ce qui suit nous rappelons le rôle de chaque paramètre et nous présentons les choix implémentés.

3.3.1.1 Modèle de classement P_c

Le modèle de classement fournit après un apprentissage supervisé préalable et indépendant de CoSC, la probabilité qu'un segment appartienne à la classe thématique recherchée. Deux types de modèles ont été testés durant cette étude :

1. Arbre de Hoefding ($P_c = \text{VFDT}$) : des arbres incrémentaux de Hoefding ou Very Fast Decision Tree (VFDT) [Hulten et al., 2001] ont été entraînés à partir d'exemples et contre-exemples (entre 10 et 15) construits de manière interactive par un expert pour chaque classe.
2. Perceptron multi-couche ($P_c = \text{MLP}$) : des perceptrons multi-couche ou Multi-Layer Perceptron (MLP) [Haykin, 1998], avec deux couches cachées et 15 nœuds par couche, ont été entraînés à partir de données de référence. L'objectif à prédire est le pourcentage de la classe en question à l'intérieur de chaque segment.

Aucune procédure d'optimisation sur les méta-paramètres de ces algorithmes n'a été effectuée. Les modèles de prédiction obtenus ne sont donc probablement pas optimaux, mais ils ont été suffisants pour démontrer l'applicabilité de la méthode proposée.

3.3.1.2 Stratégie de modification P_m

Une stratégie de modification consiste à définir les trois listes d'opérateurs de modification O_L , U_L et W_L qui seront appliqués au segment candidat s_c à chaque étape de la collaboration. Pour rappel, O_L doit contenir des opérateurs visant à réduire la sur-segmentation, U_L doit contenir des opérateurs visant à réduire la sous-segmentation et W_L peut contenir des opérateurs quelconques qui sont appliqués lorsque s_c n'est pas détecté comme sur- ou sous-segmenté. Nous avons implémenté quatre opérateurs de modification.

Fusion : cet opérateur vise principalement à réduire la sur-segmentation. Il consiste à fusionner s_c avec son plus proche voisin en termes de propriétés radiométriques. Plus concrètement, le plus proche voisin est celui dont la distance euclidienne calculée sur les moyennes de bandes radiométriques est minimale.

Scission : cet opérateur vise principalement à réduire la sous-segmentation. Il consiste à appliquer la méthode de super-pixels SLIC [Achanta et al., 2012] contrainte aux pixels de s_c . SLIC favorise la création de segments plutôt compacts et qui adhèrent aux contours présents dans l'image. Le nombre de nouveaux segments créés est calculé en fonction de la taille de s_c : plus il est grand, plus le nombre de nouveaux segments sera élevé.

Rétrécissement : Cet opérateur vise principalement à réduire la sous-segmentation. Il consiste à appliquer une opération d'érosion morphologique sur s_c en utilisant comme élément structurant un carré 3×3 . Ces pixels érodés sont redistribués aux voisins spatialement plus proches, afin de conserver une partition de l'image. Si l'érosion entraîne une rupture de la connectivité de s_c l'opération est refusée afin de respecter la définition de la segmentation.

Croissance : Cet opérateur vise principalement à réduire la sur-segmentation. Il consiste à appliquer une opération de dilatation morphologique sur s_c en utilisant comme élément structurant un carré 3×3 . Ces pixels dilatés sont retranchés aux segments auxquels ils appartenaient, afin de conserver une partition de l'image. Si la dilatation entraîne un changement dans la topologie de s_c ou de ses voisins, l'opération est refusée afin de toujours respecter la définition de la segmentation.

En partant de ces opérateurs de base, nous implémentons les deux stratégies de modification en définissant deux ensembles de listes différents.

1. Stratégie sans scission ($P_m = M_F$) :
 - O_L : Fusion ;
 - U_L : Rétrécissement ;
 - W_L : Fusion, Rétrécissent, Croissance.
2. Stratégie avec scission ($P_m = M_S$) :
 - O_L : Fusion ;
 - U_L : Scission ;
 - W_L : Rétrécissent, Croissance.

L'application de ces stratégies sont équivalentes aux algorithmes 4 et 5.

Algorithme 4 : Stratégie de modification sans scission.

Data : Segment s

Data : Liste de Segments *neighbours*

Result : Segment modifié s'

begin

if s est sur-segmenté **then**

 └ Fusionner s avec son voisin le plus similaire

else if s est sous-segmenté **then**

 └ Rétrécir s

else

 └ Appliquer à s un opérateur pris au hasard parmi fusion, rétrécissement ou croissance

Algorithme 5 : Stratégie de modification avec scission.

Data : Segment s
Data : List of Segment $neighbours$
Result : Modified Segment s'
begin
 if s est sur-segmenté **then**
 └ Fusionner s avec son voisin le plus similaire
 else if s est sous-segmenté **then**
 └ Diviser s en appliquant la méthode SLIC
 else
 └ Appliquer à s un opérateur au hasard parmi rétrécissement ou croissance

3.3.1.3 Stratégie d'évaluation P_e

La stratégie d'évaluation est la fonction d'évaluation de la qualité du résultat courant à chaque étape du processus CoSC. Autrement dit, c'est la fonction objectif que le processus vise à optimiser. Nous avons défini et étudié trois fonctions différentes.

1. Critère basé sur la classification ($P_e = Q_{clsf}$) : cette fonction se base sur la probabilité de classification des segments et cherche intuitivement à réduire le nombre de segments ambigus (définition 5), c'est-à-dire, les segments pour lesquels le classifieur n'arrive pas à déterminer avec certitude la classe. Ce critère est la probabilité moyenne qu'un segment soit correctement classé (positivement ou négativement). Elle est définie par :

$$Q_{clsf} = \frac{1}{N_R} \left(\sum_{i|P_{C_k}(R_i) > T_\epsilon} P_{C_k}(R_i) + \sum_{i|P_{C_k}(R_i) < T_\epsilon} 1 - P_{C_k}(R_i) \right) \quad (3.2)$$

où N_R est le nombre de segments dans la segmentation courante. Ainsi, plus la classe d'un segment R_i est certaine, plus R_i contribue au score. *A contrario*, les segments ambigus sont pénalisés de manière implicite puisqu'ils ne contribuent pas au score, mais ils sont pris en compte dans le facteur $\frac{1}{N_R}$. Le cas idéal où tous les segments sont classés avec 100% de certitude donne $Q_{clsf} = 1.0$. Dans le pire des cas, où tous les segments sont ambigus, alors $Q_{clsf} = 0.0$.

2. Critère basé sur la segmentation ($P_e = Q_{seg}$) : cette fonction se base sur la métrique UOA_{L_2} (Eq. 1.24). Elle ne tient compte que de la segmentation et favorise donc la réduction des erreurs de segmentation. Elle est définie par :

$$Q_{seg} = 1 - UOA_{L_2} \quad (3.3)$$

3. Critère de qualité mixte ($P_e = Q_{mix}$) : cette fonction cherche à améliorer simultanément la classification et la segmentation en combinant les deux critères définis précédemment. C'est la moyenne harmonique de Q_{clsf} et Q_{seg} . Elle est définie comme suit :

$$Q_{mix} = 2 \cdot \frac{Q_{clsf} \cdot Q_{seg}}{Q_{clsf} + Q_{seg}} \quad (3.4)$$

3.3.1.4 Sélection du segment candidat P_s

À chaque étape de la collaboration, un segment s_c est sélectionné pour être modifié ; nous appelons s_c le segment candidat. Nous avons implémenté quatre stratégies de sélection différentes.

1. Par sélection aléatoire ($P_s = s_{alea}$) : consiste à choisir s_c de manière aléatoire parmi tous les segments de \mathcal{S} .

$$s_{alea} = \mathcal{S}[\mathcal{U}\{0, N\}] \quad (3.5)$$

où \mathcal{U} dénote une distribution uniforme et N est le nombre de segments dans \mathcal{S} .

2. Au moins ambigu ($P_s = s_{amb}^-$) : consiste à choisir le segment le plus proche de T_ϵ ou de T_ζ .

$$s_{amb}^- = \arg \min_R (|P_C(R) - T_\epsilon|, |P_C(R) - T_\zeta|) \quad (3.6)$$

3. Au plus ambigu ($P_s = s_{amb}^+$) : consiste à choisir le segment le plus éloigné de T_ϵ et T_ζ .

$$s_{amb}^+ = \arg \min_R P_C(R) - \frac{T_\epsilon + T_\zeta}{2} \quad (3.7)$$

4. Au pire segmenté ($P_s = s_{seg}$) : consiste à employer l'évaluation locale à granularité fine de UOA (cf., Section. 1.3.1).

$$s_{seg} = \arg \max_R \phi_{\delta_{fine}}(R) \quad (3.8)$$

La fonction $\phi_{\delta_{fine}}$ est définie dans la section 1.3.1 (Eq. 1.22).

3.3.2 Étude du comportement de CoSC en fonction de ses paramètres

Dans cette étude, nous avons employé l'image présentée dans la figure 3.3, ainsi que les données de référence pour chaque classe. Les données de référence ont été obtenues à partir de la procédure décrite dans la section 2.1, sans attribuer un label définitif aux segments, mais en conservant le pourcentage d'intersection de chaque classe. Ce sont donc des données de référence floues. Nous avons donc choisi d'utiliser la métrique de précision floue (*fuzzy accuracy*) [Binaghi et al., 1999] pour mesurer les performances de chaque variante testée.

La méthodologie expérimentale est la suivante. Pour chaque permutation possible des différents paramètres (deux pour P_c , deux pour P_m , trois pour P_e et quatre pour P_s) et pour chacune des quatre classes, nous avons instancié un processus CoSC et avons observé l'évolution du critère d'évaluation au fil du temps (en termes d'itérations), le temps d'exécution réel et la précision floue. De fait, 192 instances différentes de CoSC ont donc été testées. Il est à noter que le coût en termes de temps pour certaines instances est assez élevé. Ainsi, nous n'avons effectué qu'une observation par instance : nous avons préféré explorer un maximum de possibilités lors de cette première étude afin de mieux comprendre l'effet des différents paramètres et ainsi repérer des configurations intéressantes. Nous étudions principalement à trois aspects : l'optimisation du critère d'évaluation, le temps d'exécution et la performance par rapport à des données de référence.

- Les figures 3.5 (resp. 3.8) montrent le comportement de CoSC pour $P_c = VFDT$ et $P_m = M_F$ (resp. $P_c = MLP$ et $P_m = M_S$) avec les différentes configurations des autres paramètres possibles. Dans chaque graphe, l'axe des x représente le nombre d'itérations effectuées à un moment donné. L'axe des y représente le score de la fonction d'évaluation en question. Prenons par exemple le sous-graphe en haut à gauche de la figure 3.5. Il correspond aux paramètres $P_c = VFDT$, $P_m = M_F$, $P_e = Q_{clsf}$ et $P_s = s_{amb}^-$. On observe que la qualité des quatre classes évolue de manière plus ou moins logarithmique, les courbes sont croissantes de manière presque linéaire

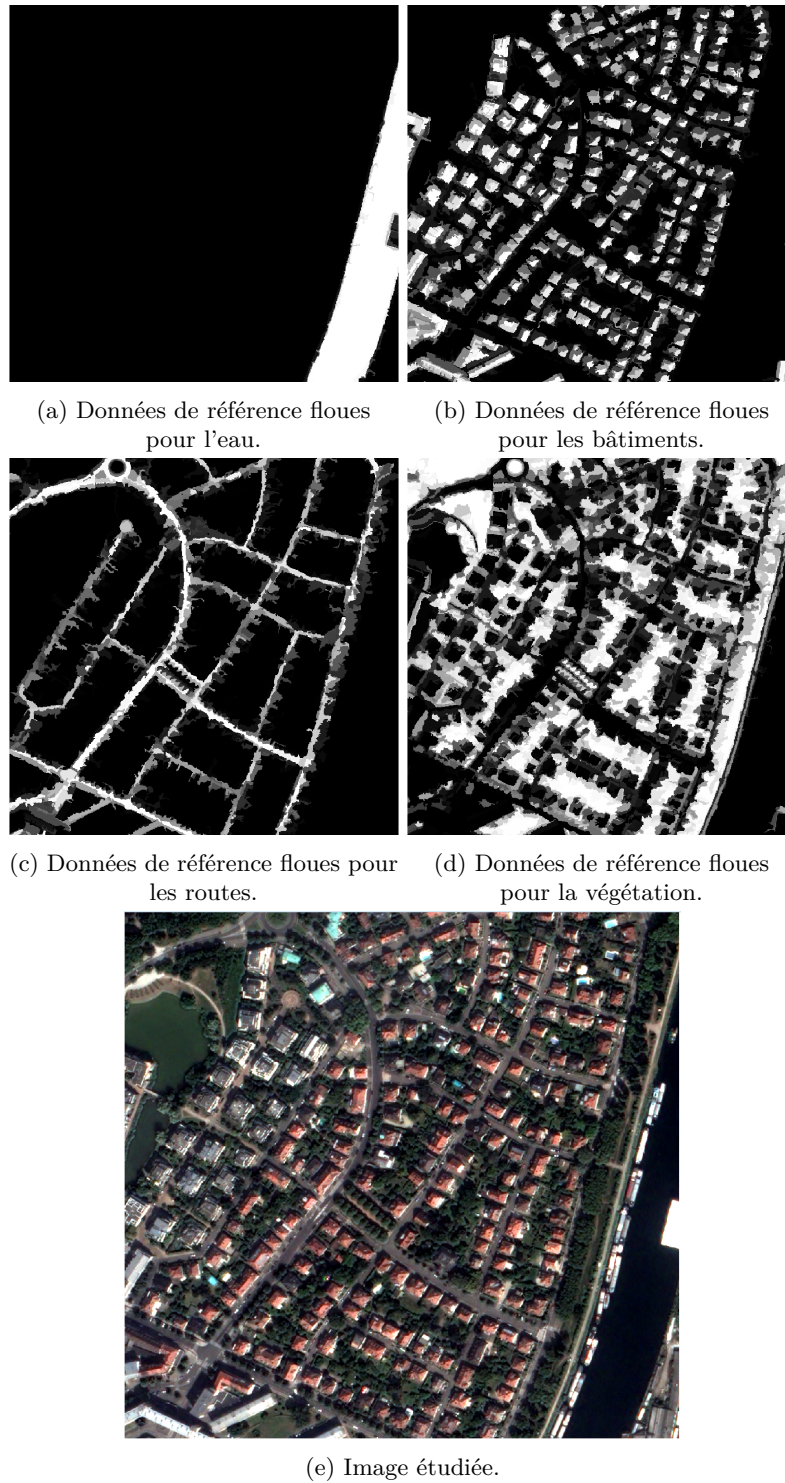


FIGURE 3.3 – Données de référence floues et l'image étudiée.

pendant un certain nombre d'itérations au début, puis elles croissent plus lentement jusqu'à l'arrêt du processus. On remarque que l'extraction de l'eau converge plus rapidement que celle des autres classes, suivie par les bâtiments, les routes et finalement la végétation. Cette différence s'explique certainement par deux facteurs :

- le premier est lié au modèle de classification. Si le taux de segments ambigus est très faible dès le début le processus convergera plus rapidement (c'est le cas pour la classe Eau) ;
- le second est lié à la capacité de la stratégie de sélection P_s à trouver des candidats pertinents pour améliorer P_e .

Pour cet exemple particulier, le processus d'optimisation se passe bien pour les quatre classes. Néanmoins, ce n'est pas toujours le cas. Regardons le sous-graphe dans la troisième ligne (du haut vers le bas) et la troisième colonne (de gauche à droite) de la figure 3.7. Il correspond à l'instance avec les paramètres $P_c = \text{MLP}$, $P_m = M_S$, $P_e = Q_{seg}$ et $P_s = s_{alea}$. Le processus d'extraction de la végétation est le seul à optimiser le critère Q_{seg} de manière remarquable. Il passe de 0.45 à 0.53 après un peu plus de 10 000 itérations. Le processus d'extraction des bâtiments commence avec un score de 0.46 mais chute brutalement en dessous 0.35, puis le score remonte lentement. Néanmoins, le processus s'arrête avant que la meilleure valeur trouvée soit améliorée à nouveau. Il en est de même pour les processus d'extraction des routes et de l'eau, qui présentent le même type de chute du score mais à des itérations différentes. Cette anomalie du comportement de Q_{seg} pourrait s'expliquer par le fait que la stratégie de scission introduit des erreurs de sur-segmentation plus rapidement que CoSC n'est capable de les corriger. Par ailleurs, une autre particularité à noter est que les valeurs et les courbes des processus utilisant le critère d'évaluation Q_{mix} reflètent bien qu'il s'agit d'une combinaison linéaire de Q_{clsf} et Q_{seg} .

Dans les tableaux 3.1, à 3.4 nous présentons une brève description des constatations les plus remarquables.

- La figure 3.9 montre le temps d'exécution de chaque processus pour les différentes configurations testées. Chaque colonne correspond à une stratégie de sélection, chaque ligne correspond à une stratégie d'évaluation et chaque sous-graphe correspond à une classe. La couleur de chaque cellule varie de blanc (temps d'exécution faible) à violet (temps d'exécution élevé). On remarque que les processus faisant intervenir l'évaluation de la segmentation (soit avec $P_e = Q_{seg}$ ou $P_e = Q_{seg}$ soit avec $P_s = s_{seg}$) ont généralement un temps d'exécution plus important. D'autres paramètres tels que le couple (modèle de classement, classe) ou la stratégie de modification semblent avoir également un effet (moins évident) sur le temps d'exécution.

- La figure 3.10 présente la précision floue calculée par rapport aux données de référence. Il n'y a pas de patron évident qui se dégage de ces figures. On constate tout de même que tous les résultats se situent entre 0.730 (obtenu pour l'extraction des routes avec $P_c = \text{VFDT}$, $P_m = M_S$, $P_e = Q_{seg}$ et $P_s = s_{seg}$) et 0.926 (obtenu pour l'extraction de l'eau avec $P_c = \text{MLP}$, $P_m = M_S$, $P_e = Q_{seg}$ et $P_s = s_{seg}$). On remarque que ces deux configurations ne diffèrent que par le modèle de classement et la classe extraite. Cela laisse à penser que ce couple de paramètres influe de manière importante sur les résultats en termes de classification. En effet, si le modèle fait une classification presque parfaite dès la première étape, alors CoSC converge très rapidement, mais sa classification sera très performante tout de même. Au contraire, si le modèle établit des prédictions aléatoires ou très peu pertinentes, CoSC va itérer un grand nombre de fois sans jamais réussir à améliorer la classification. À titre d'exemple, nous montrons dans la figure 3.4 les meilleurs résultats obtenus d'après la précision floue pour les quatre classes étudiées. Nous observons que les prédictions pour l'eau et la végétation sont effectivement plus certaines que pour les bâtiments et les routes. Cela peut expliquer les différents résultats en termes de précision floue pour chaque classe.

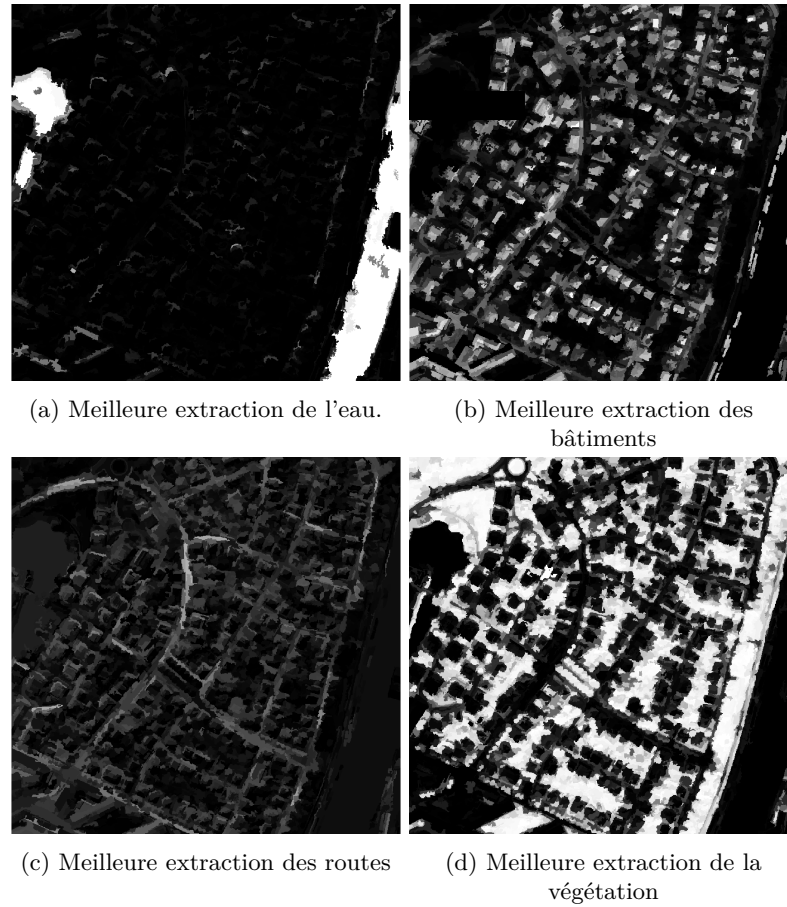
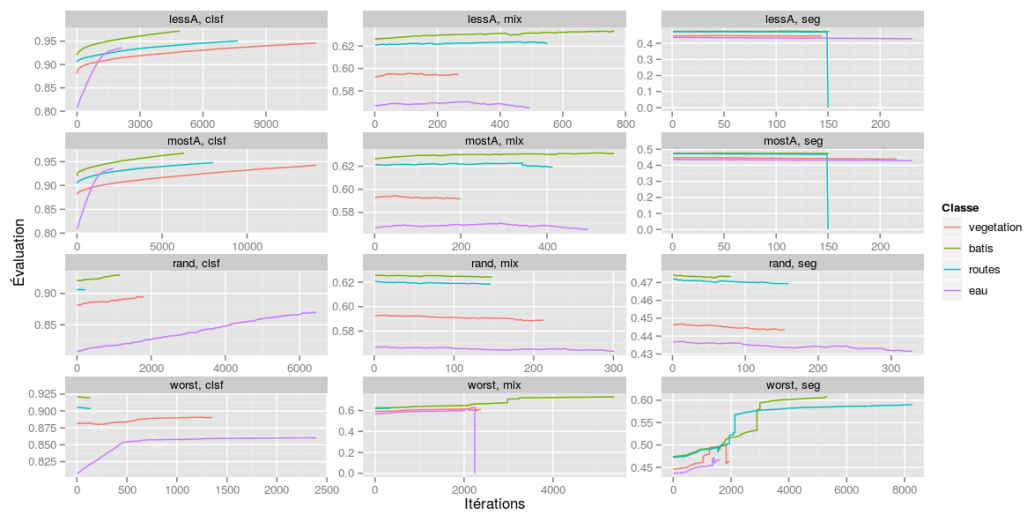
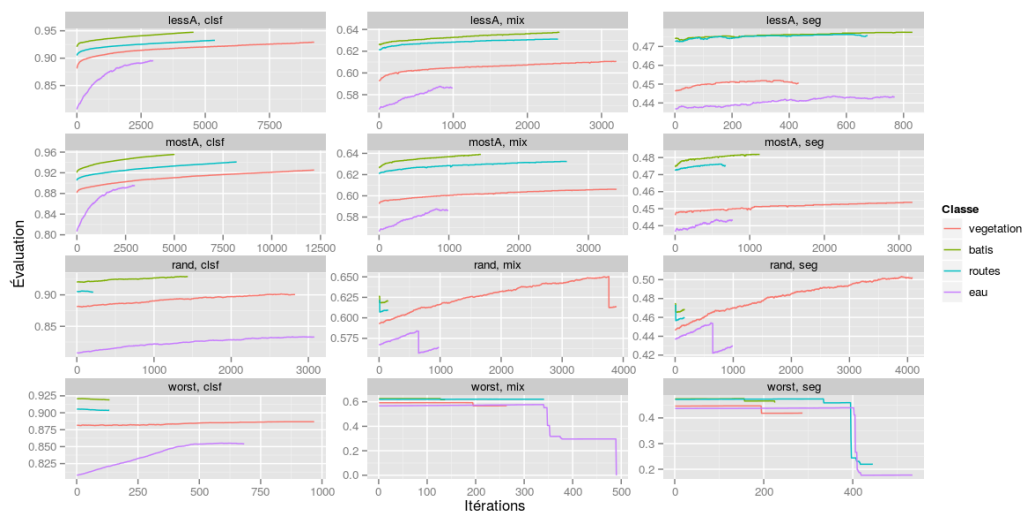
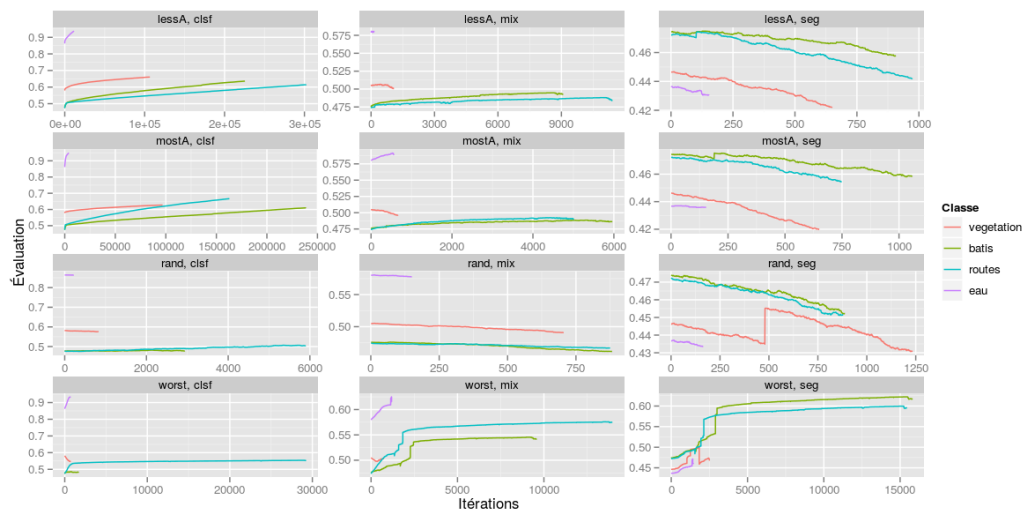


FIGURE 3.4 – Meilleures classifications mono-classes obtenues avec des stratégies différentes.

FIGURE 3.5 – Comportement de CoSC pour $P_c = \text{VFDT}$ et $P_m = M_F$.

FIGURE 3.6 – Comportement de CoSC pour $P_c = \text{VFDT}$ et $P_m = M_S$.FIGURE 3.7 – Comportement de CoSC pour $P_c = \text{MLP}$ et $P_m = M_F$.

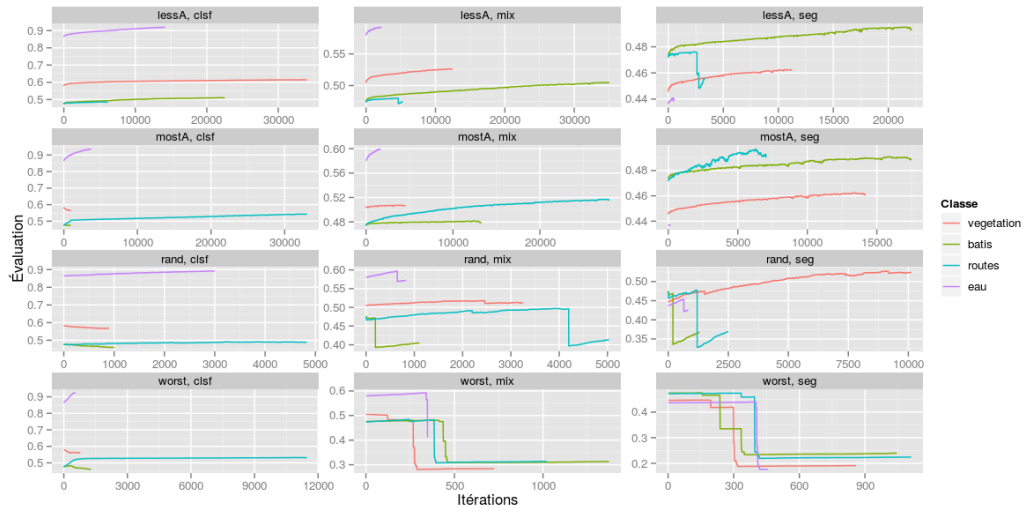
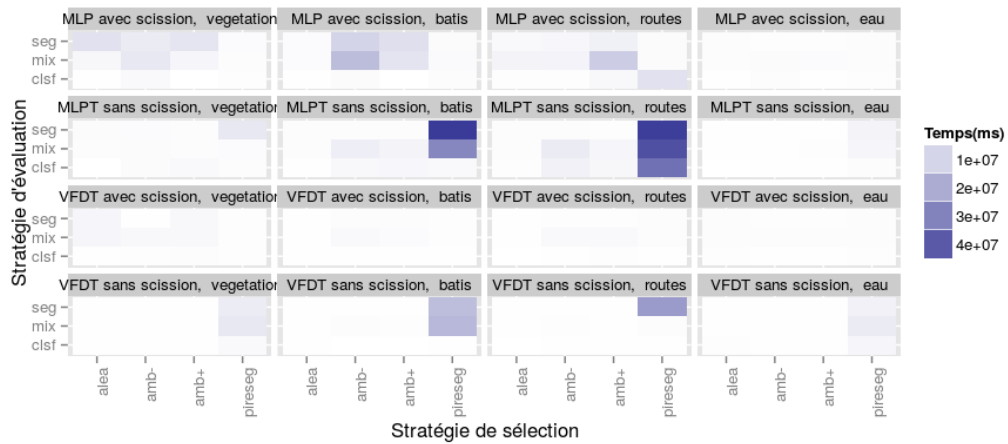
FIGURE 3.8 – Comportement de CoSC pour $P_c = \text{MLP}$ et $P_m = M_S$ 

FIGURE 3.9 – Temps d'exécution (millisecondes) de CoSC pour les différentes configurations testées.

TABLEAU 3.1 – Synthèse du comportement de CoSC avec certains paramètres.

P_c	P_m	P_e	P_s	Observations
VFDT	M_F	Q_{clsf}	s_{amb}^-	Les processus des quatre classes évoluent de manière similaire : les courbes sont croissantes de manière presque linéaire au début, puis elles croissent plus lentement jusqu'à l'arrêt du processus. Le processus d'extraction de l'eau est le plus rapide à converger (en nombre d'itérations), suivi par celui des bâtiments, puis celui des routes et enfin celui de la végétation.
			s_{amb}^+	Évolution des processus très similaire à s_{amb}^- . La seule différence semble être le nombre d'itérations avant convergence qui est ici plus élevé. L'ordre de convergence est identique.
			s_{alea}	Les courbes des quatre processus croissent de manière linéaire ; le processus s'arrête plus rapidement mais avec des valeurs de Q_{clsf} notablement plus faibles.
			s_{seg}	Les processus d'extraction des routes et des bâtiments ne parviennent pas à améliorer le critère Q_{clsf} et le processus s'est arrêté très rapidement en renvoyant la segmentation initiale. Le processus d'extraction de la végétation et de l'eau s'est exécuté pendant un nombre plus élevé d'itérations mais, il n'y a pas eu une amélioration considérable du critère Q_{clsf} .
		Q_{seg}	s_{amb}^-	Pas d'amélioration remarquable, les quatre processus s'arrêtent très rapidement, en moins de 300 itérations.
			s_{amb}^+	Pas d'amélioration remarquable, les processus s'arrêtent très rapidement, en moins de 300 itérations.
			s_{alea}	Pas d'amélioration remarquable, les processus s'arrêtent très rapidement, en moins de 400 itérations.
			s_{seg}	Les processus se sont exécutés pendant plus d'itérations qu'avec les autres stratégies de sélection. Il y a eu une amélioration considérable du critère Q_{seg} . Notamment pour le processus d'extraction des bâtiments et des routes, et un peu moins important pour celui de la végétation et de l'eau. Les courbes présentent des « sauts » assez sensibles à certains endroits.
		Q_{mix}	s_{amb}^-	Les quatre processus se sont arrêtés en moins de 800 itérations sans amélioration remarquable de Q_{mix} .
			s_{amb}^+	Les quatre processus se sont arrêtés en moins de 500 itérations sans amélioration remarquable de Q_{mix} .
			s_{alea}	Les quatre processus se sont arrêtés en moins de 300 itérations sans amélioration remarquable de Q_{mix} .
			s_{seg}	Les processus d'extraction des bâtiments, de la végétation et de l'eau se sont exécutés pendant plus de 2000 itérations avec une faible amélioration de Q_{mix} . Le processus d'extraction des routes s'est arrêté rapidement sans amélioration remarquable.

TABLEAU 3.2 – Synthèse du comportement de CoSC avec $P_c = \text{VFDT}$ et $P_m = M_S$.

P_c	P_m	P_e	P_s	Observations
VFDT	M_S	Q_{clsf}	s_{amb}^-	Les quatre courbes évoluent de manière similaire, elles croissent linéairement au début, puis plus lentement jusqu'à convergence.
			s_{amb}^+	Les quatre courbes se comportent de manière équivalente à celles obtenues avec s_{amb}^- , mais s'exécutent pendant un nombre plus élevé d'itérations.
			s_{alea}	Croissance linéaire et faible amélioration de Q_{clsf} pour l'extraction de la végétation, des bâtiments et de l'eau. Pas d'amélioration remarquable de Q_{clsf} pour l'extraction des routes.
			s_{seg}	Les quatre processus s'arrêtent très rapidement en moins de 1000 itérations. Pas d'amélioration de Q_{clsf} pour l'extraction des bâtiments, des routes et de la végétation. Faible amélioration de Q_{clsf} pour l'extraction de l'eau.
		Q_{seg}	s_{amb}^-	Évolution des courbes irrégulière, sans amélioration remarquable de Q_{seg} .
			s_{amb}^+	Évolution des courbes irrégulière avec une très faible amélioration de Q_{seg} .
			s_{alea}	Amélioration de Q_{seg} pour l'extraction de la végétation et de l'eau. Pas d'amélioration de Q_{seg} pour l'extraction des bâtiments et des routes, ces deux processus se sont arrêtés très rapidement.
			s_{seg}	Pas d'amélioration de Q_{seg} , les processus se sont arrêtés rapidement.
		Q_{mix}	s_{amb}^-	Les courbes croissent linéairement mais lentement. On constate une faible amélioration de Q_{mix} pour les quatre processus.
			s_{amb}^+	Comportement très similaire à celui de s_{amb}^- .
			s_{alea}	Les courbes ressemblent fortement à celles du processus avec $P_s = s_{alea}$ et $P_e = Q_{seg}$ mais à une autre échelle.
			s_{seg}	Pas d'amélioration du critère Q_{mix} . Les quatre processus se sont arrêtés rapidement.

TABLEAU 3.3 – Synthèse du comportement de CoSC avec $P_c = \text{MLP}$ et $P_m = M_F$.

P_c	P_m	P_e	P_s	Observations
MLP	M_F	Q_{clsf}	s_{amb}^-	Le processus d'extraction améliore le critère Q_{clsf} et converge plus rapidement que les autres processus. Les courbes croissent lentement sur un très grand nombre d'itérations.
			s_{amb}^+	Comportement très proche à celui avec s_{amb}^- mais avec une convergence en moins d'itérations.
			s_{alea}	Pas d'amélioration de Q_{clsf} pour l'extraction de la végétation et de l'eau. On observe une légère amélioration pour les routes et les bâtiments.
			s_{seg}	Pas d'amélioration de Q_{clsf} pour l'extraction de la végétation et des bâtiments. Amélioration et convergence très rapide pour l'extraction de l'eau et très lente pour l'extraction des routes.
		Q_{seg}	s_{amb}^-	Les quatre courbes ont un comportement irrégulier avec une tendance décroissante.
			s_{amb}^+	Les quatre courbes ont un comportement irrégulier avec une tendance décroissante.
			s_{alea}	Les quatre courbes ont un comportement irrégulier avec une tendance décroissante.
			s_{seg}	Les processus d'extraction des routes et des bâtiments améliorent considérablement Q_{seg} . Les deux autres processus présentent une amélioration de Q_{seg} moins importante et s'arrêtent plus tôt.
		Q_{mix}	s_{amb}^-	Faible amélioration de Q_{mix} par les quatre processus. Les processus d'extraction d'eau et de la végétation s'arrêtent rapidement.
			s_{amb}^+	Faible amélioration de Q_{mix} par les processus d'extraction de l'eau, des routes et des bâtiments. Les processus d'extraction d'eau et de la végétation s'arrêtent rapidement.
			s_{alea}	Aucun processus ne parvient à améliorer Q_{mix} .
			s_{seg}	Les processus d'extraction de l'eau, des routes et des bâtiments présentent une amélioration remarquable de Q_{mix} . Le processus d'extraction de la végétation s'arrête très rapidement sans amélioration.

TABLEAU 3.4 – Synthèse du comportement de CoSC avec $P_c = \text{MLP}$ et $P_m = M_S$.

P_c	P_m	P_e	P_s	Observations	
MLP	M_S	Q_{clsf}	s_{amb}^-	Les quatre processus améliorent très faiblement Q_{clsf} .	
			s_{amb}^+	Le processus d'extraction de l'eau améliore Q_{clsf} et converge rapidement. Le processus d'extraction des routes améliore Q_{clsf} très faiblement et converge lentement. Les processus d'extraction de la végétation et des bâtiments ne présentent pas d'amélioration de Q_{clsf} .	
			s_{alea}	Les processus d'extraction de l'eau et des routes améliorent très faiblement Q_{clsf} . Les processus d'extraction de la végétation et des bâtiments ne présentent pas d'amélioration de Q_{clsf} .	
			s_{seg}	Le processus d'extraction de l'eau améliore Q_{clsf} et converge rapidement. Le processus d'extraction des routes améliore également Q_{clsf} mais converge lentement. Les processus d'extraction de la végétation et des bâtiments ne présentent pas d'amélioration de Q_{clsf} .	
			Q_{seg}	s_{amb}^-	Amélioration remarquable de Q_{seg} pour les processus d'extraction des bâtiments et de la végétation. Faible amélioration de Q_{seg} pour l'extraction de l'eau et pas d'amélioration pour l'extraction des routes.
				s_{amb}^+	Les processus d'extraction de la végétation, des bâtiments et des routes présentent une amélioration de Q_{seg} .
		s_{alea}		Le processus d'extraction de la végétation présente une amélioration plutôt constante de Q_{seg} . Les trois autres processus améliorent Q_{seg} faiblement au début, puis la métrique chute abruptement.	
		s_{seg}		Pas d'amélioration remarquable de Q_{seg} pour aucun des processus.	
		Q_{mix}	s_{amb}^-	Faible amélioration de Q_{mix} pour les quatre processus.	
			s_{amb}^+	Les processus d'extraction de l'eau et des routes améliorent remarquablement Q_{mix} , mais celui des routes présente plus d'itérations avant convergence.	
			s_{alea}	Comportement irrégulier des quatre courbes. Des améliorations très légères de Q_{mix} pour l'extraction de l'eau, de la végétation et des routes.	
			s_{seg}	Pas d'amélioration notable de Q_{seg} pour les quatre processus.	

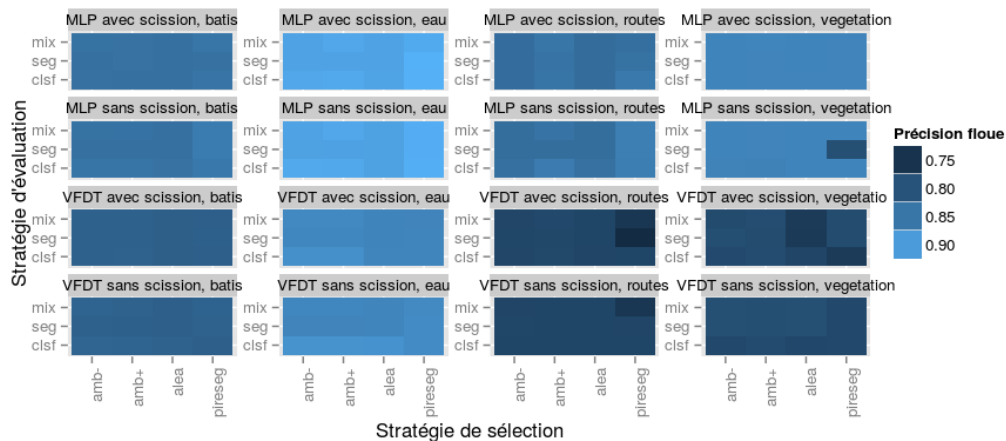


FIGURE 3.10 – Précision floue pour chaque configuration testée.

3.3.3 Bilan

De cette analyse nous pouvons tirer plusieurs conclusions.

- Le choix de la stratégie de sélection doit être fait en accord avec la stratégie d'évaluation employée. En effet, il est nécessaire de définir des stratégies permettant de sélectionner les segments les plus à même de maximiser le critère d'évaluation choisi, par exemple le plus ou le moins ambigu pour le critère basé sur la classification, ou le moins bien segmenté pour le critère basé sur la segmentation.
- À qualité égale, le type du modèle de classification n'a pas l'air d'influencer le comportement de CoSC, mais il a un impact sur le temps d'exécution. En effet les modèles plus complexes demandent plus de calculs qui se traduisent par de temps d'exécution plus élevés.
- Indépendamment du type du modèle, la qualité de celui-ci (en terme de classification) peut avoir une influence considérable sur le comportement de la méthode mais aussi sur le temps d'exécution. En effet, un modèle faisant des prédictions avec une très haute confiance va entraîner une convergence plus rapide du processus. En revanche un modèle faisant des prédictions très incertaines résultera en un processus beaucoup plus lent à converger et à des faibles améliorations du critère d'évaluation.
- La stratégie de modification avec scission ne semble pas adaptée pour améliorer le critère de segmentation, soit parce qu'elle introduit trop de sur-segmentation, soit parce qu'elle entraîne des modifications de segment inappropriées.

Nous observons que de manière générale l'extraction de la classe Eau obtient une meilleure précision floue que les autres classes. Généralement, le modèle MLP obtient de meilleures performances que le modèle VFDT pour toutes les classes, mais nous ne remarquons pas de couple de stratégies gagnant systématiquement. Pour le modèle VFDT ce sont souvent les stratégies de sélection du plus ou le moins ambigu, pour le modèle MLP c'est souvent celle du moins bien segmenté qui semblent les plus performantes. La stratégie de sélection aléatoire obtient, dans la plupart des cas, des performances plus faibles. Le critère combiné s'avère souvent une bonne option mais comme nous l'avons vu dans la section 3.3.2 il est très coûteux et demande des temps d'exécution plus élevés que les autres critères d'évaluation. En conclusion, la configuration qui

nous semble la plus stable avec des temps de calcul raisonnables est $P_e = Q_{clsf}$, $P_s = Q_{amb+}$ et $P_m = M_F$ avec le modèle de classement le plus adaptée selon la classe à extraire.

3.4 Étude détaillée d’une configuration

Afin de valider l’applicabilité de CoSC de manière plus approfondie, nous l’avons appliqué à l’extraction de la végétation dans une image de Strasbourg. Dans ce cas, il est facile de trouver un classifieur de bonne qualité. Nous réduisons ainsi le biais induit par le classifieur, puisque sa performance a un impact non négligeable sur les résultats de la méthode (voir Section 3.3.2).

3.4.1 Présentation des données

L’image étudiée (Figure 3.17a) est une image THRS Pleiades (©CNES) à 50cm, $9\,211 \times 11\,275$ (10^8 pixels) avec 4 bandes spectrales : rouge, vert, bleu et proche infra-rouge. La classe thématique recherchée est la végétation. Pour faciliter les calculs, nous avons découpé et traité l’image en 1 620 imageries de 256×256 pixels. Les résultats ont été reconstitués dans une seule image qui a été comparée à des données de référence disponibles (Figure 3.17b). Celles-ci ont été générées en superposant une sur-segmentation de l’image avec des données vectorielles de référence sur la végétation. Pour chaque pixel, on dispose d’une estimation du pourcentage de végétation qu’il contient.

Il est à noter qu’aucun traitement particulier n’a été effectué aux bords des imageries. Par conséquent, il existe quelques segments discontinus à ces endroits de l’image finale.

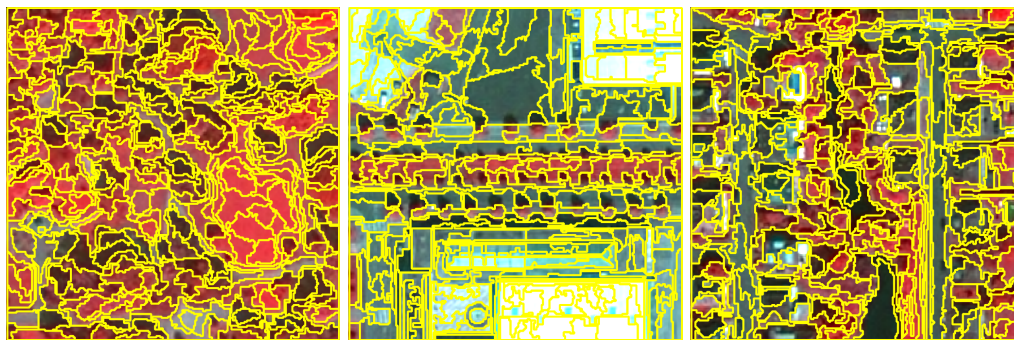
Les segmentations initiales (Figure 3.11) ont été effectuées avec un algorithme type Mean-Shift dont les paramètres ont été sélectionnés en évaluant 100 combinaisons différentes avec UOA_{L_2} .

3.4.2 Instanciation de CoSC

Pour cette expérience, nous utilisons la stratégie de sélection du plus ambigu ($P_s = s_{amb}^+$), la stratégie de modification sans scission ($P_m = M_F$), et le critère d’évaluation de classification ($P_e = Q_{clsf}$) ; ce choix constitue un bon compromis entre le temps de calcul et les performances obtenues en classification. Le modèle de classification est une régression linéaire qui a été appris en utilisant l’API de Weka [Hall et al., 2009]. Le jeu d’entraînement est un sous-ensemble des données de référence contenant des segments de végétation et de non-végétation choisis de manière aléatoire. Le pourcentage de végétation dans chaque segment a été employé comme valeur objectif de la régression. Trente caractéristiques sont employées pour décrire les segments ; elles incluent des attributs radiométriques telles que la moyenne ou l’écart-type des bandes, ainsi que des attributs géométriques tels que l’aire et des indices de forme, entre autres. Les seuils de classification utilisés sont $T_\epsilon = 0.9$ et $T_\zeta = 0.1$. Ce choix est fait pour inciter le processus à chercher une classification des segments la plus certaine possible.

3.4.3 Résultats

Nous présentons les résultats de ces expériences en deux temps : d’abord, nous étudions les résultats de CoSC sur six imageries contenant des objets variés ; puis, nous appliquons CoSC sur la totalité de l’image et nous comparons nos résultats avec une méthode orientée pixel, une méthode OBIA classique, et une méthode hybride pixel-objet.



(a) Segmentation initiale du parc public. (b) Segmentation initiale des bâtiments industriels. (c) Segmentation initiale du quartier urbain.

FIGURE 3.11 – Segmentations initiales pour trois imagettes.

3.4.3.1 Étude préliminaire

Les images employées pour cette première analyse sont illustrées dans les figures 3.13a–3.13f. Les images correspondant aux données de référence employées sont illustrées dans les figures 3.12g à 3.12l. Les figures 3.12m à 3.12r montrent les images de probabilité produites par CoSC ; et les figures 3.13g à 3.13l montrent les frontières de la segmentation obtenue à partir des zones plates de l'image de probabilité (pixels contigus ayant la même valeur). Pour ces six imagettes, nous calculons l'erreur quadratique moyenne ($RMSE$) par rapport aux données de référence, ainsi que le taux de segments ambigus ($amb\%$). Comme point de comparaison, nous classifions directement la segmentation initiale avec le modèle utilisé par CoSC.

Ces résultats sont résumés dans le tableau 3.5. Nous observons que CoSC réduit $RMSE$ et $amb\%$ dans la majorité des cas. Ces résultats constituent un premier indice sur l'applicabilité de CoSC en termes de classification. Néanmoins, le résultat le plus intéressant concerne la segmentation. En effet, CoSC a réussi à corriger de manière globale la plupart des erreurs de sur-segmentation. Observons en détail ces résultats. Les figures 3.13g à 3.13i contiennent principalement des arbres et des bâtiments en différentes proportions. On constate que tous les segments correspondant à des objets de végétation sont beaucoup mieux définis. On voit clairement une réduction des zones sur-segmentées, même pour des objets n'étant pas de la végétation tels que les routes par exemple. Dans la figure 3.13j, la segmentation des grandes zones d'arbres est améliorée ; pour les zones moins denses, la segmentation était déjà de bonne qualité et n'a pas été fortement modifiée. Dans la figure 3.13k, les deux grandes zones de forêt sont bien séparées par la rivière. Finalement, dans la figure 3.13l, qui d'après les données de référence ne contient pas de végétation, nous observons quelques légères modifications qui améliorent la segmentation de la rivière.

En somme, cette première analyse nous a permis de vérifier que CoSC réduit effectivement le taux de segments ambigus, tandis que la segmentation s'améliore considérablement de manière globale.

3.4.3.2 Étude approfondie

Il est difficile de se comparer à des méthodes présentées en section 3.1. En effet, celles-ci ne sont pas publiquement disponibles et sont complexes à implémenter ; elles ont des objectifs différents non forcément comparables aux nôtres. Enfin, à notre connaissance, il n'existe pas en télédétection, de jeu de données public et standard permettant de comparer les résultats des

TABLEAU 3.5 – Résultats de la classification pour l'étude préliminaire.

Image Figure	<i>RMSE</i>		<i>amb%</i>	
	OBIA	CoSC	OBIA	CoSC
3.12a	0,29	0,30	0,61	0,49
3.12b	0,25	0,22	0,94	0,78
3.12c	0,32	0,32	0,92	0,92
3.12d	0,40	0,32	0,59	0,40
3.12d	0,18	0,15	0,07	0,04
3.12e	0,10	0,10	0,23	0,23
Moyenne	0,26	0,24	0,56	0,48
Écart-type	0,10	0,09	0,32	0,30

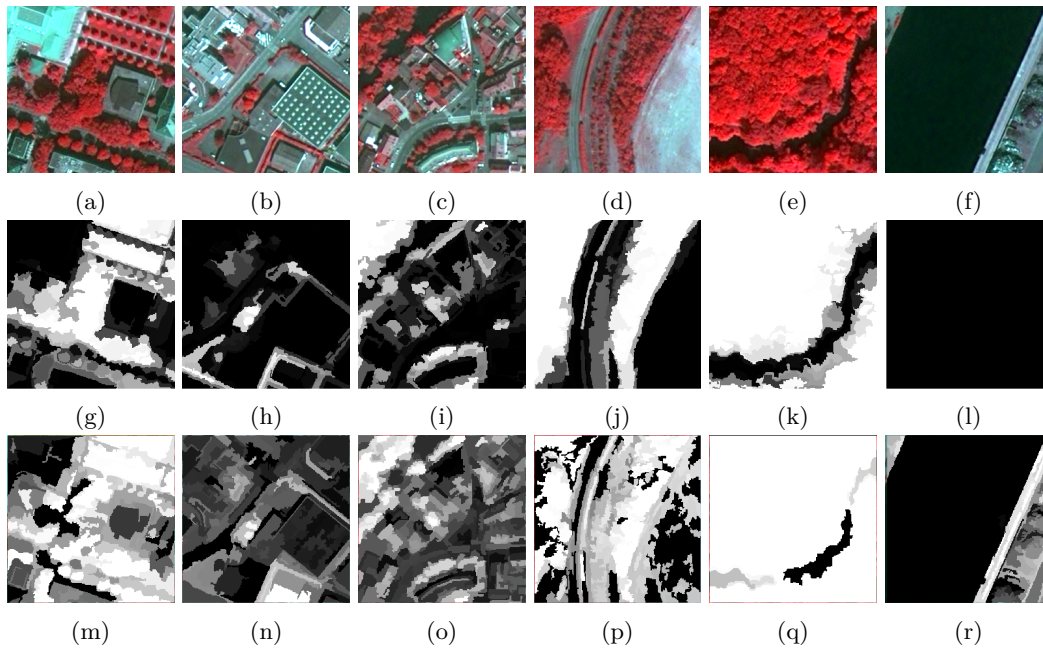


FIGURE 3.12 – Première ligne : les imagettes étudiées. Deuxième ligne : données de référence disponibles. Troisième ligne : images de probabilité obtenues avec CoSC. Le noir correspond à une probabilité de 0.0 et le blanc correspond à une probabilité de 1.0.

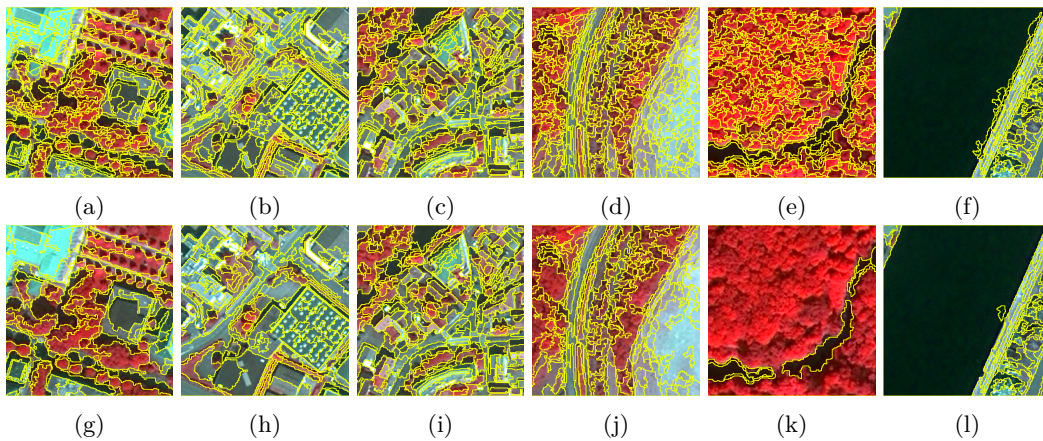


FIGURE 3.13 – Différentes imageries (Haut : segmentations initiales. Bas : segmentations finales.)

TABLEAU 3.6 – Métriques de qualité standard.

Image	Acc	Pr	Re	Fm	κ	Am%	Q_{cs}
NDVI	1,00	1,00	0,99	0,99	0,99	0,89	0,10
OBIA	0,93	0,95	0,95	0,95	0,83	0,52	0,48
CoSC	0,93	0,95	0,94	0,95	0,85	0,46	0,53

différentes contributions. Ainsi, nous avons comparé CoSC à une méthode orientée pixel simple et à une méthode OBIA classique. La méthode basée pixel a consisté à associer l'indice NDVI [Rouse et al., 1974] à chaque pixel. Cet indice est très utilisé en télédétection pour caractériser la végétation. L'image du NDVI a été normalisée entre $[0, 1]$ pour que chaque valeur de pixel puisse être vue comme une probabilité. Pour la méthode OBIA classique, nous avons classé les segmentations initiales avec le même modèle de régression que nous utilisons dans CoSC.

Analyse en termes de classification Pour valider les résultats, nous avons calculé les métriques classiques : Accuracy (Acc), Precision (Pr), Recall (Re), F-measure (Fm), et Cohen's kappa (κ). Un seuil de 0.5 a été appliqué pour déterminer si un segment des données de référence appartient ou non à la classe végétation. Le tableau 3.6 synthétise les résultats de ces métriques pour les trois méthodes évaluées. Les seuils $T_{\epsilon} = 0.9$ et $T_{\bar{\epsilon}} = 0.1$ ont été employés pour définir la zone de rejet. Notons que les segments restant dans la zone de rejet (ambigus) ne sont pas pris en compte dans le calcul des métriques. Nous observons que l'indice NDVI atteint une précision élevée, mais le pourcentage de pixels ambigus est de 89%. L'approche OBIA classique atteint des performances acceptables pour les cinq métriques et réduit le pourcentage de pixels ambigus à 52%. CoSC atteint des résultats similaires à OBIA mais avec un moindre pourcentage de pixels ambigus (46%). Cela suggère que CoSC a réussi à réduire les ambiguïtés.

Comme les résultats et les données de référence sont de nature floue, nous avons effectué l'évaluation des résultats de manière floue également. À partir de la matrice d'erreur floue [Binaghi et al., 1999], nous avons calculé les métriques \widehat{Acc} , \widehat{Pr} , \widehat{Re} et \widehat{Fm} . Ainsi, les calculs prennent en compte l'incertitude présente dans les données de référence pour obtenir une évaluation plus précise des résultats. Le tableau 3.7 synthétise les résultats de l'évaluation floue. Nous observons que l'indice NDVI est globalement moins performant que les deux autres mé-

TABLEAU 3.7 – Métriques de qualité floue.

Image	\widehat{Acc}	\widehat{Pr}	\widehat{Re}	\widehat{Fm}
NDVI	0,77	0,79	0,72	0,75
OBIA	0,81	0,75	0,92	0,83
CoSC	0,82	0,78	0,85	0,82

thodes. Encore une fois, il n’y a pas de différence remarquable entre OBIA et CoSC en termes de classification ; tous les deux atteignent une performance acceptable.

Analyse en terme de segmentation Si nous observons de près les résultats en terme de segmentation, nous constatons de nettes différences. Les figures 3.14, 3.15, et 3.16 illustrent en détail les résultats de NDVI, OBIA, et CoSC, respectivement pour trois imageries. La première ligne montre l’image de probabilités obtenue avec chaque méthode. La deuxième ligne montre les frontières des segments définis par les zones plates de l’image des probabilités, c’est-à-dire par les pixels adjacents ayant une même probabilité. Nous constatons que l’indice NDVI permet de bien différencier les zones de végétation, toutefois l’image des probabilités est très bruitée : cela provient d’une segmentation fortement sur-segmentée. L’approche OBIA classique permet également de discriminer la végétation, mais la segmentation associée est toujours sur-segmentée à des nombreux endroits. CoSC sépare nettement mieux les objets, en donnant une plus forte probabilité aux régions de végétation et une très faible probabilité aux autres régions. Si nous observons la segmentation associée, nous remarquons que celle-ci ne possède presque plus de sur- ou de sous-segmentation. Ceci est très surprenant puisque même les objets qui n’appartiennent pas à la classe végétation sont maintenant correctement segmentés. Ce peut s’expliquer par le fait que le processus prend en compte des critères d’homogénéité radiométriques pour choisir la façon dont les segments sont modifiés pendant le processus collaboratif. Le processus CoSC permet ainsi d’améliorer de manière globale la segmentation de toute l’image.

3.5 Bilan scientifique

Nous avons présenté CoSC, un cadre collaboratif de segmentation et classification. Il permet l’extraction des objets d’une classe thématique donnée grâce à l’échange d’information implicite entre ces deux paradigmes, améliorant ainsi simultanément les résultats. Nous avons réalisé une étude de sensibilité pour étudier l’influence des différents paramètres. Cette étude a permis de montrer empiriquement que le comportement de CoSC est cohérent avec diverses configurations de ces paramètres. Nous avons aussi présenté l’extraction de la végétation sur une image de la ville de Strasbourg de taille considérable. Ces études expérimentales ont permis de montrer l’applicabilité de CoSC. Les résultats montrent notamment que CoSC permet d’obtenir des bons taux de classification en améliorant remarquablement la segmentation, non seulement des objets d’intérêt mais aussi de l’image entière. Cette propriété s’avère très intéressante car plusieurs méthodes de classification reposent sur l’hypothèse qu’une correspondance parfaite entre les segments et les objets dans l’image existe. Or, une telle segmentation idéale est très difficile, voire impossible à obtenir uniquement à partir des propriétés radiométriques de l’image. CoSC permet d’injecter dans le processus de segmentation, l’information sémantique qui est implicitement encodée dans le classifieur. Ainsi, on obtient des segments qui sont plus cohérents par rapport aux objets sous-jacents.

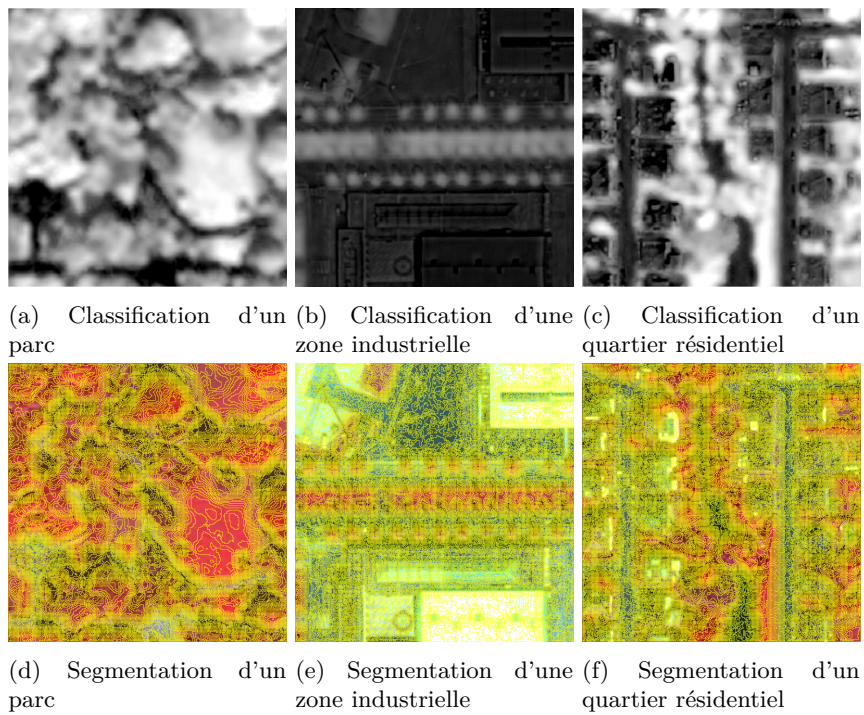


FIGURE 3.14 – Vue détaillée des résultats obtenus avec le NDVI.

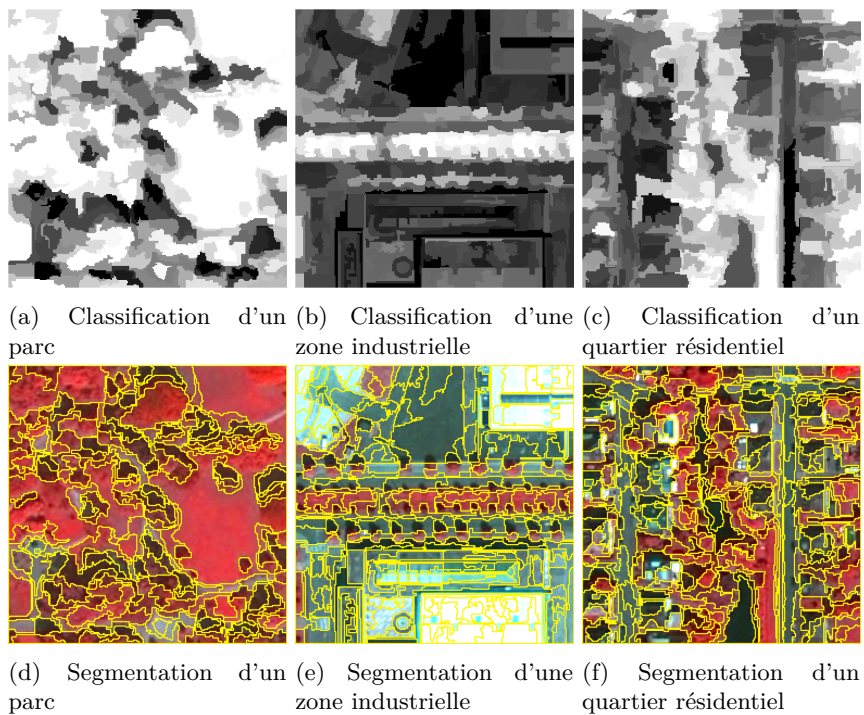


FIGURE 3.15 – Vue détaillée des résultats obtenus avec OBIA classique.

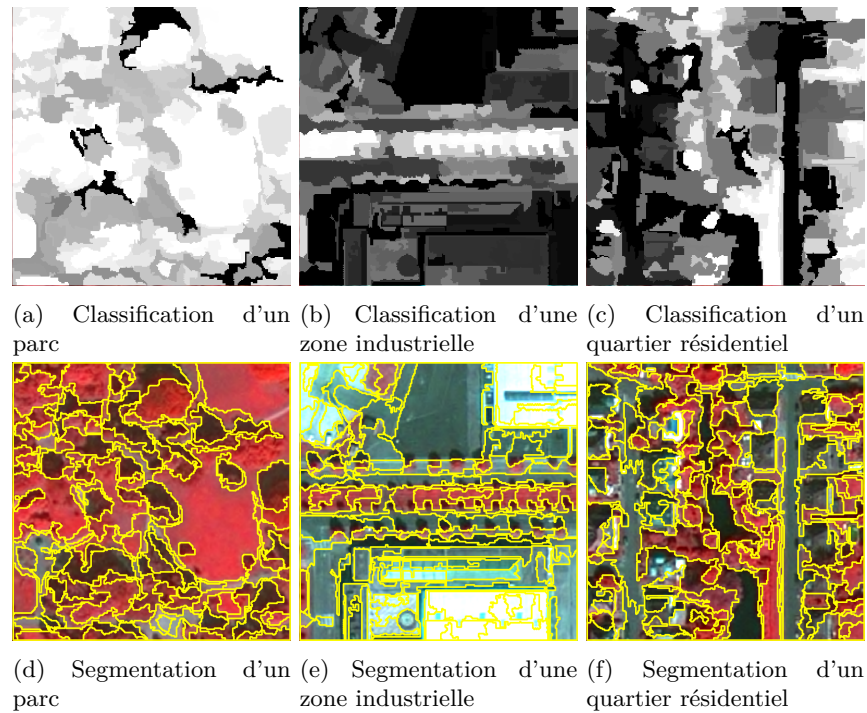
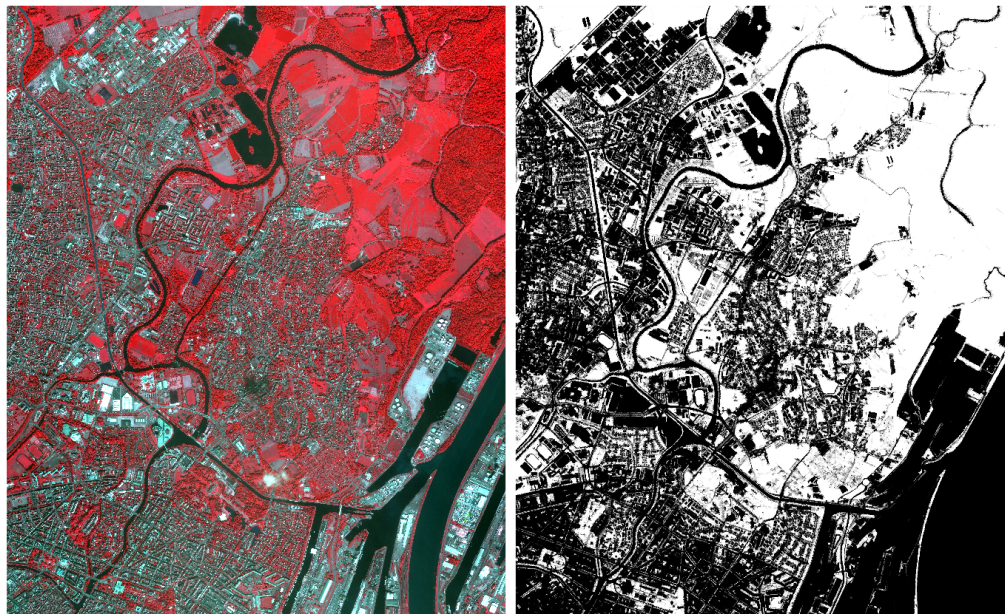


FIGURE 3.16 – Vue détaillée des résultats obtenus par CoSC.

Ces travaux ont donné lieu à deux publications dans des conférences francophones et à une publication dans une conférence internationale :

- [Troya-Galvis and Gancarski, 2015] Rencontres de Jeunes Chercheurs en Intelligence Artificielle (RJCIA2015) ;
- [Troya Galvis et al., 2016b] Rencontres francophones sur l'Extraction et la Gestion des Connaissances (EGC2016) ;
- [Troya Galvis et al., 2016a] (à paraître) conférence internationale International Conference on Pattern Recognition (ICPR 2016).

Une librairie implémentant les différentes variantes de l'approche collaborative a été développée et sera intégrée à la librairie de segmentation JSL développée et maintenue par l'équipe SDC. Un outil indépendant permettant la modification manuelle d'une segmentation en modifiant localement un segment au choix, et permettant de visualiser les résultats de la classification a également été développé (voir annexe D).



(a) Image étudiée.

(b) Données de référence.



(c) Résultat de CoSC.

FIGURE 3.17 – (a) Image Pleiades de la Communauté Urbaine Strasbourgeoise. ©CNES2012. (b) Données de référence. (c) Résultat obtenu par CoSC après reconstitution des imagettes.

Troisième partie

**Vers une approche collaborative
par agrégation d'agents
collaborants**

Approches collaboratives multi-classes

De nombreuses applications telles que la planification urbaine ou le suivi des pratiques agricoles requièrent une carte complète de l'occupation du sol pour effectuer les analyses nécessaires [Xiao and Zhan, 2009, Esch et al., 2013, Fauvel et al., 2012]. Les approches d'extraction mono-classes ne sont alors plus suffisantes, et des approches multi-classes permettant une interprétation totale de l'image sont nécessaires. Dans le domaine de la combinaison de segmentations/classifications, plusieurs méthodes ont vu le jour.

Par exemple, Lizarazo et al. ont proposé un cadre de segmentation floue à partir de classifieurs flous pour chaque classe thématique. Chaque classifieur effectue une classification pixel ; cela donne lieu à ce que les auteurs appellent des segments flous, qui sont ensuite fusionnés en une seule carte de classification dure en utilisant un ensemble défini de règles logiques [Lizarazo and Elsner, 2011].

Kurtz et al. emploient une méthode hiérarchique et multi-résolution non-supervisée pour la segmentation et l'extraction d'objets urbains complexes. Ils exploitent de façon itérative des données multi-résolution, en utilisant des images de plus basse résolution spatiale pour extraire des objets plus grossiers et des images à plus haute résolution spatiale pour extraire des objets de plus en plus complexes [Kurtz et al., 2012].

Mahmoudi et al. ont proposé un système multi-agent pour la classification des images de télédétection, où chaque agent est spécialisé dans l'extraction d'une classe thématique donnée, et un agent particulier permet la communication entre les différents agents afin de faciliter la résolution de conflits [Mahmoudi et al., 2013].

Tarabalka et al. ont proposé une méthode hybride combinant les paradigmes basés pixel et objet. Ils effectuent dans un premier temps une classification par pixel en utilisant un modèle de classification de type SVM. Une segmentation basée sur l'algorithme Expectation-Maximization (EM) est ensuite appliquée à l'image pour améliorer la classification des pixels grâce à une stratégie de vote pour attribuer une classe unique à chaque segment. Un processus de post-traitement permet de supprimer les éventuels artefacts restant [Tarabalka et al., 2009].

Récemment, Hoffman et al. ont formalisé un cadre pour l'analyse d'images basée agents ou Agent Based Image Analysis (ABIA), l'objectif étant de combiner les concepts du paradigme multi-agent aux approches OBIA, afin de faciliter l'automatisation complète du processus, spécialement lors de la sélection de règles de classification ou du paramétrage de la segmentation [Hofmann et al., 2014].

Néanmoins, la tâche consistant à segmenter et classifier une image de télédétection de manière générale est extrêmement complexe et le problème est loin d'être résolu. Toutes ces solutions ont

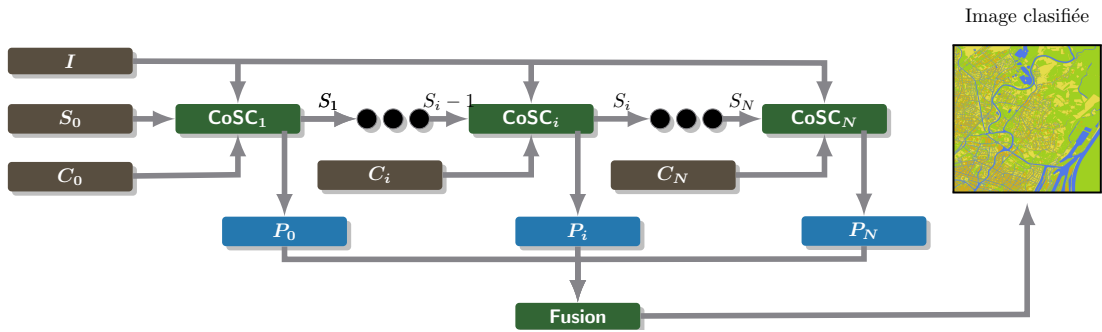


FIGURE 4.1 – Schéma de la méthode CoSCade. Les boîtes marrons correspondent à des données en entrée ; les boîtes bleu clair correspondent à des données en sortie ; les boîtes vertes correspondent à des actions.

des avantages et des inconvénients. Les systèmes multi-agents ont montré de bons résultats, mais sont en général extrêmement complexes à mettre en place. De plus, ils dépendent souvent de modèles basés sur des connaissances très liées à un domaine spécifique qui sont généralement très difficiles à obtenir ou à formaliser. Les approches hybrides pixel-objet injectent efficacement des informations de bas niveau dans le cadre des approches OBIA, mais sont fortement dépendantes de la segmentation initiale qui n'est jamais remise en question par la suite.

Dans cette thèse, nous avons étudié et mis en œuvre deux approches différentes pour la résolution du problème de l'analyse multi-classe.

1. CoSCade : une approche en cascade simple à mettre en œuvre, qui s'inspire des méthodes de *boosting* [Freund et al., 1999] en fouille de données, pour interpréter de manière incrémentale l'image. Elle consiste à appliquer de manière séquentielle plusieurs extracteurs mono-classe. L'information échangée entre les agents est donc la segmentation qui est modifiée et améliorée à chaque étape du processus (Section 4.1).
2. CoCoSC : une approche en parallèle, qui est plus élaborée et consiste en une collaboration d'agents mono-classes travaillant parallèlement chacun sur une classe différente et complémentaire. L'information échangée entre les agents concerne les zones conflictuelles ; ainsi à chaque itération, le processus se focalise sur les zones les plus difficiles à classer, jusqu'à atteindre un niveau de qualité souhaité (Section 4.2).

Les deux approches peuvent être facilement adaptées à de multiples cas d'application et semblent robustes à la segmentation initiale d'après nos expériences.

4.1 Application d'agents CoSC en cascade : CoSCade

Dans le cas où les classes recherchées sont disjointes, une extension évidente de CoSC consiste à exécuter plusieurs processus CoSC de manière indépendante et à fusionner ensuite leurs résultats. Néanmoins, cette solution ne semble pas optimale, puisque les mêmes erreurs de segmentation seraient corrigées à plusieurs reprises par chaque processus CoSC et probablement de manière contradictoire.

Or, comme nous l'avons vu dans le chapitre 3, le processus CoSC permet la labellisation d'objets appartenant à une classe thématique donnée mais surtout améliore considérablement la segmentation initiale de l'image. Cette amélioration de la segmentation ne se limite pas aux zones correspondants aux objets de la classe d'intérêt, mais porte sur l'ensemble de l'image et

notamment les zones où les segments sont ambigus. Ceci nous mène à penser que les améliorations faites sur une classe donnée peuvent s'avérer utiles pour la classification d'autres classes thématiques.

Nous proposons donc d'appliquer les processus CoSC de manière séquentielle. Chaque processus prend en entrée la segmentation finale du processus précédent. Cela permet d'exploiter le fait qu'un processus CoSC fournit une amélioration globale de la segmentation. Ainsi, on peut espérer qu'à chaque étape, la segmentation soit améliorée de manière incrémentale en donnant la priorité aux zones difficiles à classer pour chaque agent. De manière similaire aux approches de *boosting*, chaque processus CoSC profite des améliorations apportées par les précédents et permet d'améliorer davantage l'extraction des différentes classes. Nous évitons ainsi les modifications redondantes et le problème de la fusion *a posteriori* de plusieurs segmentations.

Le processus collaboratif multi-classe que nous proposons est formellement décrit par l'Algorithme 6 et illustré par la figure 4.1. Étant donné un ensemble de N classes thématiques d'intérêt, CoSCade se déroule comme suit.

1. Instancier N agents CoSC. Chaque agent est spécialisé dans l'extraction d'une classe \mathcal{C}_i différente parmi les N classes recherchées. Rappelons que chaque agent CoSC (noté $CoSC_i$) dans la chaîne possède deux sorties : (a) une segmentation modifiée (notée S_i) et (b) une image (notée P_i), donnant pour chaque pixel la probabilité d'appartenance à la classe \mathcal{C}_i .
2. Tant qu'il reste des classes à extraire, exécuter l'agent $CoSC_i$ avec la segmentation S_{i-1} comme segmentation initiale.
À la fin de cette chaîne, nous obtenons une segmentation qui a été modifiée par chacun des processus $CoSC_i$ et N images de probabilité.
3. Fusionner les différentes images de probabilité produites en attribuant à chaque pixel la classe dont la probabilité d'appartenance est la plus importante afin d'obtenir une unique image classifiée. Un seuil de probabilité minimum peut être employé (par exemple 0.5) pour éviter des erreurs de classification lorsqu'aucun classifieur n'a de prédiction avec une forte confiance. Les pixels dont la probabilité est inférieure à ce seuil sont donc labellisés comme « classe inconnue ».

L'ordre d'application des classes est choisi de manière aléatoire ou bien spécifié explicitement par l'utilisateur. Dans la section 4.1.1, nous étudions l'effet de ce choix sur le résultat final.

4.1.1 Expériences

Pour valider notre proposition, nous appliquons CoSCade pour extraire les quatre classes thématiques suivantes :

- Végétation (V) ;
- Routes (R) ;
- Bâtiments (B) ;
- Eau (E).

En vue d'étudier l'influence de différents paramètres sur le résultat, nous avons exécuté CoSCade avec toutes les combinaisons des conditions initiales suivantes :

- toutes les permutations des quatre classes recherchées ($4 \times 3 \times 2 \times 1 = 24$) ;
- deux segmentations initiales différentes ;
- deux stratégies de modification locale.

Algorithme 6 : CoSCade

Data : Ensemble des classes d'intérêt $C = \{\mathcal{C}_i, 1 \leq i \leq N\}$
Data : Image prête à l'analyse I_A
Data : Segmentation initiale S_0
Result : Image labellisée $I_{\mathcal{L}}$
begin
 foreach $\mathcal{C}_i \in C$ **do** instancier un agent $CoSC_i$;
 définir $S_{init} \leftarrow S_0$
 while C n'est pas vide **do**
 sélectionner une classe \mathcal{C}_i
 enlever \mathcal{C}_i de C
 définir $P_i, S_i \leftarrow CoSC_i.run(S_{init})$
 définir $S_{init} \leftarrow S_i$
 foreach *pixel* $p \in I_A$ **do**
 définir $label \leftarrow \arg \max_i (P_i[p])$
 if $P_i[p] < 0.5$ **then**
 définir $label \leftarrow 0$
 définir $I_{\mathcal{L}}[p] \leftarrow label$
 return $I_{\mathcal{L}}$

Nous comparons nos résultats avec ceux de la méthode de classification spectrale-spatiale (SSC) [Tarabalka et al., 2009] décrite dans la section 4. Dans ce qui suit, nous présentons en détail les données en entrée et les différents paramètres étudiés, et nous analysons les résultats en segmentation et en classification.

4.1.1.1 Données et paramètres

Image brute et données de référence. L'image étudiée est celle utilisée pour les expériences de la section 3.3.2. Pour rappel, cette image est donnée dans la figure 4.2a. C'est un extrait de $1\,122 \times 1\,085$ pixels d'une image Pléiades THRS de Strasbourg. Elle a une résolution de 50 cm sur quatre bandes spectrales : rouge, vert, bleu et proche infra-rouge. Elle montre un quartier urbain contenant des maisons individuelles, des routes, des arbres, de l'herbe ainsi qu'une portion de rivière. La figure 4.2b illustre les données de référence fournies par un expert géographe. Les routes et les bâtiments sont issus de bases de données géographiques telles que la BD Topo de l'IGN, et ne s'alignent pas toujours parfaitement avec les objets dans l'image. La végétation et l'eau ont été délimitées manuellement. Les classes étudiées ne sont pas équilibrées ; elles consistent en 11% d'eau, 20% de bâtiments, 22% de routes et 47% de végétation.

Segmentations initiales. Deux segmentations Mean-shift différentes ont été employées en entrée du processus. La première est une segmentation (notée IS_{over}) choisie par un expert géographe. On remarque que celle-ci est clairement sur-segmentée (figure 4.3a). En effet, dans les approches OBIA, il est courant de préférer la sur-segmentation à la sous-segmentation, car cela pose moins de problèmes pour la classification. La deuxième (notée IS_{opt}) a été sélectionnée (Figure 4.3b) en évaluant 100 segmentations différentes avec la métrique UOA_{L_2} (Eq. 1.24). Bien qu'elle soit choisie de manière optimale, cette segmentation possède quelques zones sur-segmentées. En particulier, la majorité des bâtiments sont sur-segmentés, tandis que les segments de route et d'eau sont beaucoup mieux délimités.



(a) Image étudiée.

(b) Données de référence.

FIGURE 4.2 – Image étudiée et données de référence associées.



(a) Segmentation choisie par l'expert.

(b) Segmentation sélectionnée de manière optimale.

FIGURE 4.3 – Segmentations initiales.

TABLEAU 4.1 – Précision, Rappel et leur moyenne harmonique par classe pour les meilleurs résultats obtenus par CoSCade avec les deux segmentations initiales et pour SSC.

Méthode	combinaison	PR	RA	F1	classe
CoSCade	VBER	0.800	0.680	0.735	Routes
		0.727	0.843	0.781	Bâtiments
		0.845	0.958	0.898	Végétation
		0.988	0.782	0.873	Eau
	EBRV	0.757	0.705	0.730	Routes
		0.721	0.830	0.772	Bâtiments
		0.860	0.958	0.906	Végétation
		0.988	0.728	0.838	Eau
SSC		0.035	0.562	0.066	Routes
		0.780	0.871	0.823	Bâtiments
		0.984	0.650	0.783	Végétation
		0.968	0.984	0.976	Eau

Modèles de classification. Pour cette expérience, un arbre VFDT [Hulten et al., 2001] a été entraîné pour chaque classe thématique. Trente caractéristiques radiométriques et géométriques couramment utilisées en télédétection sont calculées pour caractériser chaque segment. Entre dix et quinze exemples et contre-exemples ont été utilisés pour l'entraînement du modèle pour chaque classe. Aucune optimisation des méta-paramètres du modèle n'a été effectuée. Par conséquent, rien ne permet d'affirmer que ces modèles soient optimaux.

Stratégies de modification locale. Nous avons étudié le comportement de CoSCade avec les deux stratégies de modification locale présentées dans la Section 3.3.1.2 et décrites par les algorithmes 4 et 5. Dans une première variante ($P_m = M_F$), seul l'opérateur de fusion a été utilisé. En effet, celui-ci suffit pour corriger la sur-segmentation. Néanmoins, cette stratégie incite les segments à devenir de plus en plus gros et peut déboucher à une sous-segmentation de l'image. Ainsi, dans une deuxième variante $P_m = M_S$, nous avons incorporé un opérateur de scission permettant de corriger les erreurs de sous-segmentation.

4.1.1.2 Discussion

Nous analysons les résultats en termes de segmentation et de classification. En effet, même si l'objectif final est d'obtenir une classification de l'image, il nous semble intéressant d'étudier les améliorations apportées à la segmentation (sans tenir compte de la sémantique liée au domaine), celle-ci pouvant être (ré)utilisée par l'expert ou par d'autres applications. Enfin, une analyse du processus d'amélioration incrémentale du résultat semble indispensable pour mieux le comprendre.

Analyse des classifications produites. Pour chaque image labellisée et pour chaque classe, nous avons calculé la précision, le rappel et le score F1 qui correspond à la moyenne harmonique des deux premières mesures. Nous avons aussi calculé les scores micro- et macro-F1 [Tsoumakas and Vlahavas, 2007] qui permettent d'obtenir un indice de qualité globale de la classification multi-classe.

La figure 4.4 montre les scores micro- et macro-F1 pour les 24 permutations possibles des classes avec la segmentation initiale IS_{over} (en haut) et IS_{opt} (en bas) et les deux stratégies de modification testées. Le tableau 4.2 résume les meilleures et les pires configurations de classes selon les scores macro- et micro-F1.

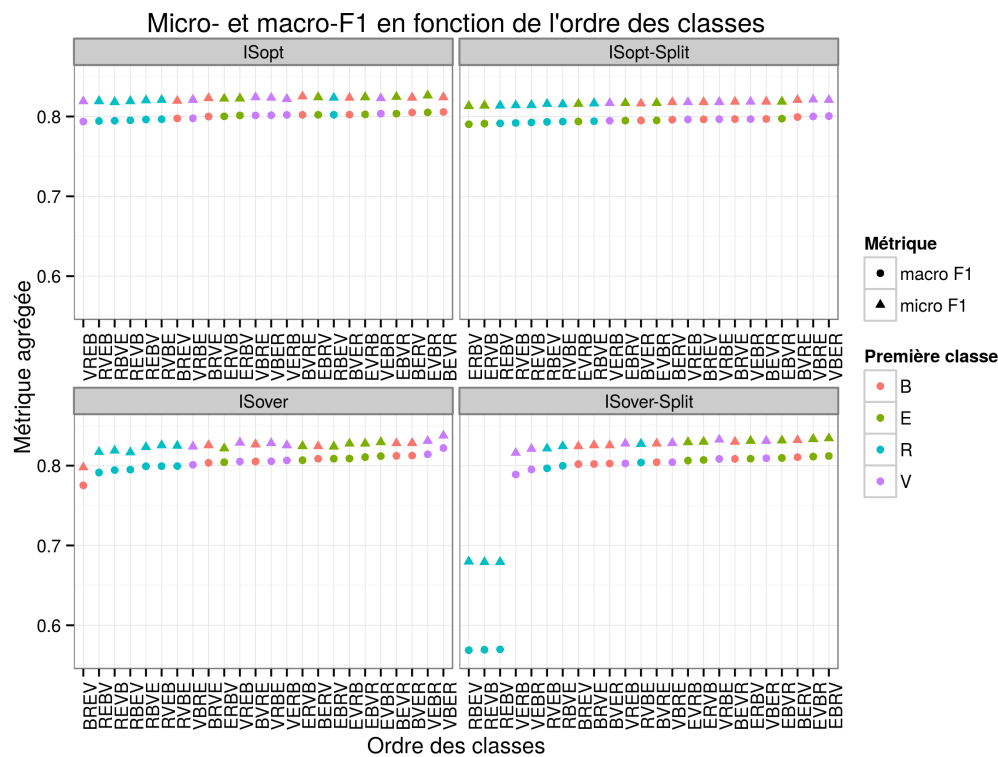


FIGURE 4.4 – Scores macro- and micro-F1 calculés pour IS_{opt} (en haut) et IS_{over} (en bas) avec $P_m = M_F$ (à gauche) et $P_m = M_S$ (à droite). Les labels sur l'axe des x sont triés en fonction des scores macro-F1 et correspondent aux 24 configurations des classes testées.

TABEAU 4.2 – Meilleurs et pires ordres de classe selon les scores macro- et micro-F1.

		Sans scission		Avec scission	
		Meilleur	Pire	Meilleur	Pire
IS_{over}	Ordre	<i>VBER</i>	<i>BREV</i>	<i>EBRV</i>	<i>RBEV</i>
	macro-F1	0.822	0.775	0.812	0.568
	micro-F1	0.837	0.797	0.834	0.679
IS_{opt}	Ordre	<i>BEVR</i>	<i>VREB</i>	<i>VBER</i>	<i>ERBV</i>
	macro-F1	0.822	0.793	0.800	0.790
	micro-F1	0.837	0.818	0.821	0.813

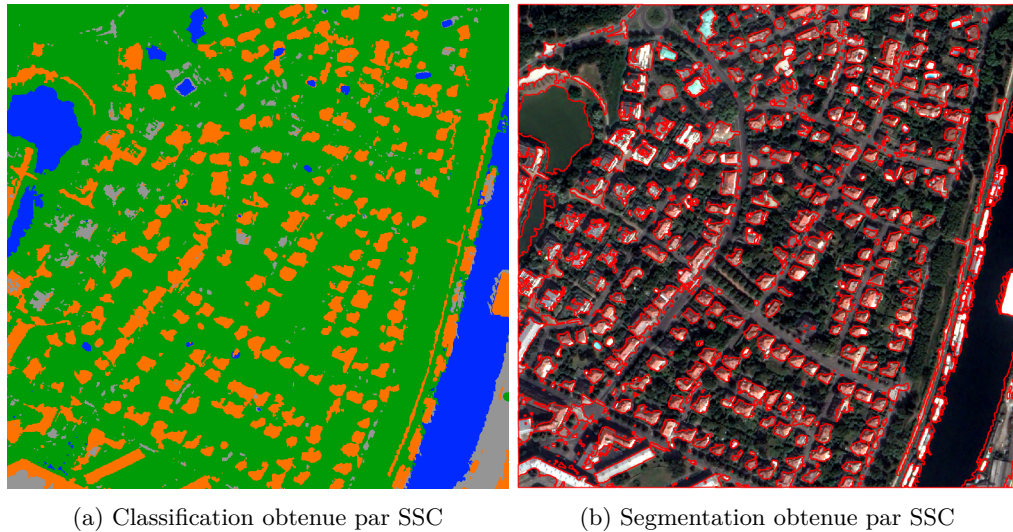


FIGURE 4.5 – Résultat obtenu par la méthode SSC.

Il n’y a pas de tendance qui se dégage concernant l’ordre des classes : si nous regardons les classements des 5 meilleures configurations (les 5 points les plus à droite dans chaque graphe), l’eau apparaît 7 fois comme la première classe de la chaîne ; les bâtiments apparaissent 6 fois ; et la végétation apparaît 6 fois également. *A contrario*, nous remarquons que les performances sont moindres lorsque le processus commence par la classe Routes. Notamment, il existe trois cas (*RBEV*, *REVB*, et *REBV* avec IS_{over} et $P_m = M_S$) où CoSCade atteint des performances anormalement basses par rapport aux autres configurations. Après vérification, nous avons constaté que dans ces cas, toute la zone d’eau a été classifiée comme de la végétation. Ces trois configurations commencent par l’extraction des routes et réalisent l’extraction de l’eau avant la végétation. Nous en déduisons que les modifications faites lors de la première étape ont produit des segments d’eau peu pertinents qui n’ont pas pu être détectés par l’agent d’extraction de la classe Eau. Par ailleurs, CoSCade semble être robuste par rapport à la segmentation initiale et à la stratégie de modification.

Influence du déséquilibre des classes. Dans la figure 4.4 on observe que les scores micro- et macro-F1 sont très corrélés. On remarque aussi que le micro-F1 est généralement supérieur au macro-F1 ; cela s’explique par la nature déséquilibrée des classes. En effet, le micro-F1 a tendance à favoriser le résultat de la classe majoritaire. En revanche, le macro-F1 pénalise plus sévèrement les erreurs de classification sur les classes minoritaires.

Comparaison avec la méthode SSC. Si nous comparons nos résultats avec ceux de la méthode SSC (macro-F1 de 0.662 et micro-F1 de 0.727), nous observons que CoSCade obtient de meilleurs résultats dans la plupart des cas. Une inspection visuelle des résultats permet de constater que SSC (Figure 4.5a) réussit à extraire avec succès l’eau et les bâtiments, mais classifie de manière erronée les routes comme étant de la végétation. Si nous observons les images labellisées par CoSCade dans les figures 4.6a et 4.6c, on remarque que les routes sont plus présentes. De plus, on constate la présence de régions qui sont labellisées comme inconnues. Plusieurs segments de petite taille et de couleur très foncée que nous pensons être de l’ombre sont mal labellisés

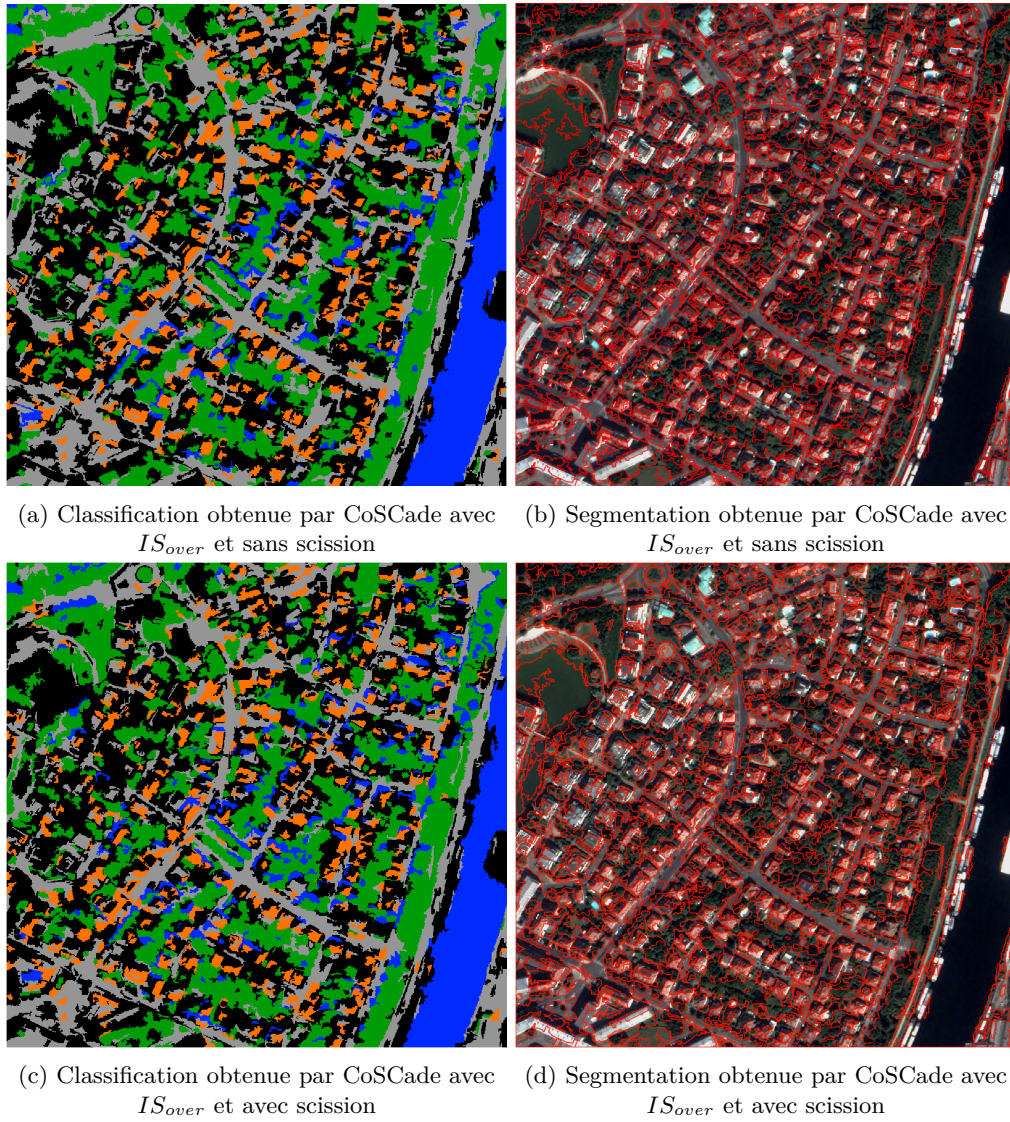


FIGURE 4.6 – Meilleurs résultats obtenus par CoSCade.



FIGURE 4.7 – Métrique $UOAL_2$ avec δ allant de 0 à 1 pour les meilleurs résultats de CoSCade (à gauche) et pour la méthode SSC (à droite).

comme de l'eau. Ceci est un problème assez courant et bien connu dans la communauté de la télédétection.

Nous pouvons corroborer ces observations dans le tableau 4.1, dans lequel nous présentons les scores détaillés par classe pour le meilleur résultat avec IS_{over} et $P_m = M_S$ ($EBRV$); et avec IS_{over} $P_m = M_F$ ($EBRV$), ainsi que pour SSC. Nous constatons qu'en général la végétation est mieux extraite, suivie par l'eau, les bâtiments et finalement les routes. SSC arrive à labelliser les bâtiments avec une précision très élevée mais se trompe très fortement sur les routes. Cela est probablement dû à la segmentation initiale qui n'est jamais remise en cause par cette méthode.

Ces expériences montrent que CoSCade arrive à des bons résultats de classification, même avec des modèles de classification faibles. Il nous semble évident que la performance de CoSCade pourrait être améliorée en optimisant le choix et l'entraînement des modèles de classification.

Analyse des segmentations produites. Nous avons décidé d'évaluer les résultats avec la métrique $UOAL_2$ (voir section 4.7) afin d'avoir une mesure quantitative et objective de la qualité. Rappelons que $UOAL_2$ est définie par $\sqrt{\Psi^2 + \Theta^2}$ avec Ψ (resp. Θ) le taux de sous- (resp. sur-) segmentation détectée en fonction d'un critère d'homogénéité H et d'un seuil δ sur celui-ci. Plus le score de $UOAL_2$ est faible, meilleure est la qualité de la segmentation.

La figure 4.7 montre les valeurs de $UOAL_2$ avec l'entropie comme indice d'homogénéité pour δ allant de 0 à 1, avec un pas de 0.01. Les segmentations suivantes y sont évaluées :

- CoSCade avec IS_{over} , $VBER$ et $P_m = M_F$;
- CoSCade avec IS_{over} , $EBRV$ et $P_m = M_S$;
- SSC.

Nous observons que la signature des deux résultats de CoSCade est très similaire. Elles atteignent une valeur minimale de 0.650 et 0.669 avec $\delta = 0.83$ et $\delta = 0.82$, respectivement. En revanche, la signature de SSC possède des valeurs qui sont plus importantes en général. Sa valeur minimale est de 0.819 avec $\delta = 0.88$. Une façon de quantifier ces résultats est de calculer l'aire sous cette courbe. Les deux résultats de CoSCade ont une aire d'environ 0.92 alors que l'aire de SSC est de 0.95. Cela suggère une amélioration de la segmentation par CoSCade. Vérifions ces résultats par une inspection visuelle. Les figures 4.6b et 4.6d montrent les deux meilleurs résultats de CoSCade et la figure 4.5b montre les résultats de SSC. On observe que SSC réussit à isoler correctement la majorité des bâtiments, mais que contrairement à CoSCade, le reste des objets est en général sous-segmenté, ce qui explique en partie les faibles performances de classification, puisque cette segmentation n'est jamais remise en cause par la méthode SSC.

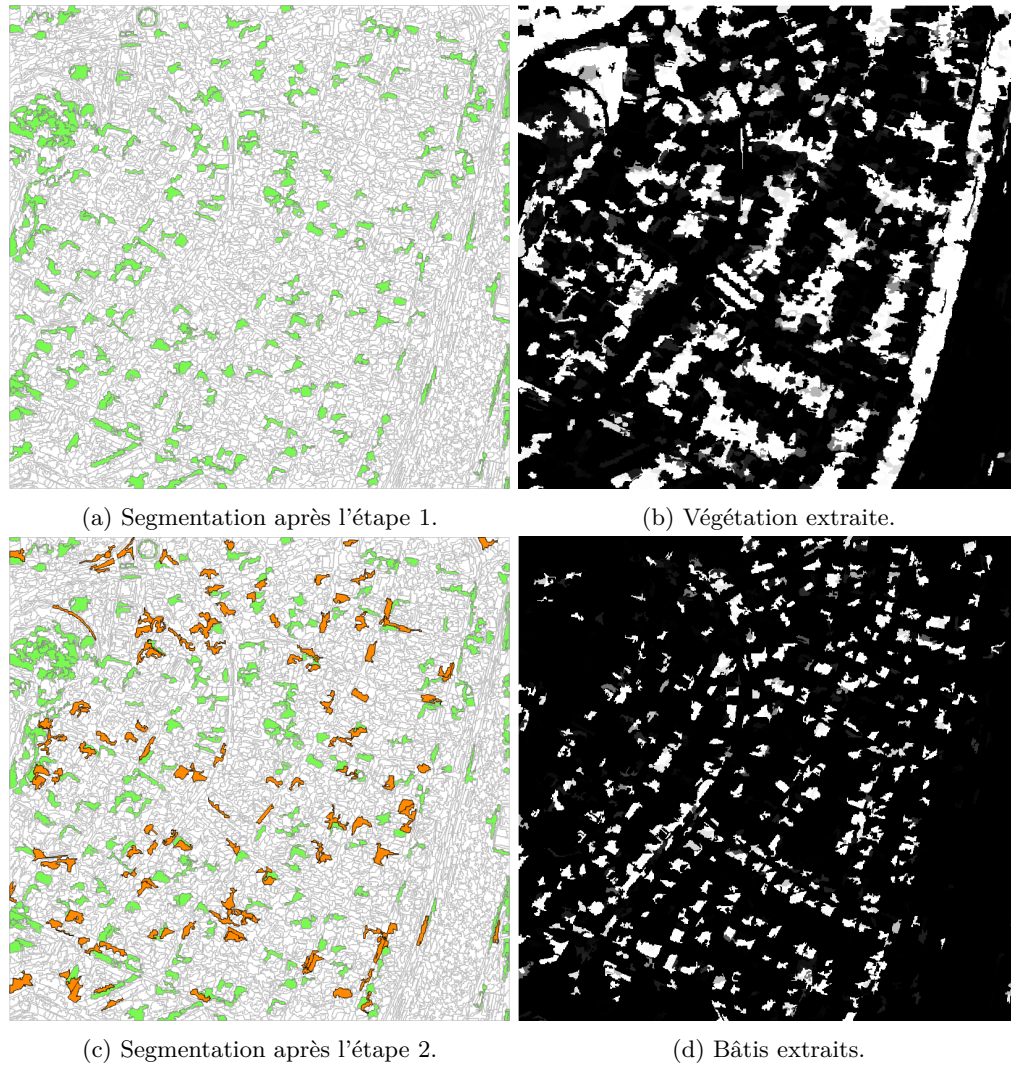


FIGURE 4.8 – Résultat étape par étape de CoSCade pour la chaîne $V \rightarrow B \rightarrow E \rightarrow R$. À gauche : l'évolution de la segmentation après chaque exécution de CoSC; les zones coloriées correspondent aux segments qui ont été modifiés lors de chaque étape. À droite : l'image de probabilité résultante à chaque étape.

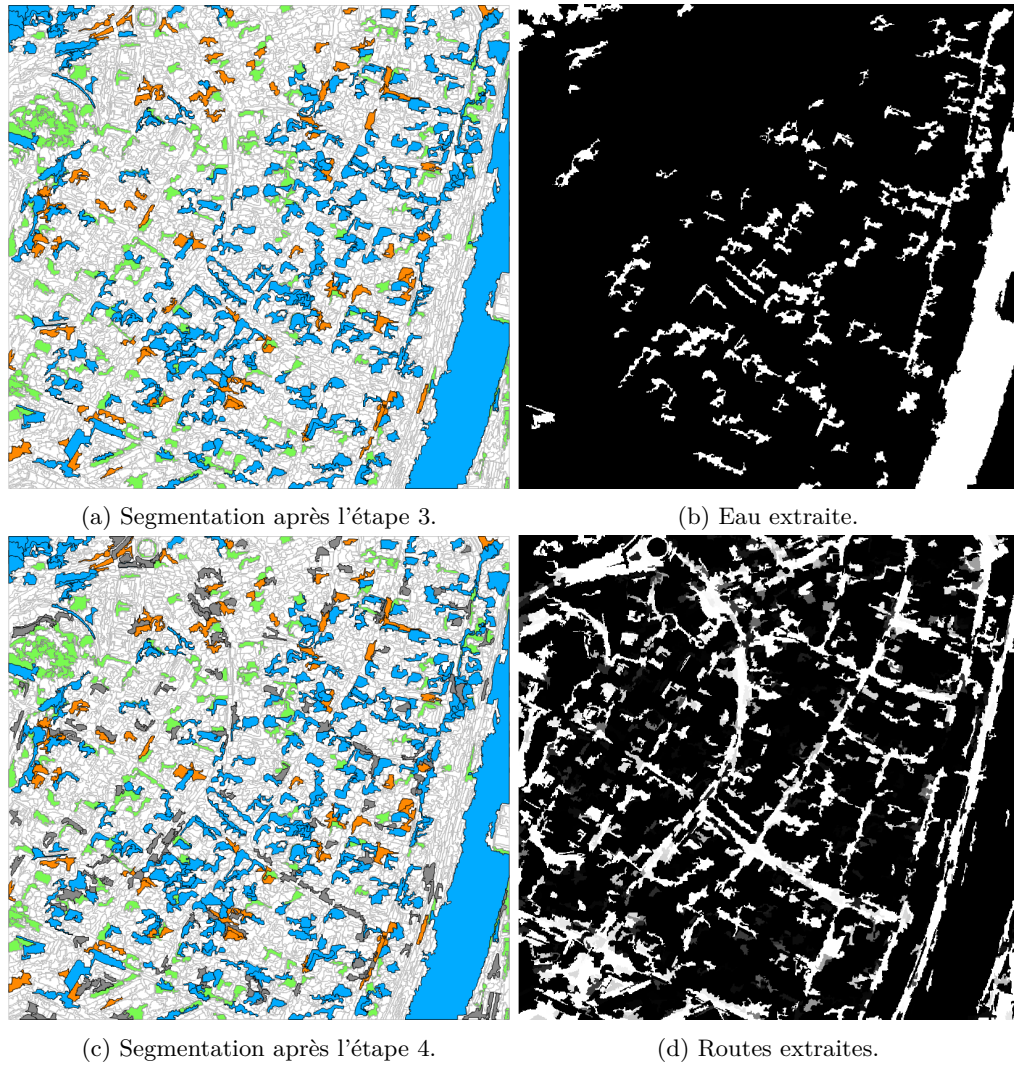


FIGURE 4.9 – (suite) Résultat étape par étape de CoSCade pour la chaîne $V \rightarrow B \rightarrow E \rightarrow R$. À gauche : l'évolution de la segmentation après chaque exécution de CoSC ; les zones colorées correspondent aux segments qui ont été modifiés lors de chaque étape. À droite : l'image de probabilité résultante à chaque étape.

Analyse du processus de collaboration. La figure 4.9 illustre les résultats intermédiaires de CoSCade pour la chaîne *VBER* avec IS_{over} et $P_m = M_F$. La colonne de gauche montre la segmentation courante après l'exécution de chaque agent CoSC. Les segments qui ont été modifiés par CoSC sont représentés avec une couleur différente en fonction de l'étape où ils ont été modifiés (étape 1 en vert, étape 2 en orange, étape 3 en bleu et étape 4 en gris). La colonne de droite montre l'image de probabilité résultante associée à l'étape. La segmentation initiale (IS_{over}) contient 5 357 segments. Après la première étape, la segmentation ne possède plus que 5 090 segments ; plusieurs segments petits et compacts sont apparus. Après la deuxième étape, il reste 4 961 segments ; des segments de petite taille sont apparus : on remarque que tous ces segments font partie d'un bâtiment ou en sont spatialement proches. Après la troisième étape, la segmentation présente 4 294 segments. On constate que tous les segments correspondant à la rivière (à droite de la segmentation) ont été fusionnés en un seul. Après la dernière étape, la segmentation a été réduite à 4 177 segments et plusieurs segments allongés sont apparus. Cela confirme notre hypothèse de départ que chaque agent CoSC est capable d'injecter des connaissances implicitement encodées dans le modèle de classification dans le processus de segmentation. En effet, à chaque étape du processus, les segments les plus à même d'améliorer la classification de la classe en question sont effectivement modifiés et améliorés. Ainsi, chaque chaînon du processus se focalise sur des régions différentes et complémentaires de l'image. Par conséquent la segmentation est améliorée de manière globale.

4.1.2 Bilan scientifique

Nous avons proposé CoSCade, une approche collaborative de segmentation et de classification faisant intervenir plusieurs agents CoSC en cascade, pour interpréter de manière automatique des images de télédétection. CoSCade permet de labelliser une image de manière incrémentale, en traitant une classe thématique à la fois. Nos expériences montrent que CoSCade atteint de bonnes performances en classification, même lorsque les modèles de classification employés ne sont pas optimaux. Nous avons montré que la segmentation initiale est améliorée remarquablement, ce qui nous amène à penser que la segmentation finale pourrait être utilisée comme entrée à d'autres méthodes OBIA.

Ces travaux ont donné lieu à l'implémentation de l'approche collaborative en cascade dans un module d'une librairie de segmentation et classification d'images de télédétection développée durant cette thèse et qui sera intégrée à la librairie de segmentation JSL développée et maintenue par l'équipe SDC. Un article soumis à une revue internationale est actuellement en cours d'évaluation.

4.2 Extraction multi-classe par l'agrégation collaborative d'agents collaboratifs

Nous proposons, dans cette section, un processus collaboratif d'agents CoSC en parallèle, qui prend en compte des segmentations indépendantes et se sert d'échanges entre plusieurs agents pour la remise en cause des résultats individuels. Nous appelons ce processus CoCoSC. Ce schéma de classification s'inspire des travaux réalisés dans l'équipe SDC sur la classification collaborative multi-stratégie, dont il reprend les grands principes [Forestier et al., 2010].

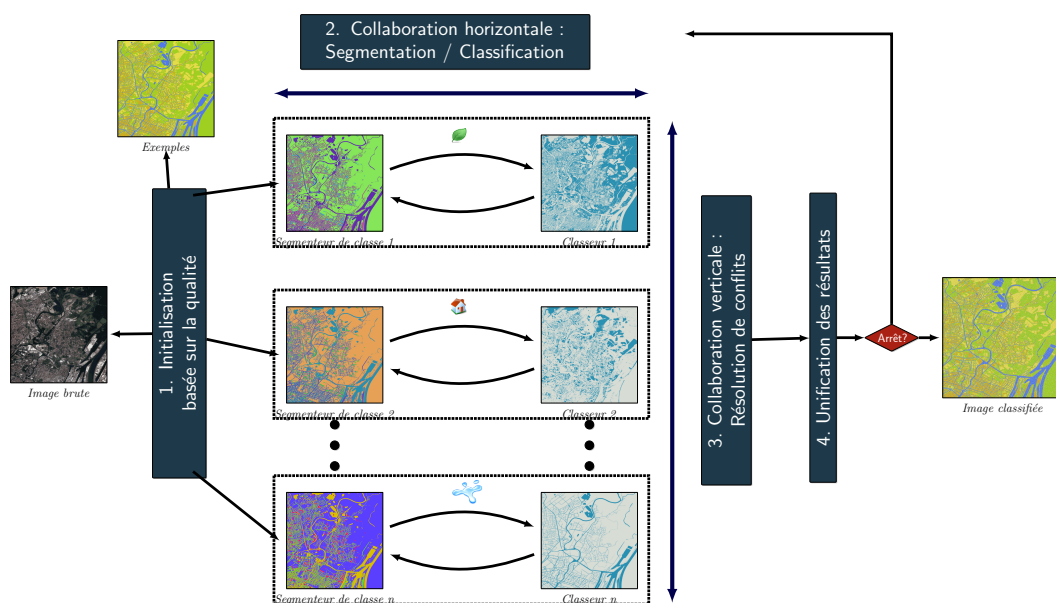


FIGURE 4.10 – Schéma collaboratif complet.

4.2.1 Collaborative CoSC : CoCoSC

Ce processus est basé sur l'utilisation d'agents CoSC sur deux niveaux de collaboration. D'une part, une collaboration dite verticale (inter-agent) mettant en œuvre l'échange d'information entre les agents et, d'autre part, une collaboration dite horizontale (intra-agent) propre au fonctionnement des agents CoSC. Comme le montre la figure 4.10, le processus CoCoSC prend en entrée une image de télédétection corrigée prête à l'analyse et se divise principalement en quatre étapes.

1. Instanciation des N agents CoSC : chaque agent se spécialise dans une classe C_i différente et complémentaire. Les paramètres de chaque agent peuvent être choisis de manière adaptée à chaque classe recherchée.
2. Collaboration horizontale : les agents CoSC s'exécutent de manière indépendante afin d'améliorer leur propre segmentation et classification.
3. Collaboration verticale : les agents CoSC collaborent entre eux afin d'améliorer conjointement leurs résultats. Elle se divise en trois sous-étapes :
 - (a) l'identification des conflits ;
 - (b) l'échange d'informations ;
 - (c) la résolution des conflits.
4. Unification des résultats : une solution intermédiaire est créée à partir des résultats de chaque agent CoSC et l'itération suivante est préparée en relaxant quelques contraintes pour chaque agent CoSC.

Les étapes 2 et 3 sont répétées jusqu'à ce qu'un critère de qualité minimum soit atteint.

En sortie, le processus produit une image labellisée avec $N + 1$ étiquettes, soit les N classes recherchées, plus une étiquette correspondant à la « classe inconnue » lorsqu'aucun des agents n'arrive à classer avec certitude un objet donné. Ci-après, nous présentons en détail les différentes étapes du processus.

4.2.1.1 Initialisation basée sur la qualité

L'étape d'initialisation permet d'obtenir N segmentations adaptées aux objets de chacune des classes recherchées. Cela permet d'initier le processus collaboratif avec des segments de taille adéquate et ayant plus de chances de correspondre aux objets géographiques. Le choix de segmentations adaptées se fait au moyen d'exemples fournis par un expert et de la métrique UOA_{L_2} (Eq. 1.24, Section 1.3.2). Pour chaque classe, les étapes suivantes sont appliquées :

1. pour tous les exemples, calculer un indice d'homogénéité H (dans les expériences que nous avons mené, nous avons utilisé l'entropie) ;
2. calculer la moyenne harmonique m de H ;
3. sélectionner, parmi un grand nombre de segmentations variées, la segmentation S_{init} qui minimise la métrique UOA_{L_2} avec $\delta = m$.

Cette procédure nous permet d'obtenir des segmentations dont l'échelle est proche de celle des objets de chaque classe. En effet, comme nous l'avons vu dans la première partie de ce document, le paramètre δ permet d'influencer le comportement de UOA_{L_2} pour favoriser des segmentations à différentes échelles. En définissant δ comme le score d'homogénéité d'exemples de segments donnés en exemple par l'expert, nous pouvons espérer que la segmentation choisie avec un tel paramètre sera à une échelle adaptée aux les objets géographiques de la classe considérée.

4.2.1.2 Collaboration horizontale

L'étape de collaboration horizontale est l'application des agents CoSC de manière indépendante. La seule différence avec la méthode CoSC présentée dans le chapitre 3 est la définition d'une nouvelle stratégie de sélection P_s par l'ajout d'une liste prioritaire de segments à modifier. Cette liste sera mise à jour lors de l'échange d'information entre agents, pendant de l'étape de collaboration verticale.

4.2.1.3 Collaboration verticale

L'étape de collaboration verticale consiste en un échange d'informations entre les différents agents CoSC afin de guider les processus mono-classes et de réduire le nombre de segments conflictuels. La figure 4.11 illustre le principe de la collaboration verticale.

La première étape consiste à identifier les éventuels segments conflictuels qui ont pu apparaître suite à la phase de collaboration horizontale. Pour ce faire, une carte de conflits au niveau des pixels est créée. Un pixel p est dit en conflit lorsqu'il appartient à deux classes thématiques \mathcal{C}_i et \mathcal{C}_j , c'est-à-dire, lorsqu'il existe i et j tels que $P_i(p) > T_{\epsilon_i}$ et $P_j(p) > T_{\epsilon_j}$, où $i \neq j$, P_i (resp. P_j) étant l'image de probabilités associée à l'agent $CoSC_i$ (resp. $CoSC_j$). La résolution d'un tel conflit s'effectue en deux temps.

1. Échange d'informations : dans cette étape, chaque segment contenant au moins un pixel conflictuel est ajouté à la liste prioritaire (introduite dans la section précédente) de chaque agent CoSC qui intervient dans le conflit. Ainsi, lors de la étape suivante les segments conflictuels seront modifiés en priorité par les agents CoSC. Cela permet de se focaliser davantage sur les zones problématiques et corriger les segments erronés.
2. Attribution d'un label : dans cette étape, on attribue à un pixel en conflit p le label k correspondant à la classe dont la probabilité $P_k(p)$ est maximale. Il est à noter que cette action n'affecte que très légèrement la classification finale. En effet, le nombre de pixels conflictuels dans un même segment est très réduit et ils se trouvent aux frontières des segments. L'attribution du label par un tel procédé permet de prendre une décision

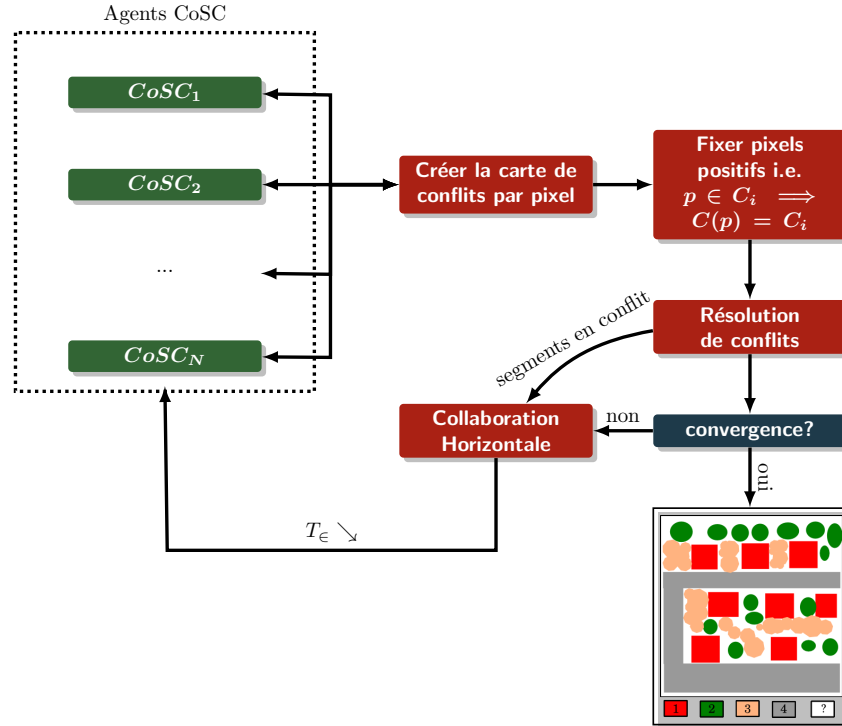


FIGURE 4.11 – Schéma du processus de collaboration verticale entre agents CoSC.

rapide concernant des pixels ayant une forte probabilité de rester en conflit, même après la prochaine étape de collaboration horizontale.

Enfin, tous les pixels qui ont été positivement classés par un seul agent $CoSC_i$ sont labellisés de manière définitive avec le label C_i .

4.2.1.4 Itération

Nous avons défini une stratégie pour itérer les phases de collaboration en améliorant la qualité globale du résultat quelle que soit la qualité individuelle de chaque classifieur. Comme il a été décrit précédemment, un agent CoSC requiert la définition de deux seuils permettant de déterminer si un segment est ambigu ou non. Or, nos expériences ont montré que les valeurs idéales de ces seuils peuvent dépendre du modèle de classification utilisé. Pour supprimer la contrainte de la définition des seuils, nous mettons en place la stratégie définie par l'algorithme 7. L'idée est de relâcher l'exigence de confiance demandée à chaque agent CoSC au fur et à mesure, en fonction de la qualité des modèles de classification. Initialement, le seuil de non-appartenance est fixé à $T_{\notin} = 0.1$ et le seuil d'appartenance est fixé à $T_{\in} = 0.99$ pour tous les classifieurs. Ce dernier est adapté à chaque itération en fonction du nombre de segments ambigus. En effet, un classifieur spécialisé dans la classe C_i a tendance à assigner aux objets de cette classe une probabilité plus élevée que celle donnée par des classifieurs spécialisés dans une classe $C_j, i \neq j$, même si cette probabilité est en réalité faible.

Dans l'étape de mise à jour des seuils nous avons défini deux termes :

1. *step* est commun à tous les agents CoSC, et calculé en fonction de Q_{gl} (Eq. 4.2) ; plus le nombre de pixels décidés est élevé, moins la mise à jour des seuils sera importante. Le fait

de réduire le seuil ainsi permet de donner plus de chances aux classifieurs faibles d'étiqueter les pixels adéquats.

2. A_i est calculé pour chaque agent. Il correspond au taux de segments non-décidés par rapport au reste des agents :

$$\begin{aligned} A_i &= u_i/U \\ U &= \sum_i^N u_i \end{aligned} \quad (4.1)$$

où u_i est le pourcentage de segments non-décidés de l'agent CoSC i . Ainsi, à chaque itération, de plus en plus de segments sont acceptés comme décidés et finalement toute (ou presque) l'image est labellisée.

4.2.1.5 Critère d'arrêt

Nous définissons un critère permettant d'estimer la qualité globale du résultat :

$$Q_{gl} = \frac{D - (K + A)}{Q} \quad (4.2)$$

où D est le nombre de pixels décidés, K est le nombre de pixels en conflit, A est le nombre de pixels ambigus et Q est le nombre total de pixels dans l'image.

Par définition, $D + K + A = Q$. Par conséquent, dans le meilleur des cas, on a $D = Q$ et donc $Q_{gl} = 1$. Dans le pire des cas, on a $D = 0$ et $A + K = Q$, donc $Q_{gl} = -1$.

Le processus s'achève lorsque plus de 95% de l'image a été labellisée ou bien lorsque tous les seuils T_{ϵ_i} sont inférieurs 0.1. On évite ainsi de labelliser des segments présentant une probabilité trop faible.

Algorithme 7 : CoCoSC

Data : Ensemble d'agents CoSC $CS = \{CS_i, 1 \leq i \leq N\}$

Data : Image prête à l'analyse I_A

Result : Image labellisée $I_{\mathcal{L}}$

begin

 minThreshold = 0.1

 initialThreshold = 0.99

 step = 1

while $Q_{gl} < 0.95$ or $\forall 1 \leq i \leq N; T_{\epsilon}^i \leq \text{minThreshold}$ **do**

 horizontalCollaboration()

 verticalCollaboration()

 step = 1 - Q_{gl}

foreach CoSC agent CS_i in CS **do** $T_{\epsilon}^i = T_{\epsilon}^i - (A_i \times \text{step});$

4.2.2 Expérimentation

Bien que la définition théorique complète du cadre CoCoSC ne soit pas entièrement finalisée, nous avons néanmoins pu mener des expériences afin d'en valider les grands principes, de vérifier sa faisabilité et d'estimer son potentiel. Pour cela, nous avons appliqué CoCoSC au même extrait d'image Pleiades utilisé dans la section précédente (Figure 4.2) pour l'extraction de l'eau, des

TABLEAU 4.3 – Différents paramètres utilisés pour l’expérience sur CoCoSC (voir Section 3.3 pour les détails).

paramètre	Eau	Bâtiments	Routes	Végétation
P_m	M_S	M_S	M_S	M_S
P_s	s_{seg}	s_{seg}	s_{seg}	s_{amb+}
P_e	Q_{seg}	Q_{mix}	Q_{clsf}	Q_{mix}
P_c	MLP	MLP	VFDT	MLP
H et δ	Entropie = 0.749	Entropie = 0.766	Entropie = 0.763	Entropie = 0.751

bâtiments, de la végétation et des routes. Nous décrivons ci-après l’instanciation des agents CoSC employés pour cette expérience ainsi que les premiers résultats obtenus.

4.2.2.1 Instanciation

La méthode CoSC est générique. De ce fait, il est possible de choisir une implémentation différente et adaptée à chaque classe recherchée. Pour mettre en évidence cette flexibilité, nous avons choisi aléatoirement des configurations différentes pour chaque classe (Tableau 4.3). La stratégie de modification avec scission a été utilisée pour tous les agents CoSC, ainsi que l’entropie comme indice d’homogénéité pour évaluer la segmentation. Les seuils sur l’entropie ont été définis en suivant la méthodologie décrite dans la section 4.2.1.1.

4.2.2.2 Résultats préliminaires

La figure 4.12 montre l’évolution de la labellisation de l’image après chaque itération. Nous observons que le processus a réussi à labelliser plus de 95% de l’image en six itérations. Lors de la première itération (figure 4.12a), la plupart de l’eau ainsi qu’une part importante des routes ont été classifiées. Nous remarquons également l’apparition de quelques zones de végétation. Lors de la deuxième itération (figure 4.12b), la majorité de la végétation apparaît, ainsi que plusieurs bâtiments. Ce n’est qu’à la troisième itération (figure 4.12c) que le processus réussit à labelliser la plupart des bâtiments. Lors de la quatrième itération (figure 4.12d), des zones de petite taille restantes sont décidées. Nous constatons que ces zones sont effectivement à la frontière des objets géographiques, d’où la difficulté de les classer rapidement. Au final, il reste peu de zones sans label, le résultat final atteint une valeur de macro-F1 de 0.803 et de micro-F1 de 0.820 par rapport aux données de référence. Ces scores sont comparables à ceux obtenus avec CoSCade mais avec la grande différence qu’avec CoCoSC, les zones non décidées sont visiblement moins nombreuses (4.5% des pixels) qu’avec CoSCade (36.7% des pixels).

Ces résultats étant très encourageants, nous avons décidé d’appliquer CoCoSC à la totalité de l’image Pléiades ($9\,211 \times 11\,275$ pixels, 50 cm de résolution et 4 bandes : R, V, B, PIR) de l’agglomération strasbourgeoise dont nous disposons (Figure 4.14). Les données de référence disponibles sont illustrées dans la figure 4.15. Elles contiennent les 4 classes (Routes en gris, Eau en bleu, Végétation en vert et Bâtiments en orange), ainsi que quelques zones (noir) qui n’appartiennent à aucune de ces classes. Nous avons appliqué CoCoSC à cette image en programmation par streaming avec des tuiles de 2048×2048 pixels. Les résultats obtenus sont fournis dans la figure 4.16. Aucun traitement particulier a été effectué aux frontières de chaque tuile ; nous constatons tout de même que l’effet du traitement par tuiles n’est pas directement observable : cela peut s’expliquer par le fait que la résolution des conflits est réalisée au niveau des pixels et non des segments, ce qui réduit notablement l’effet des bords des tuiles sur l’image résultante.

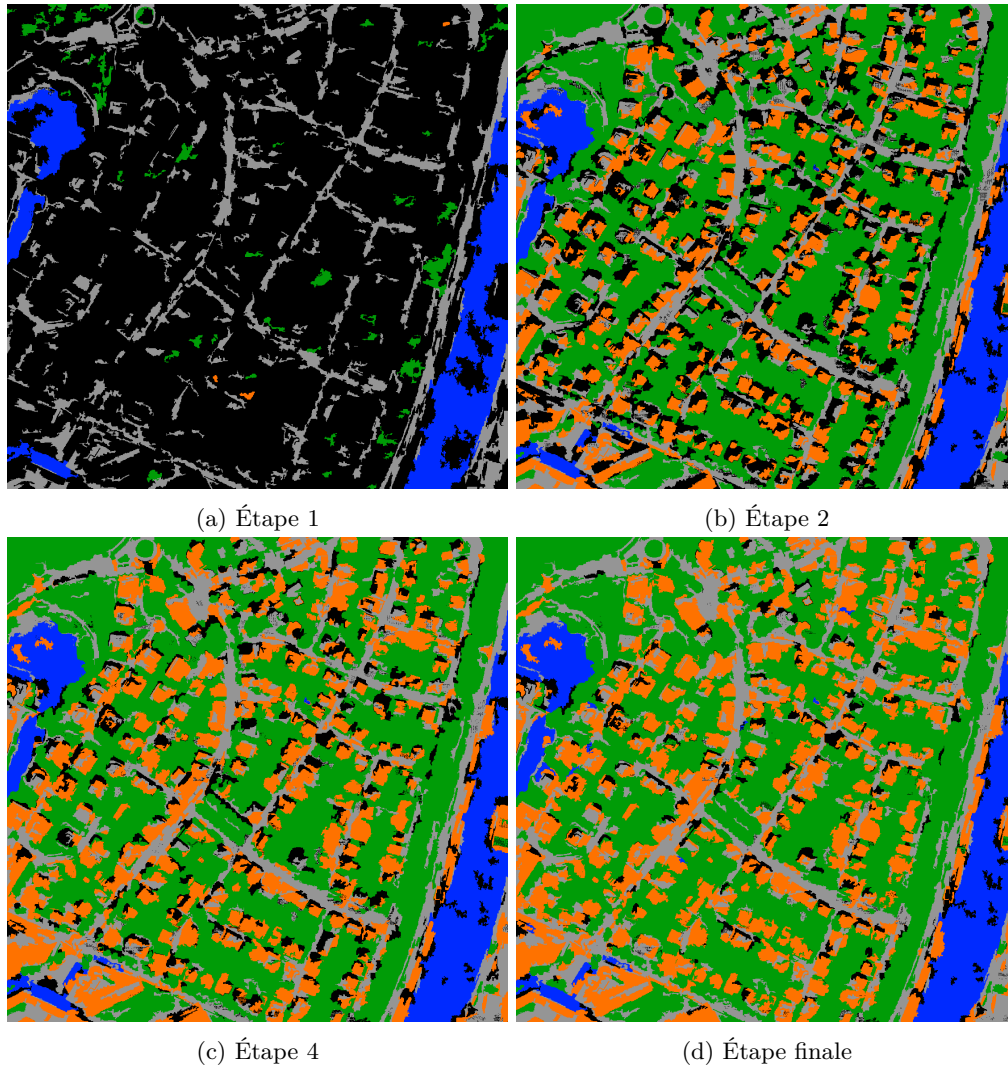
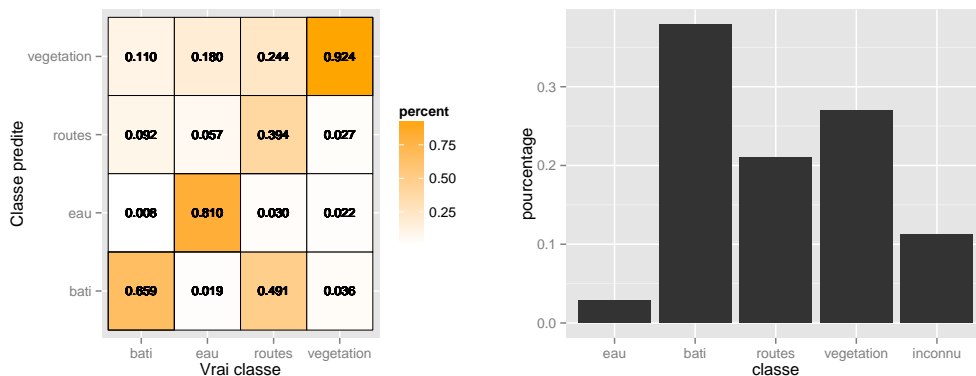


FIGURE 4.12 – Évolution des résultats de CoCoSC.



(a) Matrice de confusion des résultats de CoCoSC. (b) Répartition des pixels non-décidés.

FIGURE 4.13 – Analyse des résultats de CoCoSC.

L'évaluation par rapport aux données de référence donne une mesure macro-F1 égale à 0.693 et micro-F1 égale à 0.834. La figure 4.13a présente la matrice de confusion associée à cette classification. Nous observons que la végétation a été extraite correctement à 92%, l'eau a été extraite correctement à 81%, les bâtiments ont été extraits correctement à 66% et finalement les routes ont été extraites correctement à 39%. Ces dernières ont été confondues en partie avec de la végétation, mais surtout avec les bâtiments. En effet, on trouve dans l'image des bâtiments ayant des propriétés radiométriques très similaires à celle des routes. Ainsi, un segment de route de taille réduite et de forme similaire à celle des bâtiments a des fortes chances d'être classifié comme bâtiment au lieu de route. Une solution à ce problème serait de diviser la classe des bâtiments en plusieurs sous-classes selon la couleur des toits ou bien optimiser l'entraînement du modèle des routes.

Par ailleurs, l'image résultante contient 15.3% de pixels de classe inconnue. La figure 4.13b montre la répartition des pixels non-décidés à l'issue du processus parmi les différentes classes. Nous remarquons qu'il s'agit majoritairement de bâtiments (37.9%). Ainsi, l'amélioration du modèle de classement pour les bâtiments en créant des sous-classes plus spécifiques semblerait bénéfique pour améliorer davantage les résultats.

4.2.3 Bilan scientifique

Nous avons proposé un processus d'extraction multi-classe collaboratif qui utilise des segmentations différentes et adaptées pour chaque classe d'intérêt, et alterne itérativement une étape dite de collaboration horizontale, où plusieurs agents travaillent en parallèle pour extraire au mieux une seule classe thématique, et une étape dite de collaboration verticale, mettant en œuvre une stratégie de résolution de conflits ainsi qu'un échange d'information entre les différents agents collaborant pour mieux guider le processus global. Ce processus atteint des performances similaires à celles de la méthode CoSCade proposée dans la section précédente mais avec un pourcentage de pixels non-décidés qui est notablement inférieur à celui obtenue par cette dernière. Les premiers résultats sont donc très encourageants et prometteurs étant donnée qu'aucune optimisation des paramètres n'a été effectuée.

Ces travaux ont donné lieu à l'implémentation de l'approche entièrement collaborative dans un module d'une librairie de segmentation et classification d'images de télédétection développée



FIGURE 4.14 – Image complète de l'agglomération strasbourgeoise.



FIGURE 4.15 – Données de référence pour l'image complète.

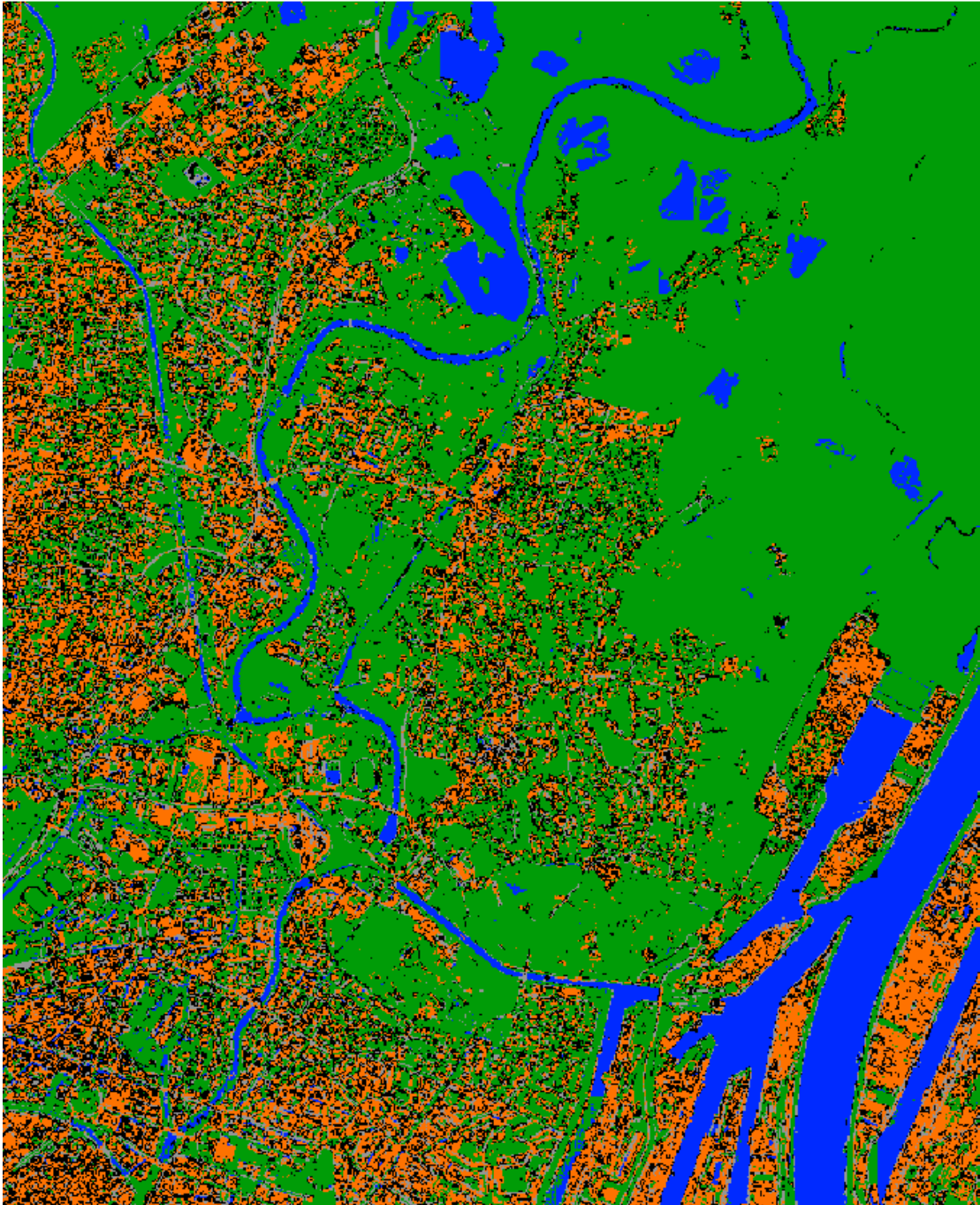


FIGURE 4.16 – Résultats de CoCoSC sur l'image complète.

durant cette thèse et qui sera intégrée à la librairie de segmentation JSL développée et maintenue par l'équipe SDC. Un module permettant de traiter des grandes images grâce à la programmation par streaming a également été développé.

Conclusion

Dans cette thèse, nous avons abordé les problématiques liées à l'interprétation automatique d'images de télédétection. L'objectif initial était le développement de méthodes multi-stratégies et collaboratives guidées par la qualité des données et des connaissances, pour extraire ou labelliser des objets géographiques de nature variée. Nous avons choisi de focaliser nos efforts sur les images optiques à très haute résolution spatiale. Par conséquent, nous avons été confrontés à un grand nombre de défis que leur traitement implique et qui restent au jour d'aujourd'hui des problématiques ouvertes.

Nos recherches, au croisement du traitement d'images, de la fouille de données et de l'observation de la Terre nous ont amenés à proposer des approches innovantes sur quatre axes portant sur :

1. l'évaluation de la qualité en segmentation d'images ;
2. la création de données de référence ;
3. l'analyse collaborative multi-paradigme mono-classe ;
4. l'agrégation multi-classe d'approches collaboratives mono-classes.

Dans la suite de ce chapitre nous résumons nos contributions relatives à chacun de ces axes de recherche, ainsi que leurs perspectives de recherche. Nous concluons enfin par des perspectives plus globales.

Évaluation de la qualité en segmentation d'images

Dans la chaîne de traitements des approches OBIA, la segmentation d'images est à la base du processus, et de ce fait, est l'une des étapes les plus importantes.

Nous avons exploré les différentes approches permettant d'évaluer la qualité d'une telle segmentation. Les approches subjectives représentent une tâche longue et répétitive ; de plus, elles fournissent des résultats entachés d'incertitude dus au biais induit par l'expertise, les connaissances et les points de vue *a priori* de l'évaluateur. Les approches supervisées fournissent des mesures objectives de la qualité pourvu qu'il existe une segmentation de référence (« idéale ») à utiliser comme point de comparaison. Or en télédétection à très haute résolution spatiale, la quasi-totalité des objets géographiques deviennent observables, mais présentent à la fois une complexité due à leur nature et une complexité importante due à la représentation de l'image (bruit des capteurs, crénelage des lignes, etc.), combinées à une diversité de ces objets en forme et en types ; cela rend quasi impossible la construction manuelle de segmentations (ou de classifications) de référence, ce qui limite drastiquement l'utilisation de telles mesures.

Les approches non-supervisées qui sont basées uniquement sur des propriétés intrinsèques de l'image et des segments, nous semblent être les plus adaptées à l'évaluation d'une segmentation mais aussi les plus difficiles à mettre en œuvre. Nos recherches menées dans ce domaine ont donné

lieu à l'approche d'évaluation non-supervisée innovante *UOA* (Under- and Over-segmentation Aware). Cette métrique, basée sur la qualité locale des segments, permet notamment de quantifier les erreurs de segmentation, mais aussi de les localiser au niveau des segments afin d'en tenir compte lors de l'étape de classification. Nos expériences ont montré que les différentes variantes de cette métrique sont supérieures sur plusieurs aspects aux métriques de l'état de l'art.

À court terme, nous envisageons d'intégrer des critères de forme en plus des critères radiométriques dans le calcul de l'évaluation.

Création de données de référence

Représenter la connaissance experte en télédétection est un processus complexe. D'une part, il est possible de modéliser les connaissances de manière explicite sous la forme de règles ou d'ontologies. D'autre part il est possible d'encoder ces connaissances de manière implicite sous la forme d'exemples. Il n'existe pas, à l'heure actuelle, de connaissances formalisées sous la forme d'ontologies réellement utilisables. Le succès des méthodes de classification supervisée repose donc sur l'existence de données labellisées de qualité. Elles sont généralement employées pour entraîner ou guider les modèles de reconnaissance d'objets, mais aussi pour réaliser une évaluation objective et précise des résultats finaux. Afin de disposer d'un tel jeu de données, nous avons mis en place un processus en trois étapes : (1) création d'une segmentation de bonne qualité mais insuffisante ; (2) évaluation locale multi-expert des segments par le biais de *crowd-sourcing* ; (3) création d'un jeu de données de (très) bonne qualité. Ces données de référence validées par des experts sont composées de 3 394 segments (labellisés chacun par leur classe thématique) et ont vocation à être distribuées gratuitement.

Les données collectées pourront être utilisées en particulier pour apprendre de manière supervisée un critère permettant de déterminer la qualité d'un segment qui soit en accord avec les avis de la foule. En effet, elles modélisent une notion de qualité multi-vue. Il semble intéressant d'utiliser cette information pour obtenir une fonction d'évaluation locale qui ne dépend pas de paramètres sensibles.

Analyse collaborative multi-paradigme mono-classe

Il existe une multitude d'approches envisageables pour intégrer de manière collaborative différentes stratégies utiles à la classification des images. Dans cette thèse, nous nous sommes focalisés sur les méthodes multi-paradigmes segmentation-classification.

Notre principale contribution dans ce domaine a été la définition d'un cadre collaboratif et générique de segmentation et classification. Le processus est guidé par la qualité de classification et de la segmentation, pour modifier de manière cohérente les segments considérés de qualité insuffisante. Le cadre est conçu pour l'extraction d'une seule classe thématique à la fois. En effet, cela permet de se focaliser sur les propriétés particulières de celle-ci pour en optimiser l'extraction. Grâce à sa généralité, le cadre peut être facilement appliqué à l'extraction de classes différentes et entraîne, d'après les expériences menées, une amélioration globale de la segmentation et de la classification, pourvu que le modèle de classification soit correctement entraîné.

L'une des limites de l'approche proposée est le fait qu'elle dépend directement du modèle de classification appris de manière supervisée. Or, cela implique un travail supplémentaire à accomplir en pré-traitement (*offline*) qui n'est pas toujours évident. Une possibilité qui nous semble intéressante pour supprimer cette contrainte est de rendre la méthode totalement non-supervisée. En effet, il serait possible d'utiliser une classification non-supervisée floue en considérant chaque cluster comme une classe thématique différente et les scores d'appartenance associés comme une

probabilité de classification. De plus, l'information d'appartenance floue pourra être à la base de la sélection des segments à modifier en favorisant, par exemple, ceux ne présentant pas une appartenance nette à une seule classe. Enfin, la qualité de la classification floue pourrait être intégrée dans le critère de qualité globale.

Agrégation multi-classe d'approches collaboratives mono-classe

Nos résultats sur l'extraction mono-classe nous ont menés à la proposition de deux approches différentes pour l'agrégation collaborative d'agents collaboratifs, en vue d'obtenir une labellisation complète de l'image. La première, dite séquentielle, consiste en l'application en cascade d'agents mono-classes. Elle permet de profiter des améliorations successives de la segmentation pour classifier l'image de manière incrémentale. À chaque étape, un agent collaboratif effectue l'extraction d'une classe thématique. Nos expériences ont montré que chaque étape du processus permet effectivement d'améliorer la segmentation et de fait la classification des objets de la classe en question.

La deuxième proposition, dite parallèle, met en place une stratégie complètement collaborative, où chaque agent travaille de manière indépendante et parallèle sur une segmentation différente. Des échanges d'information sur les conflits de classification permettent d'améliorer les segmentations dans les zones les plus ambiguës et difficiles à classer. Ce processus est itéré plusieurs fois afin de réduire le taux de segments ambigus. Nos premiers résultats montrent que l'approche parallèle atteint des performances similaires à l'approche en cascade en termes de précision, mais ne produit plus que 4.5% de pixels non décidés, alors que la version séquentielle fournit un taux de pixels non décidés qui s'élève à 36.7%.

Bien que les résultats obtenus soient très encourageants, la mise en place de stratégies d'agrégation d'agents mono-classes n'est pas triviale et le processus peut devenir rapidement très complexe à mettre en œuvre. Des études théoriques et empiriques sont donc encore nécessaires afin de formaliser le processus (informations à échanger, mécanismes à optimiser, etc.) et de mieux l'évaluer.

Perspectives à long terme

Nous pourrions considérer principalement deux perspectives de recherche. La première est l'intégration de la métrique *UOA* dans un algorithme de segmentation, pour construire progressivement la meilleure segmentation possible. Les algorithmes de segmentation hiérarchiques, tel que le Binary Partition Tree (BPT), s'avèrent être de bons candidats pour explorer cette voie de recherche. En effet, ces algorithmes sont basés sur une mesure de similarité pour construire un espace de partitionnements hiérarchique en fusionnant à chaque fois les deux partitions les plus semblables. Il serait envisageable d'y intégrer le critère d'évaluation de la segmentation pour construire l'arborescence minimisant les erreurs de segmentation à une échelle donnée.

La seconde perspective porte sur le cadre collaboratif développé dans cette thèse. Il a été conçu pour être le plus générique possible. Ainsi, il est facilement adaptable pour considérer divers critères pour guider le processus. Il est également possible d'exprimer d'autres méthodes similaires proposées dans la littérature, dans le formalisme du cadre proposé. Néanmoins, cette généralité donne lieu à une explosion des instanciations envisageables. Bien que nos études expérimentales aient montré des résultats intéressants et prometteurs, le choix de la meilleure stratégie pour extraire un ou plusieurs types d'objets donnés reste une question ouverte, qui mérite d'être

étudiée davantage. À terme, l'objectif sera de comprendre entièrement comment les paradigmes de classification et de segmentation interagissent entre eux pour formaliser un cadre théorique sur cette famille d'approches multi-paradigmes.

Bibliographie

- [Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:2274–2282.
- [Andres et al., 2009] Andres, B., Köthe, U., Bonea, A., Nadler, B., and Hamprecht, F. A. (2009). Quantitative assessment of image segmentation quality by random walk relaxation times. In *Joint Pattern Recognition Symposium*, pages 502–511. Springer.
- [Baatz and Schäpe, 2000] Baatz, M. and Schäpe, A. (2000). Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung XII*, pages 12–23.
- [Belarte et al., 2013] Belarte, B., Wemmert, C., Forestier, G., Grizonnet, M., and Weber, C. (2013). Learning fuzzy rules to characterize objects of interest from remote sensing images. In *IEEE International Geoscience and Remote Sensing Symposium – IGARSS*, pages 2986–2989.
- [Binaghi et al., 1999] Binaghi, E., Brivio, P. A., Ghezzi, P., and Rampini, A. (1999). A fuzzy set-based accuracy assessment of soft classification. *Pattern Recognition Letters*, 20(9):935–948.
- [Blaschke, 2010] Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65:2–16.
- [Breiman, 1996] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
- [Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
- [Chen and Wang, 2004] Chen, H.-C. and Wang, S.-J. (2004). The use of visible color difference in the quantitative evaluation of color image segmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP*, pages 593–596.
- [Cheng et al., 2001] Cheng, H., Jiang, X., Sun, Y., and Wang, J. (2001). Color image segmentation: Advances and prospects. *Pattern Recognition*, 34:2259–2281.
- [Cheng et al., 2004] Cheng, J., Cline, M., Martin, J., Finkelstein, D., Awad, T., Kulp, D., and Siani-Rose, M. A. (2004). A knowledge-based clustering algorithm driven by gene ontology. *Journal of Biopharmaceutical Statistics*, 14(3):687–700.
- [Chow, 1970] Chow, C. (1970). On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46.
- [Clinton et al., 2010] Clinton, N. E., Holt, A., and Gong, P. (2010). Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering & Remote Sensing*, 76:289–299.
- [Comaniciu and Meer, 1999] Comaniciu, D. and Meer, P. (1999). Mean shift analysis and applications. In *IEEE International Conference on Computer Vision – ICCV*, 2:1197–1203.

- [Coppin et al., 2004] Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., and Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25:1565–1596.
- [Corcoran et al., 2010] Corcoran, P., Winstanley, A., and Mooney, P. (2010). Segmentation performance evaluation for object-based remotely sensed image analysis. *International Journal of Remote Sensing*, 31:617–645.
- [Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- [Cristianini and Shawe-Taylor, 2000] Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.
- [Duan and Keerthi, 2005] Duan, K.-B. and Keerthi, S. S. (2005). Which is the best multiclass SVM method? an empirical study. In *Multiple Classifier Systems*, pages 278–285. Springer.
- [Dunn, 1973] Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters.
- [Duveiller et al., 2008] Duveiller, G., Defourny, P., Desclée, B., and Mayaux, P. (2008). Deforestation in central africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed landsat extracts. *Remote Sensing of Environment*, 112(5):1969 – 1981.
- [Esch et al., 2013] Esch, T., Taubenböck, H., Tal, A., Chrysoulakis, N., Feigenwinter, C., Parlow, E., and Düzgün, H. S. (2013). Exploiting Earth observation in sustainable urban planning and management – The geourban project. In *Urban Remote Sensing Event – JURSE*, pages 037–040.
- [Farmer and Jain, 2005] Farmer, M. and Jain, A. (2005). A wrapper-based approach to image segmentation and classification. *IEEE Transactions on Image Processing*, 14:2060–2072.
- [Fauvel et al., 2012] Fauvel, M., Chanussot, J., and Benediktsson, J. (2012). A spatial–spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition*, 45(1):381–392.
- [Fodeh et al., 2011] Fodeh, S., Punch, B., and Tan, P.-N. (2011). On ontology-driven document clustering using core semantic features. *Knowledge and Information Systems*, 28(2):395–421.
- [Forestier et al., 2010] Forestier, G., Gançarski, P., and Wemmert, C. (2010). Collaborative clustering with background knowledge. *Data & Knowledge Engineering*, 69(2):211–228.
- [Forestier et al., 2013] Forestier, G., Wemmert, C., and Puissant, A. (2013). Coastal image interpretation using background knowledge and semantics. *Computers & Geosciences*, 54(0):88 – 96.
- [Freund et al., 1999] Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal of Japanese Society For Artificial Intelligence*, 14(771–780):1612.
- [Gao et al., 2011] Gao, Y., Mas, J. F., Kerle, N., and Navarrete Pacheco, J. A. (2011). Optimal region growing segmentation and its effect on classification accuracy. *International Journal of Remote Sensing*, 32(13):3747–3763.
- [Gordillo et al., 2013] Gordillo, N., Montseny, E., and Sobrevilla, P. (2013). State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging*, 31(8):1426–1438.
- [Grozavu et al., 2015] Grozavu, N., Rogovschi, N., Cabanes, G., Troya-Galvis, A., and Gançarski, P. (2015). VHR satellite image segmentation based on topological unsupervised learning. In *International Conference on Machine Vision Applications – MVA*, pages 543–546.

- [Hall et al., 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18.
- [Haykin, 1998] Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition.
- [Hofmann et al., 2014] Hofmann, P., Lettmayer, P., Blaschke, T., Belgiu, M., Wegenkittl, S., Graf, R., Lampoltshammer, T. J., and Andrejchenko, V. (2014). ABIA - A conceptual framework for agent based image analysis. *South-Eastern European Journal of Earth Observation and Geomatics*, 3(25):125–129.
- [Hulten et al., 2001] Hulten, G., Spencer, L., and Domingos, P. (2001). Mining time-changing data streams. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD*, pages 97–106.
- [Jha et al., 2008] Jha, M. N., Levy, J., and Gao, Y. (2008). Advances in remote sensing for oil spill disaster management: state-of-the-art sensors technology for oil spill surveillance. *Sensors*, 8(1):236–255.
- [Johnson and Xie, 2011] Johnson, B. and Xie, Z. (2011). Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66:473–483.
- [Kekre et al., 2010] Kekre, H., Thepade, S. D., Sarode, T. K., and Suryawanshi, V. (2010). Image retrieval using texture features extracted from GLCM, LBG and KPE. *International Journal of Computer Theory and Engineering*, 2(5):695.
- [Khan and Bhuiyan, 2014] Khan, J. F. and Bhuiyan, S. M. (2014). Weighted entropy for segmentation evaluation. *Optics & Laser Technology*, 57:236–242.
- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.
- [Kuncheva, 2004] Kuncheva, L. I. (2004). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
- [Kurtz et al., 2012] Kurtz, C., Passat, N., Gançarski, P., and Puissant, A. (2012). Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology. *Pattern Recognition*, 45:685–706.
- [Lacoste et al., 2010] Lacoste, C., Descombes, X., and Zerubia, J. (2010). Unsupervised line network extraction in remote sensing using a polyline process. *Pattern Recognition*, 43(4):1631–1641.
- [Li et al., 2013] Li, H., Cai, J., Nguyen, T. N. A., and Zheng, J. (2013). A benchmark for semantic image segmentation. In *ICME, Proc.*, pages 1–6.
- [Liasis and Stavrou, 2016] Liasis, G. and Stavrou, S. (2016). Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing*, 37(5):1127–1153.
- [Lin et al., 2013] Lin, J., Peng, B., Li, T., and Chen, Q. (2013). A learning-based framework for image segmentation evaluation. In *Intelligent Networking and Collaborative Systems (INCoS), 2013 5th International Conference on*, pages 691–696.
- [Lizarazo and Elsner, 2011] Lizarazo, I. and Elsner, P. (2011). Segmentation of remotely sensed imagery: Moving from sharp objects to fuzzy regions. *INTECH Open Access Publisher*.

- [Machairas et al., 2015] Machairas, V., Faessel, M., Cárdenas-Peña, D., Chabardes, T., Walter, T. and Decencière, E. (2015). Waterpixels. In *IEEE Transactions on Image Processing*, 24(11):3707–3716
- [MacQueen et al., 1967] MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability*, 1(14):281–297.
- [Mahmoudi et al., 2013] Mahmoudi, F. T., Samadzadegan, F., and Reinartz, P. (2013). Object oriented image analysis based on multi-agent recognition system. *Computers & Geosciences*, 54:219–230.
- [Martin et al., 2001] Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV, Proc.*, pages 416–423.
- [Mather and Koch, 2011] Mather, P. M. and Koch, M. (2011). *Computer Processing of Remotely-Sensed Images: An Introduction*. John Wiley & Sons.
- [Memarsadeghi et al., 2007] Memarsadeghi, N., Mount, D. M., Netanyahu, N. S., and Le Moigne, J. (2007). A fast implementation of the isodata clustering algorithm. *International Journal of Computational Geometry & Applications*, 17(1):71–103.
- [Mnih et al., 2015] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.
- [Möller et al., 2013] Möller, M., Birger, J., Gidudu, A., and Gläßer, C. (2013). A framework for the geometric accuracy assessment of classified objects. *International Journal of Remote Sensing*, 34:8685–8698.
- [Monteiro and Campilho, 2012] Monteiro, F. C. and Campilho, A. C. (2012). Distance measures for image segmentation evaluation. In *ICNAAM, Proc.*, volume 1479, pages 794–797.
- [Musci et al., 2013] Musci, M., Feitosa, R., and Costa, G. (2013). An object-based image analysis approach based on independent segmentations. In *Urban Remote Sensing Event – JURSE*, pages 275–278.
- [Nielsen, 2015] Nielsen, M. M. (2015). Remote sensing for urban planning and management: The use of window-independent context segmentation to extract urban features in stockholm. *Computers, Environment and Urban Systems*, 52:1–9.
- [Paglieroni, 2004] Paglieroni, D. W. (2004). Design considerations for image segmentation quality assessment measures. *Pattern Recogn.*, 37:1607–1617.
- [Park and Jun, 2009] Park, H.-S. and Jun, C.-H. (2009). A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, 36(2):3336–3341.
- [Pavlidis, 1977] Pavlidis, T. (1977). *Structural Pattern Recognition*. Springer series in electrophysics. Springer-Verlag, Berlin, New York.
- [Peng and Li, 2013] Peng, B. and Li, T. (2013). A probabilistic measure for quantitative evaluation of image segmentation. *IEEE Signal Processing Letters*, 20:689–692.
- [Peng et al., 2013] Peng, B., Zhang, L., and Zhang, D. (2013). A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46:1020–1038.
- [Pham et al., 2011] Pham, H. M., Yamaguchi, Y., and Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. *Landscape Urban Plan*, 100:223–230.

- [Polak et al., 2009] Polak, M., Zhang, H., and Pi, M. (2009). An evaluation metric for image segmentation of multiple objects. *Image Vision Comput*, 27:1223–1227.
- [Pont-Tuset and Marques, 2013] Pont-Tuset, J. and Marques, F. (2013). Measures and meta-measures for the supervised evaluation of image segmentation. In *CVPR, Proc*, pages 2131–2138.
- [Promper et al., 2014] Promper, C., Puissant, A., Malet, J.-P., and Glade, T. (2014). Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios. *Applied Geography*, 53:11–19.
- [Qian et al., 2008] Qian, X., Zhang, X., Jiao, L., and Ma, W. (2008). Unsupervised texture image segmentation using multiobjective evolutionary clustering ensemble algorithm. In *Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence). IEEE Congress on*, pages 3561–3567.
- [Quinlan, 1986] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.
- [Quinlan, 2014] Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier.
- [Räsänen et al., 2013] Räsänen, A., Rusanen, A., Kuitunen, M., and Lensu, A. (2013). What makes segmentation good? A case study in boreal forest habitat mapping. *International Journal of Remote Sensing*, 34:8603–8627.
- [Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386.
- [Rouse et al., 1974] Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation*. Texas A & M University, Remote Sensing Center.
- [Rumelhart et al., 1988] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive Modeling*, 5(3):1.
- [Salembier and Garrido, 2000] Salembier, P. and 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.
- [Sarle, 1991] Sarle, W. S. (1991). Finding groups in data: An introduction to cluster analysis. *Journal of the American Statistical Association*, 86(415):830–833.
- [Sellaouti et al., 2012] Sellaouti, A., Hamouda, A., Deruyver, A., and Wemmert, C. (2012). Hierarchical classification-based region growing (HCBRG): A collaborative approach for object segmentation and classification. In *International Conference Image Analysis and Recognition*, pages 51–60. Springer.
- [Silver et al., 2016] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.
- [Smola and Schölkopf, 2004] Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.
- [Srubar, 2012] Srubar, S. (2012). Quality measurement of image segmentation evaluation methods. In *SITIS, Proc.*, pages 254–258.
- [Sublime et al., 2015] Sublime, J., Troya-Galvis, A., Bennani, Y., Cornuéjols, A., and Gançarski, P. (2015). Semantic rich icm algorithm for vhr satellite images segmentation. *14th IAPR International Conference on Machine Vision Applications (MVA)*.
- [Surowiecki, 2005] Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

- [Tarabalka et al., 2009] Tarabalka, Y., Benediktsson, J., and Chanussot, J. (2009). Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(8):2973–2987.
- [Tarabalka et al., 2010] Tarabalka, Y., Chanussot, J., and Benediktsson, J. (2010). Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367 – 2379.
- [Tari et al., 2009] Tari, L., Baral, C., and Kim, S. (2009). Fuzzy c-means clustering with prior biological knowledge. *Journal of Biomedical Informatics*, 42(1):74 – 81.
- [Troya Galvis, 2015] Troya Galvis, A. (2015). Quantification non supervisée de la sous- et sur-segmentation pour la classification d’images de télédétection. *Actes de l’atelier FDC 2015*.
- [Troya-Galvis and Gancarski, 2015] Troya-Galvis, A. and Gancarski, P. (2015). Vers une approche collaborative segmentation-classification pour l’analyse d’images de télédétection. *Rencontres des Jeunes Chercheurs en Intelligence Artificielle 2015*.
- [Troya Galvis et al., 2016a] Troya Galvis, A., Gancarski, P., and Berti-Équille, L. (2016a). Collaborative segmentation and classification for remote sensing image analysis. *Proceedings of the International Conference on Pattern Recognition ICPR 2016*.
- [Troya Galvis et al., 2016b] Troya Galvis, A., Gancarski, P., and Berti-Équille, L. (2016b). Un cadre collaboratif pour la segmentation et la classification d’images de télédétection. *Actes des rencontres francophones sur l’Extraction et la Gestion des Connaissances (EGC 2016)*.
- [Troya-Galvis et al., 2015] Troya-Galvis, A., Gancarski, P., Passat, N., and Berti-Équille, L. (2015). Unsupervised quantification of under- and over-segmentation for object based remote sensing image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5):1936–1945.
- [Tsoumakas and Vlahavas, 2007] Tsoumakas, G. and Vlahavas, I. (2007). Random k-labelsets: An ensemble method for multilabel classification. In *European Conference on Machine Learning*, pages 406–417. Springer.
- [Unsalan and Sirmacek, 2012] Unsalan, C. and Sirmacek, B. (2012). Road network detection using probabilistic and graph theoretical methods. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4441–4453.
- [Vega-Pons et al., 2011] Vega-Pons, S., Jiang, X., and Ruiz-Shulcloper, J. (2011). Segmentation ensemble via kernels. In *Pattern Recognition (ACPR), 2011 First Asian Conference on*, pages 686–690.
- [Vojodi and Eftekhary Moghadam, 2012] Vojodi, H. and Eftekhary Moghadam, A. (2012). A supervised evaluation method based on region shape descriptor for image segmentation algorithm. In *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on*, pages 018–022.
- [Vojodi et al., 2013] Vojodi, H., Fakhari, A., and Moghadam, A. M. E. (2013). A new evaluation measure for color image segmentation based on genetic programming approach. *Image and Vision Computing*, 31:877–886.
- [Wang et al., 2010] Wang, Z., Jensen, J. R., and Im, J. (2010). An automatic region-based image segmentation algorithm for remote sensing applications. *Environmental Modelling & Software*, 25:1149–1165.
- [Ward Jr, 1963] Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

- [Weidner, 2008] Weidner, U. (2008). Contribution to the assessment of segmentation quality for remote sensing applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:479–484.
- [Witharana et al., 2014] Witharana, C., Civco, D. L., and Meyer, T. H. (2014). Evaluation of data fusion and image segmentation in earth observation based rapid mapping workflows. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:1–18.
- [Xiao and Zhan, 2009] Xiao, Y. and Zhan, Q. (2009). A review of remote sensing applications in urban planning and management in china. In *Joint Urban Remote Sensing Event – JURSE*, pages 1–5.
- [Yang et al., 2014] Yang, J., Li, P., and He, Y. (2014). A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 94(0):13–24.
- [Zhan et al., 2005] Zhan, Q., Molenaar, M., Tempfli, K., and Shi, W. (2005). Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing*, 26:2953–2974.
- [Zhang et al., 2003] Zhang, H., Fritts, J. E., and Goldman, S. A. (2003). An entropy-based objective evaluation method for image segmentation. In *Electronic Imaging*, pages 38–49.
- [Zhang et al., 2008] Zhang, H., Fritts, J. E., and Goldman, S. A. (2008). Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110:260–280.
- [Zhang et al., 2012] Zhang, X., Xiao, P., and Feng, X. (2012). An unsupervised evaluation method for remotely sensed imagery segmentation. *IEEE Geoscience and Remote Sensing Letters*, 9:156–160.
- [Zhang, 1996] Zhang, Y. (1996). A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29:1335–1346.

Segmentation et classification d'images de télédétection

La segmentation d'images et la classification des segments obtenus constituent deux des étapes les plus importantes dans la chaîne de traitement des méthodes OBIA. Nous dédions ce chapitre à la définition de ces deux paradigmes et aux méthodes existantes, notamment dans le cadre de la télédétection.

A.1 Classification

La classification, de manière informelle, consiste à trouver des modèles pour prédire soit un label, soit une valeur numérique en fonction des propriétés des données. Il s'agit d'associer une sémantique aux données brutes de manière automatique.

Formellement on peut définir le problème comme suit : soit $X = \{x_0, x_1, \dots, x_p\}$ un ensemble de p vecteurs $x_i = (f_0, f_1, \dots, f_q)$ appartenant à un espace F caractérisant les propriétés de chaque donnée à classer, où chaque f_i est un attribut nominal, numérique ou binaire. Soit $L = \{l_0, \dots, l_N\}$ l'ensemble de labels correspondant aux N classes recherchées. Il s'agit de trouver un fonction

$$f : F \rightarrow L \\ x \mapsto l$$

qui associe à chaque élément $x \in X$ une classe ou un label $l \in L$.

Les techniques de classification font souvent appel à l'apprentissage automatique. Nous pouvons donc citer principalement (mais non-exhaustivement) trois types de méthodes.

A.1.1 Approches supervisées

Les approches supervisées partent d'un jeu de données $X = \{x_0, x_1, \dots, x_p\}$ pour lequel on connaît l'ensemble de labels $y = y_0, y_1, \dots, y_p$ tels que $f(x_i) = y_i$. Généralement, cet ensemble de données labellisées est alors divisé en deux sous-ensembles X_t , appelé jeu d'entraînement, et X_v , appelé jeu de validation. À partir du jeu d'entraînement, on essaye de trouver une fonction $f_s : F \rightarrow L$ minimisant $\sum_i E(f_s(x_i), y_i)$ où E est une fonction de coût quantifiant les erreurs commises par f_s . La fonction apprise f_s est souvent appelée modèle de prédiction. Finalement, le

modèle f_s appris est appliqué sur les données X_v afin de valider la capacité de généralisation de celui-ci. On appelle apprentissage automatique les techniques permettant d'optimiser la fonction de coût E afin de trouver le modèle f_s . Il existe des nombreuses techniques pour l'apprentissage du modèle f_s , nous citons ci-dessous les principales familles que l'on retrouve dans la littérature ainsi que leurs variantes les plus remarquables.

Les arbres décisionnels apprennent le modèle classification par la construction d'un arbre dont les nœuds représentent des choix faits pour séparer les données en fonction d'une propriété particulière ; les feuilles constituent l'ensemble des labels possibles. La construction de l'arbre se fait en trouvant de manière itérative l'attribut (et les valeurs de celui-ci) permettant de séparer au maximum les données en faisant le moins d'erreurs de classification possibles. Dans cette catégorie nous retrouvons des méthodes telles que l'algorithme ID3 [Quinlan, 1986], C4.5 [Quinlan, 2014] CART [Breiman et al., 1984]. Il existe également des variantes incrémentales comme l'algorithme VFDT [?], qui permettent de prendre en compte des données qui évoluent dans le temps en remettant en cause le modèle appris lorsque la performance de celui-ci se voit diminuée. L'un des avantages les plus remarquables de ces approches est l'obtention de modèles faciles à interpréter par un être humain, ce qui facilite l'analyse et la compréhension des données complexes.

Les séparateurs à vastes marges ou *Support Vector Machines* (SVM) sont une méthode de classification binaire (permettant de discriminer deux classes) linéaire qui consiste à trouver l'hyperplan permettant de séparer l'espace des données en deux groupes distincts et dont la distance entre les points les plus proches à l'hyperplan est la plus large possible [Cortes and Vapnik, 1995]. Des variantes permettant de classer des données non-linéairement séparables ont été introduites en utilisant le dit *kernel trick* [Cristianini and Shawe-Taylor, 2000]. D'autres variantes ont été proposées permettant de faire de la régression [Smola and Schölkopf, 2004] et pour gérer les problèmes de classification multi-classe [Duan and Keerthi, 2005].

Les réseaux de neurones artificiels s'inspirent du fonctionnement des neurones biologiques pour construire des modèles de prédiction. Un neurone artificiel, aussi appelé perceptron [Rosenblatt, 1958] est modélisé de manière très simple par $y = \varphi(\sum_{i=0}^m w_i x_i)$. Il prend m entrées qui sont combinées au moyen d'une somme pondérée par les poids w_i ; la sortie y d'un neurone est le résultat de la fonction ϕ , aussi appelée fonction d'activation ou de transfert, sur la somme des entrées, et correspond souvent à une fonction de type sigmoïde. Un neurone individuel calcule un hyperplan séparateur dans l'espace \mathbb{R}^m . Il est possible de connecter plusieurs neurones entre eux afin de produire des modèles plus expressifs. L'apprentissage d'un modèle est fait en propageant l'erreur de classification sur tout le réseau afin d'adapter de manière itérative les poids de tous les neurones du réseau [Rumelhart et al., 1988]. Un grand nombre de modèles basés sur des réseaux de convolutions ont été récemment proposés ; ils permettent d'apprendre automatiquement les caractéristiques les plus appropriées pour effectuer la classification et atteignent actuellement des performances sur-humaines sur des tâches comme la reconnaissance des formes [Krizhevsky et al., 2012], les jeux vidéo [Mnih et al., 2015], ou encore le jeu traditionnel Go [Silver et al., 2016]. Ces méthodes produisent des modèles très performants mais nécessitent un grand nombre de données labellisées, demandent une puissance de calcul importante et produisent des modèles très difficiles voire impossibles à interpréter.

Les méta-algorithmes, souvent appelés méthodes d'apprentissage par ensemble, font appel à plusieurs algorithmes de classification et mettent en œuvre des stratégies pour les combiner afin de profiter des avantages de chaque algorithme et ainsi obtenir des modèles plus stables et robustes. Nous y trouvons notamment la méthode de bagging [Breiman, 1996] qui consiste à diviser le jeu d'entraînement en plusieurs échantillons aléatoires et appliquer des méthodes de classification (instables de préférence) différentes sur chacun des échantillons pour finalement agréger les résultats et obtenir une unique réponse. Un autre exemple est la méthode de boosting [Freund et al., 1999] qui consiste à entraîner des classifieurs dits faibles et à les ajouter

pour obtenir une classification finale forte. Les classifieurs sont entraînés de manière séquentielle et après chaque étape, les instances sont pondérées en fonction des résultats précédents; ainsi, les nouveaux classifieurs se focalisent davantage sur les exemples qui ont été mal classés à plusieurs reprises auparavant. Pour plus de détails sur ce sujet, le lecteur intéressé peut se référer à [Kuncheva, 2004].

Les principales difficultés des approches supervisées sont :

- l'acquisition de données labellisées suffisantes et de qualité qui sont nécessaires pour l'obtention de modèles précis;
- le choix et la création de caractéristiques pertinentes pour réaliser une tâche de classification donnée;
- la mise en œuvre de stratégies permettant de réduire le sur-apprentissage et donc de généraliser le modèle de classification à des données nouvelles.

A.1.2 Approches non-supervisées

Aussi appelées méthodes de *clustering*, les approches non-supervisées n'emploient pas de données labellisées pour effectuer la tâche de classification. Elles exploitent les données brutes et essayent d'y trouver des patrons ou structures cachées pour construire des groupes de données similaires appelés *clusters*. Les méthodes de clustering répondent au besoin de disposer de techniques pour comprendre, traiter et résumer des grandes collections de données de façon automatique. Nous pouvons définir le problème du clustering comme suit : soit $X = \{x_0, x_1, \dots, x_p\}$ un ensemble de p vecteurs $x_i = (f_0, f_1, \dots, f_q)$ appartenant à un espace F ; soit $H(X_c)$ une fonction quantifiant la similarité des éléments appartenant à un ensemble X_c ; trouver un partitionnement $C = \bigcup X_i, ; i \in \{1, \dots, K\}, R_i \cap R_j = \emptyset ; \forall i, j, ; i \neq j$ en K groupes d'objets tels que $H(X_i)$ est maximale $\forall i \in \{1, \dots, K\}$ et tels que $H(X_i \cup X_j)$ est minimale $\forall i, j ; i \neq j$.

Nous y trouvons principalement deux types d'approches :

Les approches par partitionnement séparent l'espace de données directement en K sous-ensembles distincts. La méthode la plus populaire de cette famille est sans doute K-means [MacQueen et al., 1967]. Elle part de K points différents choisis aléatoirement pour initialiser les centres des clusters, la première étape consiste à calculer la moyenne des tous les points appartenant à un même cluster pour déterminer la position des nouveaux centres. La seconde étape consiste à associer chaque point au centre de cluster le plus proche. Ces étapes sont itérées jusqu'à convergence. Des multiples variantes et améliorations existent dans la littérature; on trouve par exemple : ISODATA [Memarsadeghi et al., 2007], qui ne requiert pas un nombre pré-défini de clusters pour fonctionner; Fuzzy c-means [Dunn, 1973], qui est une version floue de K-means où un point n'appartient pas obligatoirement à un unique cluster, mais l'appartenance des points aux clusters est déterminée par un vecteur d'appartenances à tous les clusters; ou K-medoïdes [Park and Jun, 2009], qui est une version de K-means plus robuste vis-à-vis des données aberrantes.

Les approches hiérarchiques cherchent des groupements imbriqués en divisant (ou fusionnant) de manière récursive des clusters intermédiaires. Les approches par regroupement, couramment connues sous le nom de classification ascendante hiérarchique (CAH) [Ward Jr, 1963] construisent un arbre en partant des feuilles (données individuelles); chaque nœud de l'arbre est construit itérativement en regroupant à chaque fois les deux éléments (ou groupes d'éléments) les plus similaires en fonction d'une mesure de distance donnée. Le processus s'arrête lorsque toutes les données ont été regroupées en un unique cluster. Au contraire, les approches par division [Sarle, 1991] partent de la racine de l'arbre (la totalité des données) et procèdent en divisant les données de sorte à maximiser la dissimilarité entre les clusters jusqu'à obtenir des singletons (les

feuilles de l'arbre). Ces approches requièrent une étape supplémentaire consistant à trouver une coupe de l'arbre permettant d'obtenir le nombre de clusters souhaité.

A.1.3 Approches semi-supervisées et guidées par des connaissances

Les approches semi-supervisées sont des méthodes hybrides qui visent à guider des processus non-supervisés en y injectant des connaissances *a priori* sur les données. Ces connaissances peuvent être introduites sous la forme de contraintes de type *must-link* ou *cannot-link* déterminant si une paire d'objets doit être dans le même cluster ou pas. Un exemple de l'utilisation de ce genre de contraintes est donnée dans [Forestier et al., 2010]; les auteurs proposent plusieurs métriques permettant de prendre en compte des contraintes sur la qualité interne des clusters, sur des sous-ensembles d'objets labellisés, et sur des relations *must-link* ou *cannot-link*, pour guider une méthode de clustering collaboratif.

Une autre façon d'intégrer des connaissances consiste à s'appuyer sur une modélisation formelle de la connaissance experte sur un domaine applicatif concret pour guider le processus de classification. Ce type de méthodes modélise la connaissance experte grâce à des ontologies, et sont très populaires en bio-informatique [Cheng et al., 2004, Tari et al., 2009] étant donnée le nombre important de ressources disponibles telles que l'ontologie Gene Ontology, ou dans le domaine du clustering de documents [Fodeh et al., 2011]. Des efforts ont également été faits pour modéliser les connaissances dans le domaine de la télédétection. En particulier, dans [Forestier et al., 2013], les auteurs présentent une approche permettant d'exploiter les connaissances expertes pour guider un processus d'interprétation d'images grâce au concept de fonction de connaissance qui fait appel à une ontologie pour déterminer le degré de cohérence entre un segment et une classe d'objets.

Il est possible de représenter une image de télédétection I comme un ensemble de pixels $p = (b_0, \dots, b_m)$ où chaque b_i correspond à la valeur de la bande radiométrique i . Ainsi, les techniques de classification sont souvent employées pour faciliter l'analyse de ce type d'images.

A.2 Segmentation d'images

La segmentation d'images est le problème consistant à partitionner une image en groupes de pixels similaires selon un critère donné. Selon le type d'application, on peut distinguer trois grandes familles d'approches : d'une part, on trouve les méthodes employées pour la détection d'objets, qui cherchent à partitionner l'image en peu de régions (généralement deux, l'objet d'intérêt et le fond). Ce type d'approches est courant en imagerie médicale, par exemple dans la segmentation de tumeurs cérébrales [Gordillo et al., 2013]. D'autre part, on trouve des méthodes plus générales qui cherchent un partitionnement complet de l'image en zones homogènes selon un critère donné. L'objectif de ces approches est d'obtenir autant de segments qu'il y a d'objets dans l'image. Finalement, des méthodes plus récentes dites de super-pixels cherchent à partitionner l'image en petits groupes de pixels très similaires et compactes en s'affranchissant de la sémantique de l'image. Elles sont utilisées principalement pour réduire le nombre d'éléments à traiter lorsque les images sont de taille importante. Dans cette famille, on retrouve la méthode SLIC [Achanta et al., 2012] ou les water-pixels [Machairas et al., 2015] par exemple.

En télédétection, nous nous intéressons principalement aux méthodes cherchant une correspondance parfaite entre les segments et tous les éléments représentés dans l'image; dans la suite du document (sauf mention contraire) nous ne faisons référence qu'à ces dernières. Pour une étude complète sur les différentes approches de segmentation, le lecteur peut se référer à [Cheng et al., 2001].

Formellement, on peut définir la segmentation d'images comme suit [Pavlidis, 1977] :

Soit I une image, et P un prédicat défini sur un ensemble de pixels $R \subset I$ tel que $P(R) = Vrai$ si et seulement si R est uniforme selon un critère d'homogénéité donné. Alors une segmentation est une partition $S = R_1, \dots, R_n$ de I telle que les axiomes suivants sont respectés :

- $I = \bigcup R_i, ; i \in \{1, \dots, n\}$
- $R_i \cap R_j = \emptyset ; \forall i, j, ; i \neq j$
- $P(R_i) = Vrai ; ; \forall i$
- $P(R_i \cup R_j) = Faux ; ; \forall i \neq j$

A.2.1 Approches orientées régions

Les approches orientées régions fusionnent des pixels voisins similaires et séparent les pixels différents. Par exemple, la méthode RISA [Wang et al., 2010] part d'un clustering K-means sur les pixels de l'image pour obtenir un premier ensemble de points intéressants. Ensuite, une étape de croissance de régions basée sur des critères spectraux et de forme est appliquée à partir des graines générées dans l'étape précédente. Finalement, une étape de fusion basée sur une mesure de distance est réalisée pour enlever les segments de petite taille. Une autre méthode très populaire et couramment employée dans des logiciels commerciaux dédiés au traitement et l'analyse d'images de télédétection est l'algorithme de segmentation multi-résolution [Baatz and Schäpe, 2000]. C'est en effet une méthode par fusion de régions basée sur des mesures d'homogénéité locale et globale. L'algorithme possède un seul paramètre dit d'échelle qui permet de contrôler la taille moyenne des segments résultants ; ce paramètre est choisi en fonction de l'application pour obtenir des segments adaptés aux objets recherchés, d'où le nom multi-résolution.

A.2.2 Approches orientées contours

Les approches orientées contours suivent les zones de forte variation (gradient élevé) qui correspondent, en théorie, aux frontières des objets. On y retrouve par exemple des algorithmes basés sur la transformée watershed ou ligne de partage des eaux [Tarabalka et al., 2010]. Ces méthodes partent d'une image du gradient qu'elles considèrent comme un relief topographique dont elles simulent l'inondation. À chaque fois que deux sources d'eau différentes se rencontrent, on empêche leur fusion par la création d'une ligne de partage. À l'issue de cette transformation on obtient un partitionnement total de l'image qui a tendance à bien suivre les contours des objets. Cependant, ces méthodes dépendent fortement de l'image du gradient employée, ainsi que des points d'inondation initiaux. Elles sont également très sensibles au bruit, puisque celui-ci entraîne souvent des zones à gradient élevé.

A.2.3 Approches hiérarchiques

Les approches hiérarchiques utilisent une représentation de l'image sous forme d'arbre, où les feuilles correspondent généralement aux pixels, et chaque nœud représente un regroupement de tous les pixels descendant de ce nœud. L'une des méthodes les plus connues dans cette catégorie est l'arbre binaire de partitions ou *Binary Partition Tree* (BPT) [Salembier and Garrido, 2000]. Celle-ci part des pixels de l'image et, de manière similaire à un clustering hiérarchique ascendant, regroupe deux à deux les pixels (ou groupes de pixels) adjacents les plus proches en fonction d'une métrique de similarité donnée. Le regroupement se fait itérativement jusqu'à obtenir un seul groupe contenant la totalité des pixels de l'image.

Ces méthodes permettent de construire un espace de partitionnements que l'on peut exploiter pour trouver des représentations à des niveaux de détail différents en fonction de l'application.

Néanmoins le choix de la coupe de l'arbre la plus adaptée est un problème difficile et souvent dépendant du contexte applicatif.

Pour plus de détails sur les approches hiérarchiques notamment celles basées sur les graphes, le lecteur peut se référer à [Peng et al., 2013].

A.2.4 Approches par clustering

Les approches par clustering (cf. Section A.1.2), regroupent les pixels similaires dans un espace de données quelconque (généralement colorimétrique) sans se soucier dans la plupart des cas de leurs relations spatiales, telles que la méthode Mean-shift [Comaniciu and Meer, 1999] qui remplace la valeur de chaque pixel par la moyenne des pixels dans un rayon spatiale donnée et dont les valeurs se trouvent à une distance (en termes de couleur) déterminée de la moyenne.

Des techniques plus sophistiquées telles que celles décrites dans [Vega-Pons et al., 2011], appliquent les techniques de clustering par ensemble à la segmentation d'images. Le but est de trouver une partition P^* telle que la similarité entre P^* et toutes les partitions P_i de l'ensemble selon un critère donné soit maximisée. Un exemple similaire est donné dans [Qian et al., 2008]; les auteurs présentent un algorithme de segmentation d'images texturées divisé en deux phases : la première phase est un algorithme évolutionnaire visant à trouver des solutions maximisant la compacité des clusters, et la similarité entre pixels voisins ; la deuxième phase consiste à déterminer la meilleure solution par un algorithme de meta-clustering.

Les approches de segmentation par clustering apportent des solutions et simples efficaces permettant d'obtenir un partitionnement sur des critères de similarité divers. Néanmoins, elles ne tiennent pas compte des relations et des contraintes spatiales ; cela entraîne une perte d'information cruciale pour la segmentation.

Métriques supervisées classiques pour évaluer la segmentation d'images

Nous listons ci-dessous les métriques supervisées pour évaluer la segmentation d'images existant dans la littérature jusqu'aux années 2000. Pour une étude complète sur ces différentes métriques, le lecteur peut se référer à [Monteiro and Campilho, 2012].

Distance de Hamming : mesure le degré de superposition entre les régions de deux segmentations. La métrique est donc définie par :

$$D_H(S \rightarrow R) = \sum_{r_i \in R} \sum_{s_k \neq s_j, s_k \cap r_i \neq \emptyset} |r_i \cap s_k|$$

$$p = 1 - \frac{D_H(S \rightarrow R) + D_H(R \rightarrow S)}{2 \times |S|}$$
(B.1)

Local Consistency Error : permet de quantifier la consistance entre deux segmentations à des niveaux de détail différents. Elle est définie par :

$$LCE(S, R, p_i) = \frac{1}{N} \sum_i \min\{E(S, R, p_i), E(R, S, p_i)\}$$
(B.2)

où $E(S, R, p)$ mesure la similarité des segmentations S et R au pixel p .

Bidirectional Consistency Error : adaptation de LCE pour pénaliser les différences. Elle est définie par :

$$BCE(S, R, p_i) = \frac{1}{N} \sum_i \max\{E(S, R, p_i), E(R, S, p_i)\}$$
(B.3)

Partition Distance : soient S_1, S_2 deux partitions, la distance d_{sym} est définie comme le nombre de points de S_1 qu'il faut supprimer pour que $S'_1 = S_2$, alors $d_{sym}(S_1, S_2) = 0 \Rightarrow S_1 = S_2$.

Rand Index : exprime le problème comme la comparaison de deux étiquetages. Elle est définie par :

$$\frac{1}{\binom{N}{2}} \sum_{i,j;i \neq j} \mathbb{I}(l_i = l_j \wedge l'_i = l'_j) + \mathbb{I}(l_i \neq l_j \wedge l'_i \neq l'_j)$$
(B.4)

l et l' étant les étiquetages à comparer, et \mathbb{I} l'indice mesurant le nombre de fois où les segmentations sont en accord.

Distance Distribution Signatures : prend en compte la distance entre les frontières de régions correspondantes dans deux segmentations différentes. Elle est définie par :

$$b = 1 - \frac{D_B(B_S, B_R) + D_B(D_R, D_S)}{c \times (|R| + |S|)} \quad (\text{B.5})$$

où B_S et B_R sont les ensembles des points dans les frontières de R et S , et D_B mesure la distance entre deux frontières.

F-measure : applique les concepts de précision P et rappel R pour mesurer l'accord entre les frontières de deux segmentations. Elle est définie par :

$$F = \frac{PR}{\alpha R + (1 - \alpha)P} \quad (\text{B.6})$$

Earth Mover's Distance : est définie comme étant la somme minimale des coûts de « transport » des pixels entre les signatures de deux segmentations. Elle est définie par :

$$EMD(S_1, S_2) = \frac{\sum_i \sum_j f_{ij} d_{ij}}{\sum_i \sum_j f_{ij}} \quad (\text{B.7})$$

où f_{ij} est le poids (associé à chaque pixel) transporté entre les deux points.

Perceptual Discrepancy Measure : prend en compte les faux positifs et les faux négatifs en les pénalisant. Elle est définie par :

$$w_p = \frac{\alpha_p \log(1 + d_p)}{D} \quad w_n = \frac{\alpha_n d_n}{D} \quad (\text{B.8})$$

où d_p et d_n sont la distance d'un faux positif (ou négatif) à la frontière de référence, et D est la longueur de la diagonale de l'image ; la métrique finale est $s_w = 1 - (w_p + w_n)$.

Métriques non-supervisées classiques pour évaluer la segmentation d'images

Nous listons ci-dessous les métriques non-supervisées pour évaluer la segmentation d'images existant dans la littérature jusqu'aux années 2000. Pour une étude complète sur ces différentes métriques le lecteur peut se référer à [Srubar, 2012].

Goodness Uniformity : la différence des valeurs d'un pixel doit être minimale par rapport aux valeurs moyennes de son segment. Elle est définie par :

$$GU(I, S) = \sum_i \sum_{j \in s_i} [I_j - \mu(s_i)]^2 \quad (C.1)$$

$$\mu(s_i) = \frac{1}{|s_i|} \sum_{x \in s_i} I_x$$

Gray-level Contrast : dans des méthodes par seuillage, la moyenne des valeurs des segments doit avoir un certain contraste. Elle est définie par :

$$GC(I, S) = \frac{|\mu(s_f) - \mu(s_b)|}{\mu(s_f) + \mu(s_b)} \quad (C.2)$$

la formule n'est définie que pour un seuillage en 2 classes; elle peut se généraliser en calculant un contraste local deux à deux entre tous les segments et en faisant une somme pondérée de ces contrastes; ainsi :

$$LN(I, S) = \frac{\sum_i \left(|s_i| \cdot \sum_j \frac{|bord(s_i) \cap bord(s_j)|}{bord(s_i)} \cdot \frac{|\mu(s_i) - \mu(s_j)|}{\mu(s_i) + \mu(s_j)} \right)}{\sum_i |s_i|} \quad (C.3)$$

Figure of Certainty : mesure la différence entre les pixels d'un segment et la valeur de leur représentant dans le segment. Elle est définie par :

$$FOC(I, p, q) = \frac{1}{N} \sum_j \frac{1}{1 + p \cdot |I_j - I_{s_i}|^q} \quad (C.4)$$

avec N le nombre de pixels, I_{s_j} la valeur du représentant du segment, p et q des paramètres de mise à l'échelle.

Zeboudj Contrast : évalue le contraste d'un segment ainsi que le contraste à la frontière du segment. Elle est définie par :

$$\begin{aligned}
 ZC(I, S) &= \frac{1}{N} \sum_i |s_i| \cdot C(s_i) \\
 C(s_i) &= \begin{cases} 1 - \frac{IR(s_i)}{ER(s_i)} & si \quad 0 < IR(s_i) < ER(s_i) \\ ER(s_i) & si \quad IR(s_i) = 0 \\ 0 & sinon \end{cases} \\
 IR(s_i) &= \frac{1}{|s_i|} \sum_{x \in s_i} \max_{y \in s_i \cap neigh(x)} c(x, y) \\
 ER(s_i) &= \frac{1}{|bord(s_i)|} + \sum_{x \in bord(s_i)} \max_{y \in neigh(x), y \notin s_i} c(x, y) \\
 c(x, y) &= \frac{|I_x - I_y|}{L}
 \end{aligned} \tag{C.5}$$

avec $bord(s_i)$ les pixels à la frontière de s_i , L la valeur d'intensité maximale, et I_x l'intensité du pixel x .

Rosenberger criterion : fait la différence entre segments lisses et avec texture. Elle est définie par :

$$\begin{aligned}
 RC(I, S) &= \frac{\bar{D} - \underline{D}}{2} \\
 \underline{D} &= \frac{1}{|S| \cdot N} \sum_i |s_i| \cdot D_i
 \end{aligned} \tag{C.6}$$

\underline{D} étant la disparité intra-segment, représentée par l'écart-type pour les segments lisses et un ensemble de caractéristiques pour les segments texturés ; \bar{D} est la disparité inter-segment et correspond à la différence de deux segments lisses ou à la distance euclidienne entre les caractéristiques de deux segments texturés.

Erdem contrast : calcule le contraste aux bords des segments à l'aide d'une fenêtre pour réduire la sensibilité au bruit. Elle est définie par :

$$E(I, s_i) = \frac{1}{L \cdot \sqrt{3} \cdot |bord(s_i)|} + \sum_{j \in bord(s_i)} \|\mu_{in}(j) - \mu_{out}(j)\| \tag{C.7}$$

μ_{in} et μ_{out} étant les intensités moyennes dans la fenêtre à l'intérieur et à l'extérieur du segment.

Liu-Yang criterion : prend en compte le nombre de segments et leur taille. Elle est définie par :

$$LY(I, S) = \frac{\sqrt{|S|}}{1000 \cdot N} \sum_i \frac{\left(\sum_{j \in s_i} \|I_j - \mu_i\| \right)^2}{\sqrt{|s_i|}} \tag{C.8}$$

Borsotti criterion : adaptation de la formule précédente pour pénaliser des segments de même taille. Elle est définie par :

$$BC(I, S) = \frac{\sqrt{|S|}}{1000 \cdot N} \sum_i \left(\frac{\left(\sum_{j \in s_i} \|I_j - \mu_i\| \right)^2}{1 + \log|s_i|} + \left(\frac{eq(|s_i|)}{|s_i|^2} \right)^2 \right) \quad (\text{C.9})$$

:

Annexe **D**

Prototype développé

En complément des bibliothèques implémentant les différentes méthodes introduites lors de cette thèse, un logiciel indépendant, permettant de visualiser et manipuler les segmentations et les classifications obtenues, a été développé.

Le prototype permet notamment d'évaluer une segmentation selon plusieurs critères et de visualiser les résultats de cette évaluation comme l'illustre la figure D.1.

Le prototype permet également d'entraîner de manière interactive des modèles de prédiction incrémentaux, et de visualiser de plusieurs manières les résultats de la classification, comme le montre la figure D.2.

Finalement, le prototype permet de manipuler directement les segments pour corriger manuellement une segmentation, comme le montre la figure D.3. Quatre opérateurs sont disponibles pour modifier un segment : croissance, rétrécissement, fusion, et scission.

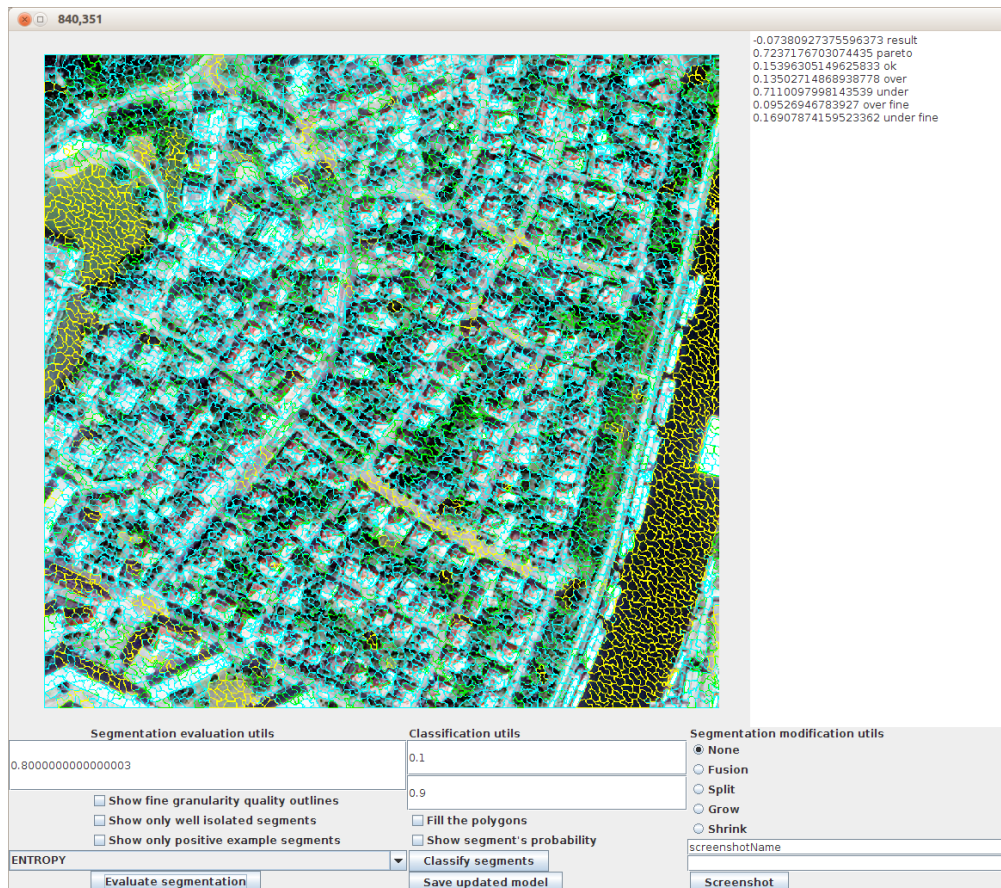


FIGURE D.1 – Visualisation de l'évaluation de la segmentation.

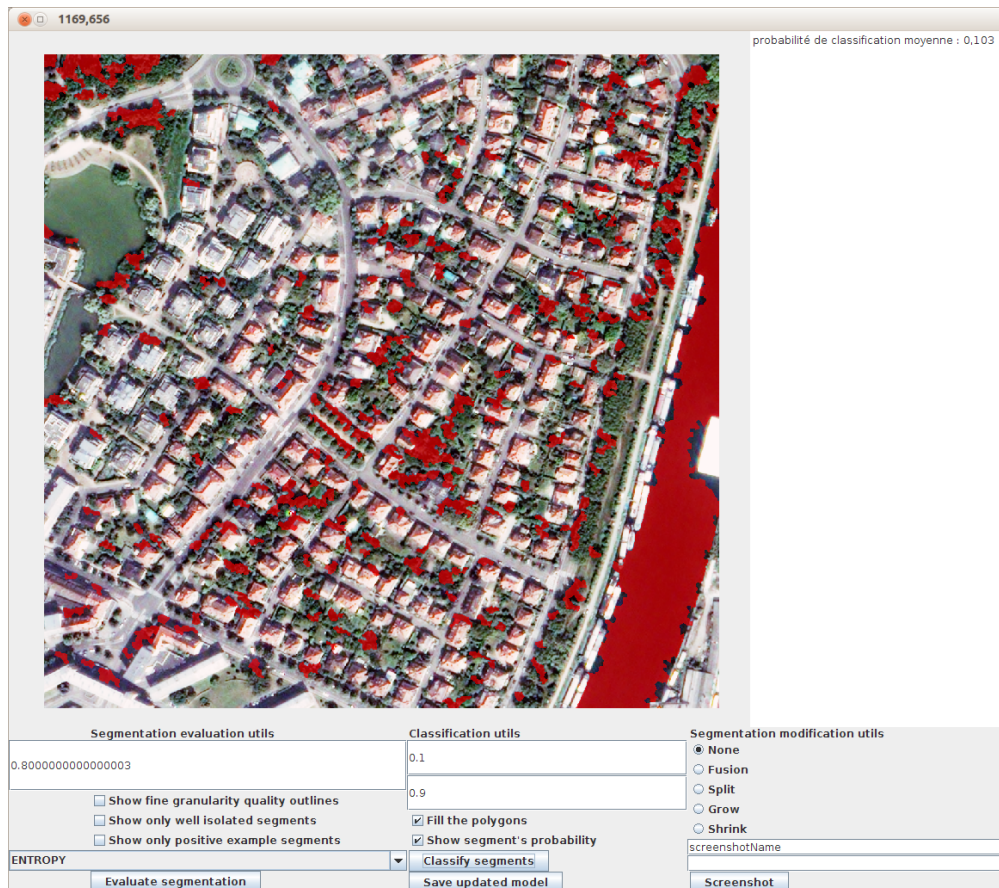


FIGURE D.2 – Visualisation du résultat d'une classification.

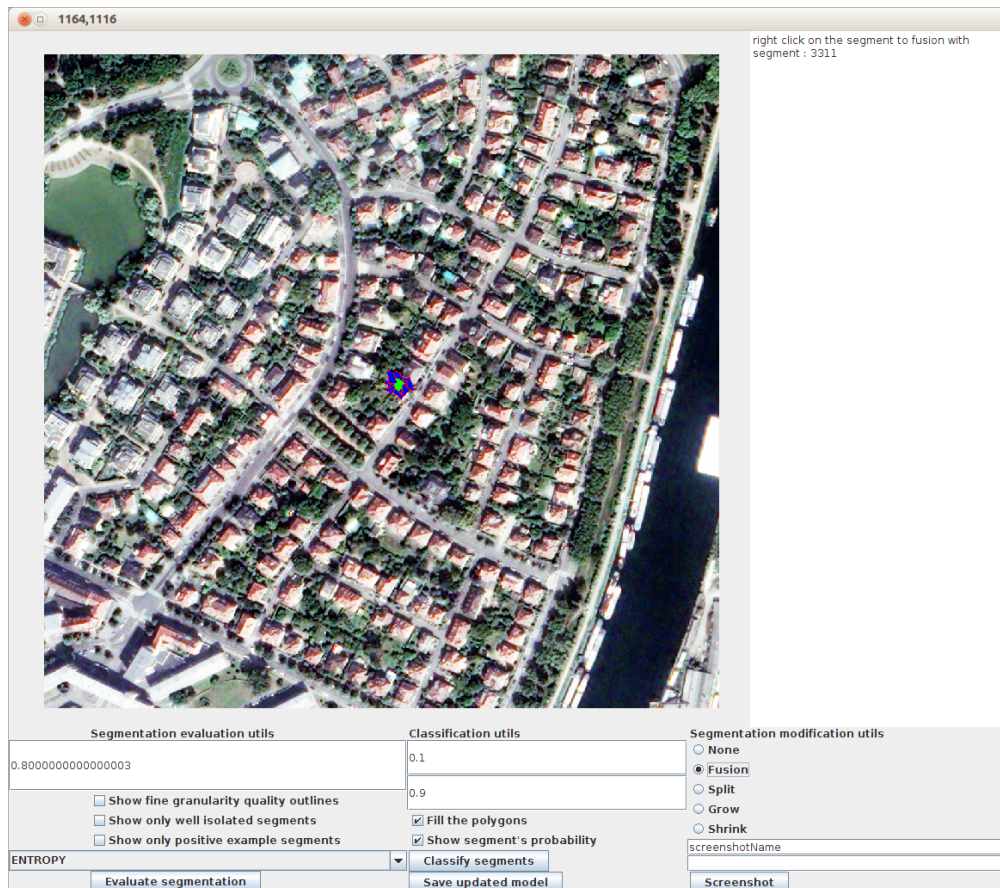


FIGURE D.3 – Visualisation de l'opérateur de fusion d'un segment avec l'un de ses voisins.

Approche collaborative et qualité des données et des connaissances en analyse multi-paradigme d'images de télédétection

Résumé

L'interprétation automatique d'images de télédétection à très haute résolution spatiale est une tâche complexe mais nécessaire. Les méthodes basées objet sont couramment employées pour traiter ce type d'images. Elles consistent à construire les objets d'intérêt au moyen d'une étape de segmentation puis à les classifier en utilisant des méthodes de fouille de données.

La majorité des travaux entrepris dans ce domaine considèrent la segmentation et la classification de manière indépendante. Or, ces deux étapes cruciales du processus sont fortement liées. Dans cette thèse, nous proposons deux approches différentes basées sur la qualité des données et des connaissances, pour initialiser, guider et évaluer un processus collaboratif de manière objective :

1. Une première approche basée sur une stratégie d'extraction mono-classe qui permet de se focaliser sur les propriétés particulières d'une classe donnée afin de mieux labelliser les objets de cette classe par rapport au reste.
2. Une deuxième approche multi-classe offrant deux stratégies différentes d'agrégation d'extracteurs mono-classes qui permet l'obtention d'une image entièrement labellisée de manière automatique.

Mots clés : approches collaboratives, segmentation, classification, télédétection

Abstract

Automatic interpretation of very high spatial resolution remotely sensed images is a complex but necessary task. Object-based image analysis approaches are commonly used to deal with this kind of images. They consist in applying an image segmentation algorithm in order to construct the objects of interest, and then classifying them using data-mining methods.

Most of the existing work in this domain consider the segmentation and the classification independently. However, these two crucial steps are closely related. In this thesis, we propose two different approaches which are based on data and knowledge quality in order to initialize, guide, and evaluate a segmentation and classification collaborative process.

1. The first approach is based on a mono-class extraction strategy allowing us to focus on the particular properties of a given thematic class in order to accurately label the objects of this class.
2. The second approach deals with multi-class extraction and offers two strategies to aggregate several mono-class extractors to get a final and completely labelled image.

Keywords: collaborative approaches, segmentation, classification, remote sensing