Représentation d'images hiérarchique multi-critère

Jimmy Francky Randrianasoa

UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

ÉCOLE DOCTORALE SCIENCES TECHNOLOGIE SANTE (547)

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Discipline : ***INFORMATIQUE***

Présentée et soutenue publiquement par

Tianatahina Jimmy Francky RANDRIANASOA

Le 8 décembre 2017

Représentation d'images hiérarchique multicritère

Thèse dirigée par **Nicolas PASSAT**

JURY

M. Jocelyn CHANUSSOT,	Professeur,	Grenoble INP,	**Président**
M. Nicolas PASSAT,	Professeur,	Université de Reims Champagne-Ardenne,	**Directeur de thèse**
M. Sébastien LEFEVRE,	Professeur,	Université de Bretagne-Sud,	**Rapporteur**
M. Philippe SALEMBIER,	Professeur,	Université Polytechnique de Catalogne,	**Rapporteur**
M. Pierre GANÇARSKI,	Professeur,	Université de Strasbourg,	**Examinateur**
Mme Yukiko KENMOCHI,	Chargée de recherche CNRS,	Université Paris-Est Marne-la-Vallée,	**Examinateur**
M. Éric DESJARDIN,	Maître de conférences,	Université de Reims Champagne-Ardenne,	**Examinateur**
M. Camille KURTZ,	Maître de conférences,	Université Paris-Descartes,	**Examinateur**

Acknowledgments

I would like to express my thanks and my gratitude to my supervisors Nicolas Passat, Éric Desjardin and Camille Kurtz for their belief in me, their encouragements and their advises while I was going through difficult moments. I fully appreciate their teachings, their scientific rigour, their commitments, their humanity and their solidarity. I really have a great chance to have them as guides bringing me at the end of this long journey.

Thanks to the Université de Reims Champagne-Ardenne (URCA), the doctoral school (ED-STS), the laboratory Centre de Recherche en STIC (CReSTIC EA 3804) and all their authorities.

Many thanks to the members of the previous team SIC and the new team RVM of the CReSTIC laboratory. It has been a pleasure to work in a such both professional and convivial mood.

I would like also to thank the secretary of the CReSTIC laboratory that helped a lot for administration matters.

Special thanks to my family for the great moral support they provided.

Thanks also to the jury members who accepted to evaluate this thesis.

This work was partially supported by the French *Agence Nationale de la Recherche* under Grant ANR-12-MONU-0001.

Représentations d'images hiérarchique multi-critère

La segmentation est une tâche cruciale en analyse d'images. L'évolution des capteurs d'acquisition induit de nouvelles images de résolution élevée, contenant des objets hétérogènes. Il est aussi devenu courant d'obtenir des images d'une même scène à partir de plusieurs sources. Ceci rend difficile l'utilisation des méthodes de segmentation classiques. Les approches de segmentation hiérarchiques fournissent des solutions potentielles à ce problème. Ainsi, l'Arbre Binaire de Partitions (BPT) est une structure de données représentant le contenu d'une image à différentes échelles. Sa construction est généralement mono-critère (i.e. une image, une métrique) et fusionne progressivement des régions connexes similaires. Cependant, la métrique doit être définie a priori par l'utilisateur, et la gestion de plusieurs images se fait en regroupant de multiples informations issues de plusieurs bandes spectrales dans une seule métrique. Notre première contribution est une approche pour la construction multi-critère d'un BPT. Elle établit un consensus entre plusieurs métriques, permettant d'obtenir un espace de segmentation hiérarchique unifié. Par ailleurs, peu de travaux se sont intéressés à l'évaluation de ces structures hiérarchiques. Notre seconde contribution est une approche évaluant la qualité des BPTs en se basant sur l'analyse intrinsèque et extrinsèque, suivant des exemples issus de vérités-terrains. Nous discutons de l'utilité de cette approche pour l'évaluation d'un BPT donné mais aussi de la détermination de la combinaison de paramètres adéquats pour une application précise. Des expérimentations sur des images satellitaires mettent en évidence la pertinence de ces approches en segmentation d'images.

Hierarchical multi-feature image representation

Segmentation is a crucial task in image analysis. Novel acquisition devices bring new images with higher resolutions, containing more heterogeneous objects. It becomes also easier to get many images of an area from different sources. This phenomenon is encountered in many domains (e.g. remote sensing, medical imaging) making difficult the use of classical image segmentation methods. Hierarchical segmentation approaches provide solutions to such issues. Particularly, the Binary Partition Tree (BPT) is a hierarchical data-structure modeling an image content at different scales. It is built in a mono-feature way (i.e. one image, one metric) by merging progressively similar connected regions. However, the metric has to be carefully thought by the user and the handling of several images is generally dealt with by gathering multiple information provided by various spectral bands into a single metric. Our first contribution is a generalized framework for the BPT construction in a multi-feature way. It relies on a strategy setting up a consensus between many metrics, allowing us to obtain a unified hierarchical segmentation space. Surprisingly, few works were devoted to the evaluation of hierarchical structures. Our second contribution is a framework for evaluating the quality of BPTs relying both on intrinsic and extrinsic quality analysis based on ground-truth examples. We also discuss about the use of this evaluation framework both for evaluating the quality of a given BPT and for determining which BPT should be built for a given application. Experiments using satellite images emphasize the relevance of the proposed frameworks in the context of image segmentation.

Keywords
Binary partition tree, hierarchical morphology, image segmentation, multi-feature, graph-based image processing, supervised quality evaluation

Mots-clés
Arbre binaire de partitions, hiérarchie morphologique, segmentation d'image, multi-critère, traitement d'image par graphe, évaluation de la qualité supervisée

Thèse réalisée au au sein du

Centre de Recherche en STIC (CReSTIC) – EA 3804

UFR Sciences Exactes et Naturelles

Moulin de la Housse

51687 REIMS cedex 2

Table des matières

Contents

Introduction

Contexte

Les images numériques sont de plus en plus utilisées dans notre vie quotidienne. Diverses tâches telles que l'extraction d'objets, le suivi d'objets, la reconnaissance de formes sont utiles pour exploiter leur contenu. L'utilisation d'images est maintenant presque un facteur clé pour de nombreuses applications dans de nombreux domaines tels que la robotique, la télédétection, la cartographie géographique, l'imagerie médicale, la vision par ordinateur, la gestion de la sécurité.

Pour exploiter pleinement le contenu de ces images, les méthodes de traitement et d'analyse d'images sont requises. En particulier, ils permettent de faciliter la tâche des experts du domaine en réduisant les tâches d'interprétation photographique manuelles / visuelles qui sont longues et fastidieuses.

Initialement, l'analyse d'image classique était principalement basée sur un paradigme de pixel jusqu'au tout début des années 90. Avec l'arrivée de nouvelles générations d'images de grandes dimensions (spectrale, spatiale, temporelle), un nouveau paradigme a émergé : il élargit le champ de l'analyse d'image en considérant les régions des images (c'est-à-dire des groupes de pixels homogènes) au lieu de considérer les pixels indépendamment. En général, les suites de traitements d'analyses d'images basées sur la notion de région commencent par l'acquisition d'images, suivi d'une tâche de prétraitement à partir de laquelle des rectifications et / ou une réorientation des images initiales sont nécessaires. Une fois que l'image est préparée, une tâche de segmentation y délimitant des objets d'intérêt peut être effectuée. Le résultat de segmentation est souvent utilisé pour une tâche de classification qui regroupe par classe les objets partageant des caractéristiques similaires (par exemple une étiquette sémantique). Le résultat de la classi-

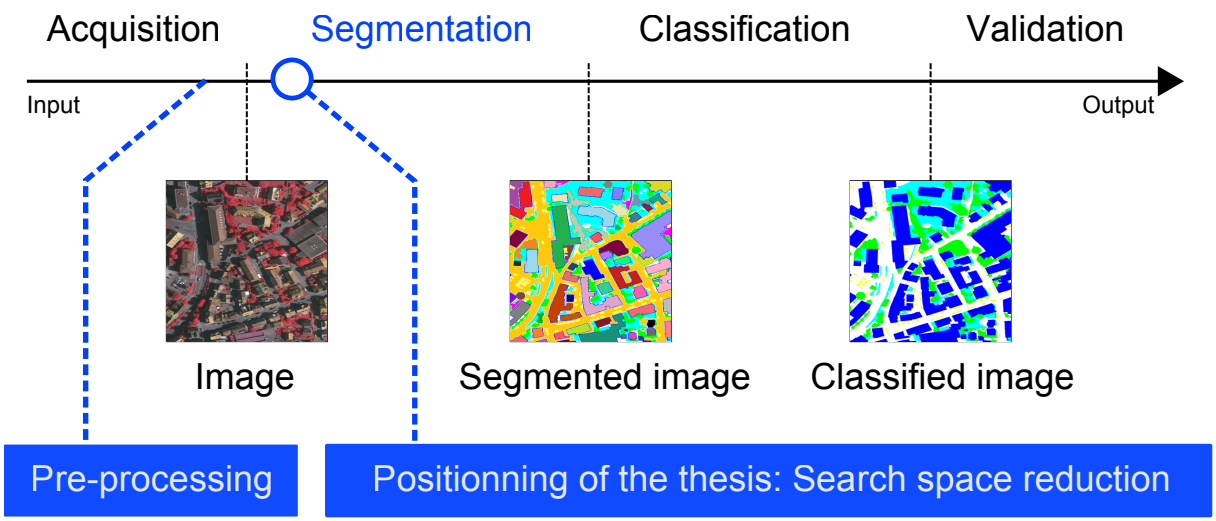


FIGURE 1 – General pipeline of region-based image processing and analysis.

fication est évalué et interprété par l'expert selon l'étude qu'il / elle veut réaliser. La Figure 1 illustre un pipeline de base de l'analyse d'une image basée sur la notion de région.

Dans le contexte du traitement et de l'analyse de l'image, la segmentation de l'image est une tâche cruciale. Elle peut être définie selon deux paradigmes : (1) délimiter un (ou plusieurs) objet (s) spécifique (s) dans une image ; et (2) subdiviser une image en zones homogènes significatives. Le deuxième paradigme, qui est étudié dans cette thèse, est particulièrement pertinent pour les applications qui nécessitent une analyse globale du contenu d'image. C'est, par exemple, le cas dans de nombreux domaines de la vision par ordinateur et en particulier dans les applications de télédétection.

Problématique

La segmentation d'image est connue comme étant un problème mal posé. De nombreux résultats peuvent être obtenus en utilisant n'importe quelle méthode de segmentation donnée et il n'est pas toujours trivial de déterminer lequel est meilleur que les autres.

Par ailleurs, l'évolution rapide des dispositifs d'acquisition impliqués dans divers domaines apporte de nouvelles générations d'images avec des résolutions plus élevées, et augmente la complexité et l'hétérogénéité des objets traités. C'est le cas par exemple de la télédétection qui est le domaine

d'application considéré dans cette thèse. En effet, la résolution spatiale des images satellitaires sont améliorées de plus en plus en partant de la faible résolution spatiale (LSR : $>$ 30 m), suivie de la résolution spatiale moyenne (MSR : 30 à 5 m), puis de la haute résolution spatiale (HSR : 5 à 1 m) et maintenant à la très haute résolution spatiale (VHSR : $<$ 1 m). Le niveau des détails fournis par ces nouvelles générations d'images pourraient être une mine d'informations apportant plus de précision. Cependant, trop d'informations peuvent devenir plus difficiles à traiter pour les méthodes classiques de segmentation d'images. En fait, la résolution spatiale grossière des images antérieures a permis l'analyse de certaines structures d'intérêt en tant qu'objets simples. Maintenant, ces objets sont perçus comme complexes car ils peuvent souvent être composés d'ensembles de pixels hétérogènes (c'est-à-dire ayant des propriétés différentes). Comment gérer les informations détaillées fournies par les nouvelles générations d'images à des fins de segmentation d'images ?

En outre, ce nouveau niveau de détails sur le contenu de l'image permet le calcul des propriétés qui sont basées sur des caractéristiques de région (radiométriques, géométriques...). Dans de nombreux travaux proposés par la communauté du traitement d'image, une caractéristique peut aider à concevoir un critère d'homogénéité relatif à la similarité ou à la dissimilarité entre les régions voisines. Ceci est important, par exemple, pour la construction de représentation hiérarchique d'image où les pixels sont regroupés en fonction de leur similarité. Habituellement, une telle construction se fait en utilisant un critère unique basé sur une seule caractéristique. Certaines contributions tirant parti de la variété des informations ont été proposées en combinant certaines fonctionnalités. Ceci est généralement fait en calculant une combinaison linéaire des caractéristiques qui détermineront toujours un critère unique. Comment en utiliser plusieurs, peut-être non comparables, dans la problématique de la segmentation d'image ?

L'acquisition d'un certain nombre d'images provenant de diverses sources est également plus facile. Il a déjà été montré que l'utilisation d'images différentes représentant une même scène apportait des informations supplémentaires utilisables. Les images sont généralement combinées pour en obtenir une seule. En fait, la construction d'une structure de données à des fins de segmentation d'image est traditionnellement réalisée en

suivant un paradigme « mono-image, mono-critère ». Comment gérer un paradigme multi-images / multicritères lors de la construction dans un espace hiérarchique de représentations d'une image ?

De nombreuses méthodes évaluant la qualité des résultats de la segmentation ont déjà été proposées dans l'état de l'art. Cependant, bien que de formidables résultats de segmentation ont été obtenus à partir de structures hiérarchiques, il n'est pas simple de déterminer quelles sont celles qui faut prendre en considération. Étonnamment, il existe très peu de travaux consacrés à l'évaluation de la qualité des structures hiérarchiques. En d'autres termes, il n'existe pas de cadre qui nous permet de déterminer si de telles structures sont définies de manière pertinente (c'est-à-dire avec la / les métrique(s) de construction adaptée (s)) par rapport à une application de segmentation d'image spécifique. Le dernier problème étudié dans cette thèse concerne cette interrogation : comment peut être évaluée la capacité par de structures hiérarchiques à fournir des résultats de segmentation pertinents.

Motivations

Les approches de segmentation hiérarchique peuvent fournir des solutions potentielles pour traiter des objets d'intérêt complexes et hétérogènes. Une solution classique consiste à définir —avant le processus de segmentation— un sous-espace de partitions candidates, à l'intérieur de l'intégralité de l'espace de recherche de toutes les partitions possibles. Une façon classique de définir de tels espaces de partition réduit est de construire des hiérarchies de partitions. Le but est de pré-calculer certaines régions imbriquées de l'image dont l'union correspond à une partition du support image. Cela évite l'explosion combinatoire induite par le grand nombre de combinaisons de sous-ensembles dans les partitions. Il conduit souvent à des structures de données basées sur des graphes organisées en arbres qui modélisent les partitions organisées hiérarchiquement en fonction de la relation de d'amélioration. Cette thèse ne se concentre pas sur l'ensemble du processus de segmentation mais sur la proposition de nouvelles représentations hiérarchique d'image comme méthodes de réduction de l'espace de recherche qui peut alors être utilisé à des fins de segmentation (voir le positionnement

de la thèse dans la Figure 1).

En particulier, l'arbre de binaire de partition (BPT) est une structure de données hiérarchique modélisant le contenu d'une image à différentes échelles. Plus précisément, c'est une représentation hiérarchique d'une image modélisée comme une arborescence, dont chaque nœud est une région connectée. Chacun de ces nœuds est soit une feuille —une région élémentaire— ou un nœud interne, modélisant l'union des régions de ses deux nœuds enfants. Sa racine est le nœud correspondant à l'ensemble support de l'image. En pratique, un BPT est construit à partir de ses feuilles, fournies par une partition initiale du support de l'image, jusqu'à sa racine, de bas en haut.

La structure du BPT permet aux utilisateurs d'explorer l'image à différentes échelles. Ça peut être utilisé pour diverses tâches telles que la segmentation, la recherche d'informations, la reconnaissance d'objets ou encore la navigation visuelle. Une coupe appliquée sur le BPT fournit une partition totale de l'image qui peut être considérée comme le résultat d'une segmentation d'image.

Comme d'autres structures hiérarchiques, le BPT était principalement conçu pour traiter une image à la fois. En outre, contrairement à la plupart d'entre elles (par exemple, les arbres à composants, les arbres de formes) intrinsèquement définies à partir du contenu de l'image, le BPT est également conçu pour intégrer une métrique extrinsèque qui est utilisée, en combinaison avec l'image, pour construire une modèle mixte image / connaissances. En d'autres termes, un BPT est généralement construit pour une image et une métrique. Le BPT a déjà démontré sa pertinence pour le traitement d'images et les tâches d'analyse, comme par exemple dans les domaines de l'analyse vidéo, de la télédétection ou de l'imagerie médicale. Cependant, comme indiqué ci-dessus, il reste principalement limité à un paradigme « une seule image, une seule métrique ». En effet, d'une part la métrique ?à savoir le critère utilisé pour décider lequel des nœuds fusionner ? est une fonction généralement scalaire, qui impose de fusionner divers éléments de connaissances d'experts (par exemple des critères d'hétérogénéité colorimétrique et géométrique). Ça signifie en particulier que cette métrique doit être soigneusement pensée au préalable par l'utilisateur, ce qui constitue une tâche complexe. D'autre part, du point de vue de l'imagerie, la

gestion de plusieurs images est généralement traitée soit en définissant une super-image en les fusionnant, soit en assemblant les multiples informations fournies par diverses bandes spectrales en une seule métrique.

Notre première motivation dans cette thèse est de définir une représentation hiérarchique multicritère d'une scène observée, impliquant une / plusieurs images et une / plusieurs métriques. Une telle représentation hiérarchique rassemble les informations pertinentes d'une manière unifiée permettant d'obtenir des résultats de segmentation d'image pertinents. Dans cette thèse, nous ne nous concentrons pas sur la manière d'obtenir des résultats de segmentation à partir des structures hiérarchiques, mais bien sur la façon de construire une structure unifiée à partir de laquelle de tels résultats de segmentation pourraient être obtenus.

Une large littérature a été consacrée à la segmentation par BPT, aux méthodes et à leurs variantes, en particulier dans le contexte difficile de la télédétection. Divers paramètres ont été étudiés (spectraux, spatiaux, géométriques, etc.) pour construire des BPT. La conception de ces métriques (choix et combinaison) influence fortement la structure des BPT obtenus (c'est-à-dire la forme de l'espace de recherche réduit qui sera utilisé lors d'une segmentation plus poussée ultérieurs, et donc la qualité du résultat de cette segmentation). Cependant, il n'existe que peu de travaux consacrés à l'évaluation et / ou la comparaison de ces structures hiérarchiques.

Notre deuxième motivation est l'évaluation de la qualité de ces structures hiérarchiques, afin de déterminer si leur utilisation est pertinente ou non pour d'autres tâches telles que la segmentation d'image ou l'extraction d'objets. Une telle évaluation de la qualité peut être utile pour comparer différentes structures par évaluation, mais peut aussi aider à améliorer leur construction par prédiction.

Comme le montre la Figure 1, cette thèse se concentre sur la réduction de l'espace de recherche pour la segmentation d'images. En particulier, notre recherche s'appuie sur les BPT dont la construction dépend fortement des connaissances antérieures de l'utilisateur.

Contributions

Les méthodes de traitement et d'analyse d'images sont développées pour aider l'utilisateur en réduisant son / ses tâches. Ces méthodes peuvent être supervisées ou non supervisées, mais l'utilisateur doit toujours être impliqué car il / elle a des connaissances utiles sur l'application. Dans cette thèse, notre objectif est d'élaborer des environnements aidant l'utilisateur à améliorer les résultats de la segmentation d'images. Deux axes de recherche basés sur la représentation hiérarchique des images sont présentés dans cette thèse. Le premier axe est un cadre généralisant la construction de BPT dans une approche multicritère. Le deuxième concerne l'évaluation de la capacité d'un BPT à fournir un résultat de segmentation pertinent.

Cadre généralisé pour la construction BPT dans une approche multicritère Le BPT est généralement construit de manière mono-métrique. En effet, un BPT peut être construit à partir d'une métrique et d'une image choisies. La métrique doit être définie par l'utilisateur via une fonction déterminée par ses connaissances préalables sur l'étude qu'il / elle veut mener. En relâchant les contraintes restrictives de ce paradigme vers une approche « multi-image, multicritère », nous pourrons apporter une certaine flexibilité et réduire la perte d'informations tout en aidant l'utilisateur dans son travail d'analyse. Dans cette thèse, notre but est d'étendre la construction BPT afin d'obtenir une structure hiérarchique unifiée rassemblant des informations pertinentes en impliquant une ou plusieurs métriques et une / plusieurs images. La façon classique d'utiliser diverses mesures dans la construction BPT est généralement réalisée en demandant à l'utilisateur de construire une nouvelle métrique qui en est une combinaison linéaire. Également, la manipulation de plusieurs images nécessite des méthodes de fusion d'images menant à une super-image qui est une combinaison de plusieurs d'entre-elles. Ces méthodes de fusion de métriques et d'image ont déjà montré leur utilité dans le contexte du traitement et de l'analyse de l'image. Cependant, le moyen de les appliquer doit être réfléchi par l'utilisateur, ce qui n'est pas une tâche facile. Effectivement, la pertinence de la super-métrique dépend de la formulation de la combinaison linéaire conçue par l'utilisateur. De plus, la perte d'information induite par la fusion de nombreuses images dépend de la méthode de fusion d'image

choisie. Pour réduire la charge sur l'utilisateur, notre objectif est de fournir un cadre de construction de BPT pouvant traiter chacune des caractéristiques et avec chacune des images. Notre approche repose sur une stratégie de mise en place d'un consensus entre les —potentiellement conflictuelles et incertaines— caractéristiques et les images, nous permettant d'obtenir un espace de segmentation hiérarchique unifié. Cette proposition fournit des alternatives à la problématique de la construction de métriques complexes relatives à plusieurs caractéristiques, éventuellement non comparables.

Environnement d'évaluation de la qualité des BPTs Malgré leur utilité dans le contexte du traitement et de l'analyse d'images, très peu de travaux ont été consacrés à l'évaluation de la capacité pour un BPT de fournir des résultats de segmentation pertinents. D'une manière classique, les résultats de segmentation obtenus à partir d'un BPT peuvent être évalués directement en utilisant une segmentation différente des méthodes d'évaluation (par exemple, à partir d'une carte de vérité-terrain). De nombreuses évaluations des méthodes de segmentation ont été proposées et ont déjà prouvé leur utilité. Cependant, évaluer le résultat d'une segmentation obtenue à partir d'un BPT ne permet pas d'accéder d'une manière générale à sa qualité. En effet, un grand nombre de résultats de segmentation peuvent être obtenus à partir d'un BPT en lui appliquant plusieurs coupes. La détermination de ces coupes n'est pas triviale et les appliquer n'est pas une tâche facile. Ainsi, une évaluation basée sur les résultats d'une segmentation issue d'un BPT ajoute un biais induit par la façon dont les coupes ont été réalisées. Une mauvaise segmentation résultant d'une mauvaise coupe ne signifie pas systématiquement que le BPT n'est pas en mesure de fournir des résultats de segmentation satisfaisants. Pour éviter un tel biais, notre objectif est de fournir un environnement permettant d'évaluer directement la qualité du BPT. La capacité d'un BPT à fournir une segmentation acceptable ne dépend pas seulement de l'application, mais aussi de l'attente de l'utilisateur. En effet, l'utilisateur connaît mieux l'application et a une idée des résultats qu'il / elle veut obtenir. Notre environnement d'évaluation est ainsi supervisé car nous avons besoin de l'utilisateur pour définir ses attentes. L'idée principale est de calculer un score de qualité d'un BPT en évaluant sa structure et sa capacité à contenir dans sa hiérarchie des

objets intéressants. En outre, un tel environnement d'évaluation est utile pour comparer notre BPT multicritère (MBPT) avec sa forme classique. Dans un but d'amélioration de notre cadre de construction de BPT multicritère, impliquer notre approche d'évaluation dans le processus pourrait être une solution intéressante, aidant l'utilisateur à choisir judicieusement les propriétés et la stratégie consensuelle.

Structure du manuscrit

Ce manuscrit est divisé en deux parties principales, plus trois annexes, brièvement décrites ci-après.

Dans la partie I, nous introduisons d'abord le domaine de la segmentation d'image hiérarchique. En particulier, nous présentons le BPT de différents points de vue et discutons de ses limites dans le contexte du traitement et de l'analyse d'images. Afin d'apporter quelques solutions aux problèmes précédemment présentés, nous proposons ensuite une nouvelle généralisation de la construction BPT dans une approche multicritères. Dans un but d'optimisation de l'algorithme proposé, nous discutons également de deux heuristiques basées sur une approche séquentielle et une approche distribuée. Certaines expériences dans le cadre d'images de télédétection soulignent qualitativement la pertinence des environnements proposés dans le contexte de la segmentation d'image basée sur des BPT.

Dans la partie II, nous présentons brièvement quelques méthodes d'évaluation de la segmentation d'images et analysons leurs limites pour évaluer la qualité d'un BPT. Nous proposons ensuite un nouvel environnement d'évaluation estimant la capacité d'un BPT à fournir de résultats pertinents de segmentation pour l'utilisateur. Certaines expériences sur les images de télédétection montrent l'utilité de notre approche et une validation quantitative de notre construction BPT multicritère est présentée.

Dans le cadre de ce travail, nous présentons dans l'annexe A le développement d'une bibliothèque en java mettre en œuvre nos méthodes proposées. L'annexe B décrit les principales structures de données impliquées dans cette implémentation. L'annexe C détaille les publications scientifiques et les communications liées à cette thèse.

Introduction

Context

Digital images are becoming more and more used in our daily life. Various tasks such as object extraction, object tracking, pattern recognition are useful to exploit their contents. The use of images is now almost a key factor for different applications in many domains such as robotics, remote sensing, geographical mapping, medical imaging, computer vision, security monitoring.

To fully exploit the content of these images, image processing and analysis methods are required. In particular, they enable to ease the task of the domain experts by reducing manual / visual photo interpretation tasks which are tedious and time consuming.

Initially, classical image analysis was mainly based on a pixel paradigm until the very beginning of the 90's. With the arrival of new generations of images having huge dimensions (spectral, spatial, temporal), a new paradigm has emerged : it lightens the image analysis process by considering image regions (i.e., groups of homogeneous pixels) instead of considering pixels independently. In general, the region-based image analysis pipelines start from the image acquisition, followed by a pre-processing task from which some rectifications and / or reorientation of the initial image(s) are required. Once the image is prepared, a segmentation task can be carried out, delineating some objects of interest from the image. The segmentation result is often used for a classification task by grouping by class the objects sharing similar characteristics (e.g., semantic label). The classification result is assessed and interpreted by the expert according to the study he / she wants to proceed. Figure 2 illustrates a basic pipeline of a region-based image analysis.

In the context of image processing and analysis, image segmentation is

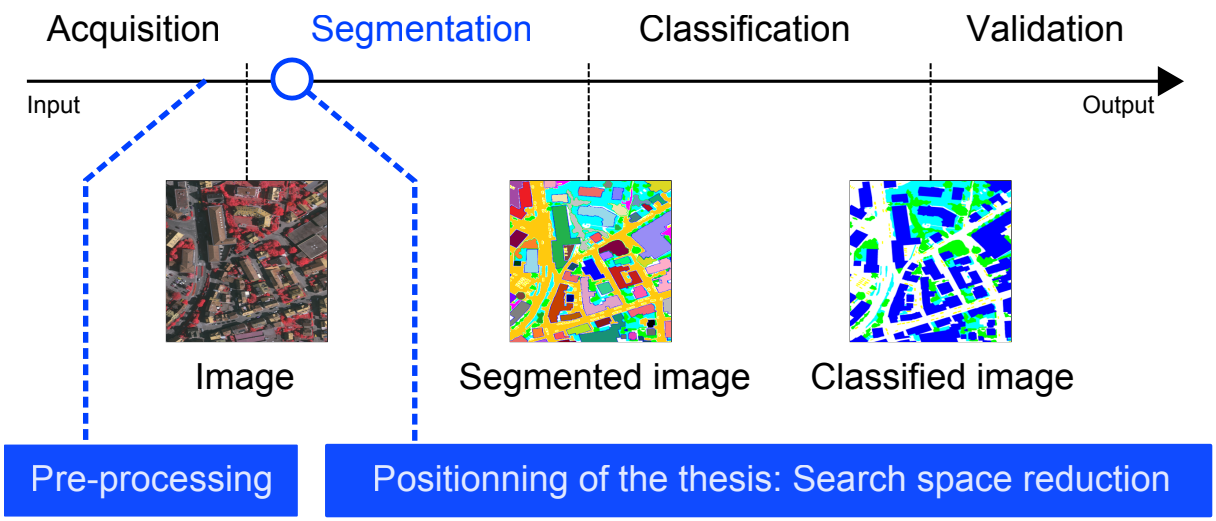


FIGURE 2 – General pipeline of region-based image processing and analysis.

a crucial task. It can be defined under two paradigms : (1) delineating one (or many) specific object(s) in an image ; and (2) subdividing an image into meaningful homogeneous zones. The second paradigm, which is studied in this thesis, is particularly relevant for applications that require a global analysis of an image content. This is, for instance, the case in many computer vision areas and in particular in remote sensing applications.

Problematic

Image segmentation is known as an *ill-posed* problem. Several results can be obtained by using any given segmentation method and it is not always trivial to determine which one is better than the others.

Besides, the rapid evolution of the acquisition devices involved in different thematic domains brings new generations of images with higher resolutions, and increasing complex and heterogeneous objects. This is for instance the case in remote sensing which is the application domain considered in this thesis. Indeed, the spatial resolution of satellite images has been improved increasingly from the low spatial resolution (LSR : $>$ 30 m) through the medium spatial resolution (MSR : 30 to 5 m), the high spatial resolution (HSR : 5 to 1 m) and now to the very high spatial resolution (VHSR : $<$ 1 m). The level of details provided by such new generations of images could be a wealth of information bringing more precisions. However, too much information can become more difficult, for the classical

image segmentation methods, to be handled. As a matter of fact, the coarse spatial resolution of the previous images allowed directly the analysis of some structures of interest that were considered as simple objects. Now, these objects are considered being complex as they may often be composed by sets of heterogeneous pixels (i.e. having different properties). How to handle detailed information provided by such new generations of images for image segmentation purposes ?

In addition, this new level of details about the image content allows for the computation of features that are based on the characteristics (radiometrical, geometrical...) of the regions. In many works proposed by the image processing community, one feature can help to design an homogeneity criterion which relates to the similarity or the dissimilarity between neighbouring regions. This is important, for instance, for the construction of hierarchical image representation where pixels are grouped according to their similarity. Usually, such construction is done by using a unique criterion based on a single feature. Some contributions taking advantage of the variety of information have been proposed by combining some features. This is generally done by computing a linear combination of the features that will still determine a unique criterion. How to use various, possibly non-comparable, features in the problematic of image segmentation ?

The acquisition of a number of images from various sources is also now easier. It has already been shown that the use of different images representing a same scene would bring additional information on the study. The images are usually combined to get one unique image. In fact, the construction of a data-structure for image segmentation purpose is traditionally done by following a "mono-image, mono-feature" paradigm. How to manage a multi-image / multi-feature paradigm during the construction of a hierarchical image representation space ?

Many methods evaluating the quality of the results of the segmentation have been already proposed in the state of the art. However, since tremendous segmentation results can be obtained from one hierarchical structure, it is not straightforward to determine which ones to consider. Surprisingly, there exist very few works devoted to evaluate the quality of hierarchical structures. In other words, there does not exist any framework allowing us to determine if such structures are defined in a relevant way (i.e., with

the adapted construction metric(s)) with respect to a given segmentation application on a specific image. The last problem studied in this thesis can be how to evaluate the capacity of hierarchical structures to provide relevant segmentation results.

Motivations

Hierarchical segmentation approaches can provide potential solutions for dealing with complex and heterogeneous objects of interest. A classical solution consists of defining —before the segmentation process— a subspace of candidate partitions, within the whole search space of all the possible partitions. A popular way of defining such reduced partition spaces is to build hierarchies of partitions. The purpose is to pre-compute some nested regions of the image whose union corresponds to a partition of the image support. This avoids the combinatorial explosion induced by the huge number of combinations of subsets within partitions. It often leads to graph-based data-structures organized as trees that model partitions hierarchically organized with respect to the refinement relation. This thesis does not focus on the whole segmentation process but on the proposition of new hierarchical image representation as search space reduction that can be further used for segmentation purposes (see the positionning of the thesis in Figure 2).

Particularly, the Binary Partition Tree (BPT) is a hierarchical data-structure modeling an image content at different scales. Precisely, it is a hierarchical representation of an image modelled as a tree structure, whose each node is a connected region. Each of these nodes is either a leaf —an elementary region— or an internal node, modelling the union of the regions of its two children nodes. Its root is the node corresponding to the entire support of the image. Practically, a BPT is built from its leaves, provided by an initial partition of the image support, to its root, in a bottom-up fashion.

The BPT structure allows users to explore the image at different scales. It can be used for various tasks such as segmentation, information retrieval, object recognition and visual browsing. A cut applied on the BPT provides a total partition of the image that can be considered as an image

segmentation result.

Like other hierarchical structures, the BPT was mainly designed to process one image at a time. Furthermore, in contrast with most of them (e.g., component-trees, trees of shapes) that are intrinsically defined from the image content, the BPT is also designed to embed an extrinsic metric that is used, together with the image, to build a mixed image / knowledge model. In other words, a BPT is generally built for one image and one metric.

The BPT has already demonstrated its relevance for challenging image processing and analysis tasks, e.g. in the domains of video analysis, remote sensing or medical imaging. However, as stated above, it remains mostly limited to a *one image, one metric* paradigm. Indeed, on the one hand, the metric —namely the merging criterion used to decide which nodes to fuse— is a generally scalar function, that imposes to fuse various elements of expert knowledge (e.g. colorimetric and geometric heterogeneity criteria). This means in particular that this metric has to be carefully thought beforehand by the user, which constitutes a complex task. On the other hand, from the imaging point of view, the handling of several images is generally dealt with either by defining a super-image by merging them, or by gathering the multiple information provided by various spectral bands into a single metric.

Our first motivation in this thesis is to define a multi-feature hierarchical representation of an observed scene, involving one / many images and one / many metrics. Such hierarchical representation gathers pertinent information in a unified way helping to obtain relevant image segmentation results. In this thesis, we are not focusing on the way of getting segmentation results from the hierarchical structures, but on the way to build a unified structure from which such segmentation results could be obtained.

A large literature has been devoted to BPT-based segmentation, methods and their variants, in particular in the challenging context of remote sensing. Various metrics were investigated (spectral, spatial, geometric, etc.) for building BPTs. The design of these metrics (choice and combination) strongly influences the resulting BPTs (i.e., the reduced search space for further segmentation, and thus the quality of the subsequent segmentation result). However, there exists only few works devoted to the evaluation

and / or the comparison of such hierarchical structures.

Our second motivation is the evaluation of the quality of these hierarchical structures, in order to determine if their use is relevant or not for other tasks such as image segmentation, object extraction. Such quality assessment can be useful to compare different structures by evaluation, but can also help to improve their construction by prediction.

As shown in Figure 2, this thesis focuses on the search space reduction for image segmentation. In particular, our research is relying on the BPTs whose construction depends highly on prior knowledge from the user.

Contributions

Image processing and analysis methods are developed to help the user by reducing his / her tasks. Those methods can be supervised or unsupervised but the user has always to be involved since he / she has useful knowledge about the application. In this thesis, our purpose is to elaborate some frameworks helping the user to improve image segmentation results. Two research axes based on hierarchical image representation are presented in this thesis. The first axis is a framework generalizing the BPT construction in a multi-feature way. The second axis is about the evaluation of the ability of a BPT to provide a pertinent segmentation result.

Generalized framework for the BPT construction in a multi-feature fashion The BPT is generally built in a mono-metric way. Indeed, one BPT can be built from one chosen metric and one image. The metric has to be defined by the user via a feature determined by his / her prior knowledge about the study he / she wants to lead. Relaxing this paradigm into a "multi-image, multi-feature" one would bring some flexibility and could reduce the loss of information while helping the user in his / her study. In this thesis, our purpose is to extend the BPT construction in order to get one unified hierarchical structure gathering pertinent information by involving one / many metrics and one / many images. The classical way to use various metrics in the BPT construction is usually done by asking the user to build a super-metric that is a linear combination of many. Likewise, the handling of several images requires image fusion methods leading to

a super-image which is a combination of many. These metric and image fusion methods have already shown their utility in the context of image processing and analysis. However, the way to apply them has to be thought carefully by the user, which is not an easy task. Indeed, the pertinence of the super-metric depends on the formulation of the linear combination by the user. In addition, the loss of information induced by the fusion of many images depends on the chosen image fusion method. To reduce the charge on the user, our aim is to provide a BPT construction framework dealing with each feature and with each image. Our approach relies on a strategy setting up a consensus between the —potentially conflicting and uncertain— features and the images, allowing us to obtain a unified hierarchical segmentation space. This proposition provides alternatives to the complex metric construction issue from several, possibly non-comparable, features.

Framework evaluating the quality of BPTs Despite their usefulness in the context of image processing and analysis, very few works were devoted to evaluate the ability of the BPT to provide pertinent segmentation results. In a classical way, the segmentation results obtained from a BPT can be directly assessed by using different segmentation evaluation methods (e.g., from a ground-truth map). Many evaluation of segmentation methods have been proposed and have already proven their utility. However, evaluating one segmentation result obtained from one BPT could not essentially assess its general quality. Indeed, a huge number of segmentation results can be obtained from one BPT by applying several cuts on it. The cuts are not trivial to determine and applying them is not an easy task. Hence, an evaluation analysis based on the segmentation results from a BPT adds some bias induced by the way the cuts were applied. A bad segmentation result from a wrong cut does not mean systematically that the BPT is not able to provide satisfactory segmentation results. To avoid such bias, our aim is to provide an evaluation framework assessing directly the quality of the BPT without finding segmentation results. The capacity of a BPT to provide an acceptable segmentation does not depend only on the application, but also on the expectation of the user. Indeed, the user knows better the application and has an idea about the results he / she wants to get. Our

evaluation framework is then supervised since we need the user to define his / her expectations. The main idea is to compute a quality score of a BPT by assessing its structure and its ability to contain in its hierarchy some interesting objects. Such evaluation framework is, in addition, helpful for comparing our multi-featured BPT (MBPT) with the classical one. For some purpose of improvements on our multi-feature BPT construction framework, involving our evaluation approach in the process could be an interesting solution, helping the user to chose wisely the features and the consensual strategy.

Structure of the manuscript

This manuscript is divided in two principal parts, plus three appendices, briefly described below.

In Part I, we introduce first the hierarchical image segmentation domain. In particular, we present the BPT in various points of view and discuss about its limits in the context of image processing and analysis. In order to bring some solutions to the issues previously presented, we then propose a new generalization of the BPT construction in a multi-feature way. For an optimization purpose of the proposed algorithm, we also discuss about two heuristics based on a sequential and a distributed approach. Some experiments on remote sensing images emphasize qualitatively the relevance of the proposed frameworks in the context of BPT-based image segmentation.

In Part II, we first introduce briefly some image segmentation evaluation methods and discuss about their limits for assessing the quality of a BPT. We then propose a new evaluation framework estimating the capacity of a BPT to provide pertinent segmentation results for the user. Some experiments on remote sensing images show the usefulness of our approach, and a quantitative validation of our multi-featured BPT construction is presented.

As a part of the work, we present in the Appendix A the development of a library implementing our proposed methods. Appendix B describes the main data-structures involved in this implementation. Appendix C details the scientific publications and communications related to this thesis.

Première partie

Multi-criteria Binary Partition Tree construction

Dans cette partie, nous introduisons premièrement le paradigme de segmentation d'image hiérarchique par graphe. Ensuite, nous présentons l'arbre binaire de partitions (BPT) et discutons des limites de sa construction. Afin de franchir ces limites, nous présentons le BPT d'un point de vue graphe et proposons une nouvelle approche pour la construction de BPT multicritère. Notre approche permet l'obtention d'un unique espace de recherche unifié en utilisant dans sa construction un / plusieurs attribut(s) et une / plusieurs image(s). Afin de réduire les coûts en termes de temps de calcul de l'algorithme que nous proposons, nous présentons deux heuristiques d'optimisation comme perspectives. Des expériences utilisant des images de télédétection montrent l'utilité de notre approche dans le contexte de la segmentation d'images.

In this part, we first introduce the graph-based and the hierarchical image segmentation paradigms. Then, we present the Binary Partition Tree (BPT) and discuss the limits of its classical construction. To go beyond those limits, we present the BPT from a graph-based point of view and propose a new multi-criteria BPT construction framework involving one / many feature(s) and one / many image(s) that allows us to obtain a unified hierarchical search space for image segmentation. In order to reduce the cost of our proposed algorithm, we propose two optimization heuristics as perspectives. Experiments using remote sensing images emphasize the usefulness of our framework for image segmentation purpose.

Chapter 1

Hierarchical image representation

Contents

In this chapter, we first describe the interest of the reduction of the search space for image segmentation purpose (Section 1.1). We start by introducing some notions on graph-based and hierarchical image segmentation (Section 1.2). We then introduce the Binary Partition Tree (BPT), its specificity and its limitations (Section 1.3).

1.1 Search space reduction

From a data point of view, the segmentation process can be seen as an information retrieval procedure in a defined search space. In particular, this search space corresponds to one (or potentially many) image(s) containing the data. As the technologies of the image acquisition devices are progressing rapidly, it is now easier and common to get a new generation of images having huge dimensions (e.g., it is common to encounter images of

$10\,000 \times 10\,000$ pixels in the domain of remote sensing). The high dimensions of the images bring more information and wider search spaces that make some classical segmentation processes costly. An alternative way to facilitate the image partitioning in such huge search spaces is to reduce their size while preserving as much pertinent information as possible. The pertinence of the information depends on the application in which the reduced search space will be involved. Despite the obvious information loss of such approach, the search space reduction has already shown its utility for some purposes as the image segmentation. Indeed, if the reduction of the search space is well done, the loss of information is minimized, so the result obtained from its manipulations could be comparable to that obtained from the initial search space.

The next section (Section 1.2) introduce a common way of reducing the search space by using graphs and by building hierarchical structures representing image contents.

1.2 Graph-based and hierarchical image segmentation

Image processing and analysis problems, and in particular segmentation, are often considered in a discrete way via some concepts from the graph theory. Practically, image points (i.e., pixels, voxels) are considered as the vertices of a graph, while the spatial / neighbouring relations between them are modelled by graph edges. This paradigm, democratized since the early 70's [1], led to the development of a wide range of segmentation approaches, based on basic graph manipulations.

In this context, image segmentation could be viewed as a partial (e.g., region growing [2]), or total partitioning problem with monotonic (e.g., watersheds [3]) or non-monotonic transformations (e.g., split and merge [4]). Some of these approaches led to the development of optimization schemes (e.g., graph-cuts [5, 6, 7], random walks [8], power watersheds [9]). In the framework of mathematical morphology, these graph-based approaches gave rise to the notion of connected operators [10] that are used by region-based morphological image processing techniques [11]. A work for a comparative review of the computation algorithms of some structures

related to the connected operators has been investigated in [12].

Graph-based segmentation allows us to obtain one segmentation result from a given image. In order to tackle the ill-posed problem of segmentation, hierarchical approaches were developed to compute families of nested partitions, providing adapted solutions at different scales. These notions of hierarchies take their origin in image models initially devoted to optimize the access and space cost of the carried information (e.g., octrees [13]). These regular models were progressively shifted toward image / content-guided, irregular hierarchies [14].

From this point of view, several hierarchical image models were developed, mainly in the framework of mathematical morphology. The most popular are the component-tree [15], the tree of shape [16, 17], the hierarchical watersheds [18], the hyperconnected component-trees [19], the α-tree [20, 21] and the binary partition trees [22] (see Section 1.3). Since they provide a space of potential segmentations, instead of a single result, these hierarchical models were progressively involved in attribute-based [23, 24, 25] or optimization schemes [26, 27, 28] for segmentation purpose. Solutions for object extraction from hierarchical representation of an image were also proposed for specific domains such as remote sensing [29]. Most of the hierarchical structures proposed in the literature are models intrinsically deriving from the image signal. For instance, component-trees represent the inclusion of the successive level-sets; trees of shapes represent the image level-lines; while hierarchical watersheds rely on saliency measures similar to gradients. These models strongly rely on a functional, often derived from the image intensity. Hence, one unique structure of each of these models can be obtained from one image (see left part of Figure 1.1). In some applicative contexts, the regions contained in such trees may not correspond to objects of interest in image content, in particular when dealing with complex images. Compared to the image signal dependent hierarchical structures, the binary partition tree model relies on a measure of dissimilarity / similarity between adjacent elements of the graph that is used for the construction of a hierarchy.

Based on these image models, generally designed as trees, further developments were proposed to allow for a better flexibility in image and parameter handling. The case of multiband images was considered, leading

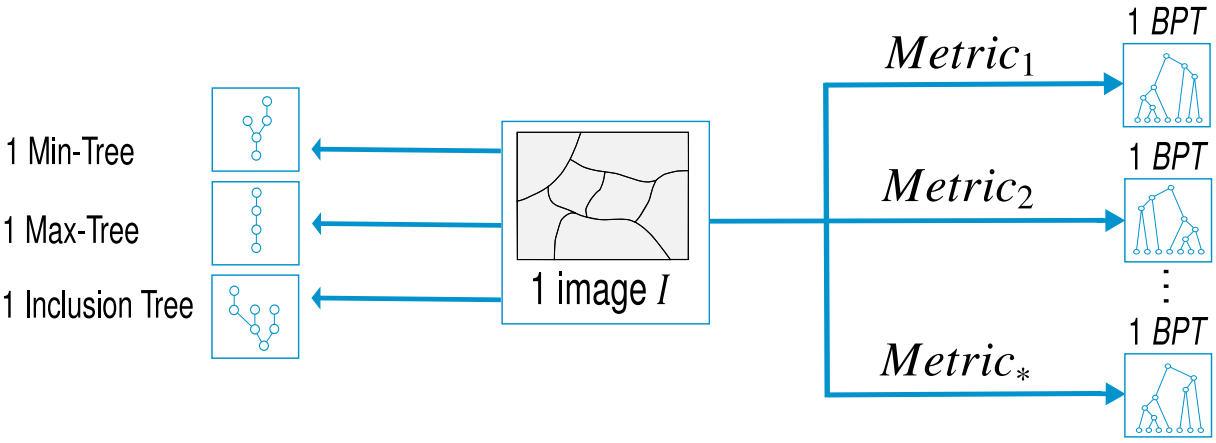


Figure 1.1: Comparison between the component-trees and the BPT. From one image equiped with a canonical order relation, only one min-tree, one max-tree and one inclusion tree can be obtained. However, from one image, several BPTs can be obtained according to the metric used.

to data-structures such as component-graphs [30], multivalued component-trees [31] or multivariate trees of shapes [32]. Topological handling was also investigated, by allowing connectivity hierarchies in component-hypertrees [33], or dealing with asymmetric hierarchies [34] allowing for non-directed graph as image models. The way to embed semantic information as image values was also pioneered via the notion of shaping [35, 36].

In the next section (Section 1.3), we present the Binary Partition Tree (BPT), which is the hierarchical model studied in the sequel of this manuscript.

1.3 Binary Partition Trees

Image

An image I is defined on a nonempty, finite set of points Ω. In general, such points are pixels (resp. voxels) in 2D (resp. 3D) images. The structural properties between these points have to be explicitly considered.

Practically, the relevant information is the existence of a neighbourhood relation between two points. To model the fact that two points x and y of Ω are neighbours, let $\frown$ be an adjacency (i.e., irreflexive, symmetric) binary relation on Ω. In other words, $\mathfrak{G}_\Omega = (\Omega, \frown)$ is a graph that models the structure of the image space.

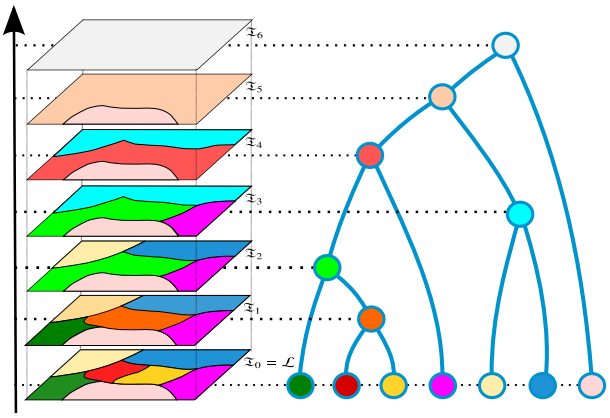


Figure 1.2: Hierarchical representation of an image I by a Binary Partition Tree (BPT) $\mathfrak{T}$.

For any partition $\mathcal{P}$ of Ω, we define an adjacency inherited from that of Ω. We say that two disjoint sets $N_1, N_2 \in \mathcal{P}$ are adjacent if there exist $x_1 \in N_1$ and $x_2 \in N_2$ such that (x_1, x_2) is an edge of $\frown$, i.e., x_1 and x_2 are adjacent in $\mathfrak{G}_\Omega = (\Omega, \frown)$. This new adjacency relation $\frown_\mathcal{P}$ is also irreflexive and symmetric.

The image I takes its values in a space $\mathbb{V}$. There is no actual requirement with regard to this value space. In general, $\mathbb{V}$ is a scalar space (e.g. a subset of $\mathbb{R}$ or $\mathbb{Z}$) in the case of monovalued images, or a Cartesian set of such spaces in the case of multi- / hyperspectral images. Practically, an image can then be formalized as a function $I : \Omega \to \mathbb{V}$ that associates to each point x of the finite set Ω a value $I(x)$ of the set $\mathbb{V}$.

Definition

Considering an image I defined on a given finite support (i.e., a set of pixels Ω), a Binary Partition Tree [22] (BPT) is a hierarchical representation of an image I, organized as a binary tree (see Figure 1.2), noted $\mathfrak{T}$. Each node N of a BPT is a connected region corresponding to a subset of Ω (i.e., $N \subseteq \Omega$). A node is either a leaf of the tree (i.e., an elementary region, possibly a single pixel) or an internal node (i.e., the union of two nodes modelling two adjacent regions). The root of the BPT is the node corresponding to the entire support Ω of the image. Practically, a BPT is built from its leaves (provided by an initial partition of Ω) to its root, in a bottom-up fashion, by iteratively choosing and merging two adjacent regions which optimize a criterion (called a metric in the remainder of this

manuscript) expressing their likeness.

Image / knowledge model

The BPT relies on a mixed image / knowledge model. It can then easily embed expert-defined knowledge, in addition to image-based information. From a structural point of view, BPTs present similar properties with Binary Space Partition Trees [37], designed to efficiently model an image space, mainly in computer graphics. Indeed, BPTs provide hierarchies of nested total partitions of an image.

From an algorithmic point of view, a BPT is built by progressively merging elementary image segments (e.g., flat zones), based on a given metric. In particular, this metric requires an a priori knowledge from the expert, its choice is crucial, as it has a direct impact on the adequacy of the built BPT with the targeted application.

In comparison with the component-trees, several BPTs can be obtained from one image, depending on the metrics chosen by the expert (see right part of Figure 1.1).

This gave rise to several works, ranging from theoretical contributions [38] to experimental assessments [39]. The basic criteria (i.e., features) used in most of image segmentation approaches are generally radiometric or geometric region similarities (or their fusion into a single metric). Various metrics were investigated for building BPTs, as their design (choice and combination) strongly influences the resulting BPTs (i.e., the reduced search space for further segmentation) and thus the quality of the subsequent segmentation result.

Except for contributions in the field of object recognition [40, 41, 42], BPTs were mainly involved in segmentation / classification cases. More precisely, the wider application field of BPTs is remote sensing [43]. In this context, BPTs were involved in multiresolution / multiscale image segmentation and classification [44, 45]; coupled optical / LIDAR data analysis [46]; hyperspectral images [47, 48]; polarimetric SAR [49, 50]; mixed SAR / hyperspectral images [51]; or multi-temporal SAR image analysis [52].

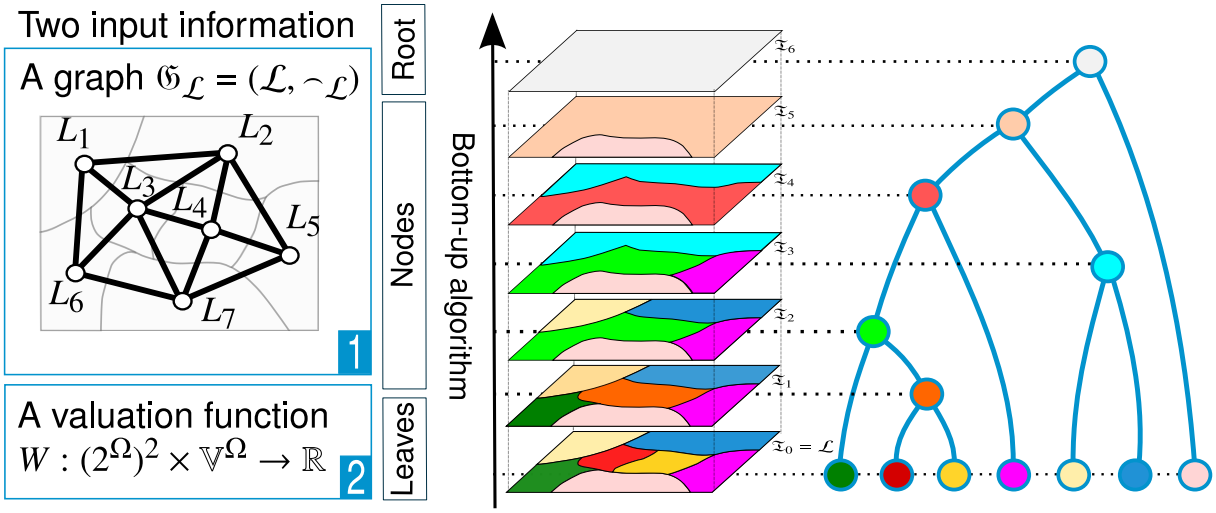


Figure 1.3: Illustration of the construction of a BPT from one image. Left (input): (1) a graph $\mathfrak{G}_{\mathcal{L}} = (\mathcal{L}, \frown_{\mathcal{L}})$ models the initial partition of the image; (2) a valuation function $W : (2^{\Omega})^2 \times V^{\Omega} \to \mathbb{R}$ allows us to iteratively choose the pair of nodes to be merged (see Section 2.1). Right: progressive, bottom-up, creation of the tree from $\mathfrak{T}_o$ to $\mathfrak{T}_6$ by iterative fusions of two neighbouring regions.

The standard Binary Partition Tree construction

The BPT construction is a bottom-up process. It proceeds from the determination of the leaves —provided by an initial partition of the image— to the root. This is done by choosing and merging iteratively two adjacent regions that minimize a criterion reflecting their likeness. This merging sequence is stored in a hierarchical structure, namely the BPT, that models the image at different scales (see the right part of Figure 1.3). A huge number of distinct BPTs may be obtained for a given initial partition of Ω. In order to determine a relevant BPT, it is then necessary to decide of the priority of the fusions between nodes. For a partition $\mathcal{P} \subseteq \Omega$, let $N_i, N_j \in \mathcal{P}$ be two distinct and adjacent regions / nodes. A BPT generation relies on two main notions: a *region model* $M_r(N_i)$ which specifies how a region N_i is characterized (e.g., colour, shape); and a *merging criterion* $O_r(N_i, N_j)$ which defines the similarity of neighbouring regions N_i, N_j and thus the merging order.

A strategy commonly adopted for representing each region is to consider their average colour in a given space (e.g., RGB, HSV), and to merge pairs of adjacent regions either similar one to each other, or similar to the region model of the novel region built from their union. Another strategy such as

in [53] considers as region model a linear combination of radiometric and geometric features, with a merging criterion where the trade-off between these features evolves during the construction of the BPT. In any case, the merging criterion is a scalar function, which requires to fuse various elements of expert knowledge in a single metric.

Tree cuts and partitions

Defining a *cut* within a tree consists of defining a subset of nodes which are all independent with respect to the hierarchical organization of the tree. In other words, considering $\mathcal{N}$ the set of all nodes of the tree, two nodes $N, N' \in \mathcal{N}$ such that N' is a child of N, cannot both belong to a same cut. Trivial examples of cuts of a tree $\mathfrak{T}$ are the singleton cut $\mathcal{R} = \{N_\Omega\}$ formed by the only root, and the cut $\mathcal{L} = \{N_x\}_{x\in\Omega}$ formed by the leaves.

A cut is said total if it is maximal for the inclusion relation, i.e. if we cannot add a new node hierarchically independent from all nodes of the cut. Otherwise, a cut is said partial. For instance, $\mathcal{R} = \{N_\Omega\}$ is a total cut: any other node is linked to N_Ω by the $\searrow$ relation. This is also the case of $\mathcal{L} = \{N_x\}_{x\in\Omega}$: any other node is linked to at least one (in fact, at least two) nodes of $\mathcal{L}$ by the $\searrow$ relation. However, any strict subset of $\mathcal{L}$ is a partial cut.

Let $\mathcal{C} \subseteq \mathcal{N}$ be a cut of $\mathfrak{T}$. By construction of a tree, any two nodes $N_1, N_2 \in \mathcal{C}$ are disjoint: $(N_1 \neq N_2) \Rightarrow N_1 \cap N_2 = \emptyset$. If $\mathcal{C}$ is a total cut, then it defines a partition of Ω: any node is nonempty, any two nodes are non-intersecting, and the union of all nodes covers the whole set Ω. If $\mathcal{C}$ is a partial cut, it defines a partial partition of Ω, i.e., a partition of a strict subset of Ω.

Advantages and limits

BPT as a search space for segmentation

By construction, any cut $\mathcal{C}$ of the BPT provides a partition of the image support Ω (see Figure 1.4).

The equivalence between cuts of $\mathfrak{T}$ and partitions of Ω allows for the use of BPTs as a data-structure for image segmentation. Indeed, segmenting the image I consists of subdividing its support Ω into (spectrally, seman-

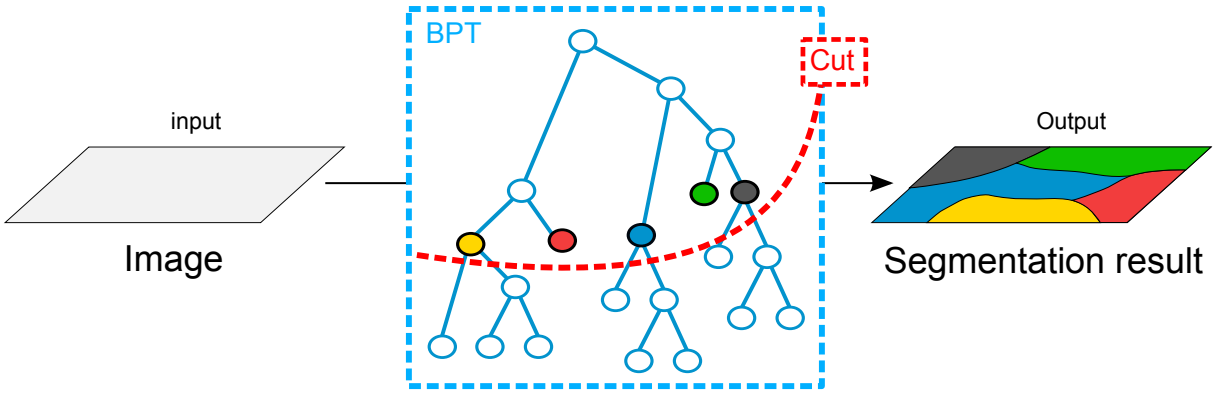


Figure 1.4: One cut on the BPT provides one partition that can be considered as an image segmentation result.

tically, etc.) homogeneous regions, i.e., providing a relevant partition (or partial partition) of Ω.

Based on the above remarks, a BPT constitutes a relevant data-structure for image segmentation. In particular, any of its cuts providing a partition, it constitutes a search space containing potential segmentation results, associated to a given image.

The main virtues of the BPTs are its low size and its hierarchical structure. First, it contains $2|\mathcal{L}| - 1$ nodes. Since the number $|\mathcal{L}|$ of leaves is bounded by the number of points of the image (i.e., $|\mathcal{L}| \leq |\Omega|$), the size of $\mathfrak{T}$ is indeed linear with respect to the image size: $|\mathcal{N}| = \mathcal{O}(|\Omega|)$. Second, the BPT is a (binary) tree structure. This organization of nodes enables the development of efficient "divide and conquer" strategies, that can lead to low computational complexity algorithms.

In fact, BPTs appear as good data-structures for image segmentation, allowing us to obtain a cut / partition / segmentation in a fast and efficient way. Actually, BPTs are indeed very good data-structures for that purpose, *provided they are correctly built.*

Limits of the Binary Partition Tree

Generally, the BPT is built in a mono-feature way (i.e., one image, one metric).

On the one hand, the use of a single metric for the construction of a BPT often leads to consider a single criterion. Indeed, the criteria used in most of image segmentation approaches are generally textural similarities

[54] or multi- / hyper-spectral homogeneity [47]. Until now, the principal solution for considering several criteria was a (linear) combination in a single region similarity measure [44]. As the metric has to be carefully thought, the conception of such (linear) combination is not an easy task for the expert.

On the other hand, the handling of several images is generally dealt with either by defining a super-image, or by gathering multiple information provided by various spectral bands into a single metric. The rate of information loss is up to the image fusion method applied.

In fact, the metric combination and the image fusion impacts the BPT construction process and the pertinence of the possible segmentation results (by extracting cuts) depends directly on their definitions.

To deal with these issues and limits, we propose in the next part a framework generalizing the construction process of BPTs in a multi-feature (i.e., one / many image(s), one / many metric(s)) way.

1.4 Résumé : Représentation hiérarchique d'images

Dans ce chapitre, nous nous avons montré, dans un premier temps, l'intérêt de la réduction de l'espace de recherche utile pour la segmentation d'image. Nous avons commencé par introduire des notions de segmentation d'images hiérarchique par graphe. Puis, nous avons introduit l'arbre binaire de partitions (BPT), ses spécificités et les limites.

Réduction de l'espace de recherche La segmentation d'image peut être considérée comme une procédure de recherche d'informations dans un espace de données défini. En particulier, cet espace correspond ici à une ou plusieurs images contenant les données. De nos jours, l'évolution des capteurs permet l'acquisition de nouvelles générations d'images ayant des dimensions élevées. Ainsi, les espaces de recherche deviennent volumineux et difficiles à gérer par les processus de segmentation classiques. Une alternative pour faciliter le partitionnement d'images à partir de tels espaces est de réduire leur taille tout en essayant de préserver au maximum les informations pertinentes. La réduction d'espace de recherche a déjà montré son utilité pour des tâches comme la segmentation d'images.

Segmentation d'images hiérarchique par graphes Les problèmes en analyse et traitement d'images, en particulier la segmentation, sont souvent considérés dans des cas discrets au travers de concepts basés sur la théorie des graphes. En pratique, les points d'une image (i.e., pixels, voxels) sont considérés comme des sommets d'un graphe et les relations spatiales entre les voisins sont modélisées par des arêtes du graphe. Ce paradigme, très utilisé depuis le début des années 70, a mené au développement de nombreuses approches de segmentation, basées sur la manipulation des graphes. Généralement, ces approches fournissent un seul résultat de segmentation pour une image donnée. La segmentation étant un problème mal posé, les approches hiérarchiques ont été développées afin d'obtenir diverses solutions adaptées à différentes échelles. De ce point de vue, beaucoup de modèles hiérarchiques d'images ont été développés, principalement dans le domaine de la morphologie mathématique. La plupart des structures hiérarchiques proposées dans la littérature sont modélisées directement à partir d'informations intrinsèques des images (e.g., signal, intensité, gradient). Dans certains contextes applicatifs, les régions contenues dans ces structures ne correspondent pas aux objets d'intérêt, en particulier quand les images sont complexes. Comparé à ces structures, les arbres binaires de partitions (BPT) sont modélisés à la fois à partir d'informations intrinsèques de l'image mais aussi à partir d'informations extrinsèques caractérisées par des connaissances a priori de l'utilisateur sur l'application visée. Ces spécificités nous motivent à nous baser sur les BPTs dans nos travaux de recherches.

Arbres binaires de partitions Considérant une image I définie sur un support fini (i.e., un ensemble de pixels Ω), un arbre binaire de partitions (BPT) est une représentation hiérarchique de I, organisée en arbre binaire noté $\mathfrak{T}$. Chaque nœud d'un BPT est une région correspondant à un sous-ensemble de Ω (i.e., $N \subseteq \Omega$). Un nœud est soit une feuille de l'arbre (i.e., une région élémentaire, un pixel) ou un nœud interne (i.e., union de deux nœuds modélisant deux régions adjacentes). La racine du BPT est un nœud correspondant au support entier Ω de l'image. En pratique, un BPT est construit d'une manière ascendante en partant des feuilles (définies par une partition initiale de Ω), en choisissant itérativement deux

régions adjacentes à fusionner qui minimisent une mesure de similarité que l'on appelle métrique. En particulier, cette métrique nécessite des connaissances a priori provenant de l'utilisateur. Son choix est crucial et influence la pertinence du BPT construit. Par comparaison aux arbres de composantes (e.g., min-tree, max-tree, inclusion-tree), plusieurs BPTs peuvent être construits à partir d'une image en fonction de la métrique choisie par l'utilisateur. Pour une partition $\mathcal{P} \subseteq \Omega$, avec $N_i, N_j \in \mathcal{P}$ deux régions (i.e., nœuds) distinctes, la construction d'un BPT nécessite la définition d'un modèle de région $M_r(N_i)$ pour modéliser les caractéristiques d'une région N_i (e.g., couleur, forme) ; et d'un critère de fusion $O_r(N_i, N_j)$ qui permet de définir la similarité des régions voisines N_i, N_j, et ainsi permettre un ordre de fusion. Par construction, une coupe $\mathcal{C}$ appliquée sur un BPT permet d'obtenir une partition du support Ω de l'image. Si le BPT est bien construit, il peut être considéré comme une structure de données pertinente pour la segmentation d'images par partitionnement. Cependant, le BPT est généralement construit à partir d'une image et d'une métrique. D'une part, l'utilisation d'une seule métrique induit la considération d'un seul critère (i.e., attribut). Afin de pouvoir prendre en compte plusieurs critères, une métrique basée, par exemple, sur leur combinaison linéaire doit être définie par l'utilisateur. Le choix de cette métrique est important et une telle combinaison n'est pas évidente pour l'utilisateur. D'autre part, l'utilisation de plusieurs images nécessite le choix d'une méthode de fusion les combinant en une seule. Afin de franchir les limites induites par le paradigme mono-critère de la construction d'un BPT, nous proposons une approche multicritère permettant la construction d'un MBPT à partir d'une / plusieurs images et une / plusieurs métriques.

Chapter 2

Multi-feature BPT construction framework

Contents

In this chapter, our first contribution is to present the BPT in a graph-based point of view (Section 2.1). We then propose a multi-feature generalization of the classical BPT construction leading to the notion of multi-feature BPT (MBPT) (Section 2.2). We also discuss about technical details of the proposed algorithm with a complexity analysis (Section 2.3).

2.1 Binary Partition Tree: A graph-based point of view

In this thesis, we show that the BPT can be seen in a graph-based point of view. We first define it, from this point of view, and then describe its structure by using graph-based formulations. Such a new representation of the BPT is required for its further generalization.

Graph-based definition of the Binary Partition Tree

Let us consider a graph $\mathfrak{G} = (\Omega, \frown)$ that satisfies the properties of the notion of an image: Ω is a nonempty, finite set, and $\frown$ is an irreflexive, symmetric relation. Such a graph $(\Omega, \frown)$ can be the graph of an image, but it can also be a graph obtained by gathering nodes / points of an image graph (e.g., initial partition by superpixel computation); in both cases, we will still note $(\Omega, \frown)$, for the sake of readibility. This is a frequently used way of reducing the space cost of the further image processing procedures. These two strategies —full image graph vs. superpixel graph— may induce different behaviour in terms of BPT quality. A huge number of distinct BPTs may be obtained for a given initial partition of Ω.

A BPT associated to $(\Omega, \frown)$ is a tree $\mathfrak{T} = (\mathcal{N}, \searrow)$ (i.e., a directed, connected, acyclic graph). In particular, the set $\mathcal{N}$ is partitioned into three subsets:

- $\mathcal{R} = \{N_\Omega\}$ such that $N_\Omega = \Omega$;
- $\mathcal{B} \subset 2^\Omega$ (2^Ω is the power-set of Ω);
- $\mathcal{L} = \{N_x\}_{x \in \Omega}$ such that $\forall x \in \Omega, N_x = \{x\}$;

and the nodes of these subsets satisfy the following properties:

$$\forall N \in \mathcal{R} \cup \mathcal{B}, d^+(N) = 2 \tag{2.1}$$
$$\forall N \in \mathcal{L}, d^+(N) = 0 \tag{2.2}$$
$$\forall N \in \mathcal{B} \cup \mathcal{L}, d^-(N) = 1 \tag{2.3}$$
$$\forall N \in \mathcal{R}, d^-(N) = 0 \tag{2.4}$$
$$\forall N \in \mathcal{R} \cup \mathcal{B}, N = \bigcup_{N \searrow N'} N' \tag{2.5}$$

where $d^+(N) = |\{N' \in \mathcal{N}, N \searrow N'\}|$ and $d^-(N) = |\{N' \in \mathcal{N}, N' \searrow N\}|$.

We use the standard terminology of trees: the elements of $\mathcal{N}$ are the nodes of $\mathcal{B}$ of $\mathfrak{T}$, the node Ω is the root of $\mathcal{R}$ of $\mathfrak{T}$, the singleton nodes of $\mathcal{L}$ are the leaves of $\mathfrak{T}$.

Less formally, a BPT $\mathfrak{T}$ of $(\Omega, \frown)$ provides a family $\mathcal{N}$ of subsets of Ω. These subsets are hierarchically organized from the whole set Ω, to the singleton sets $\{x\}$, $x \in \Omega$, with respect to the inclusion relation. This hierarchical organization is characterized by the fact that an element N

of $\mathcal{N}$, which is not a singleton set, is associated, via $\searrow$, to exactly two elements N_1, N_2 of $\mathcal{N}$ that form a binary partition of N. In other words, we have $N = N_1 \cup N_2$, $N_1 \cap N_2 = \emptyset$ and $N_1, N_2 \neq \emptyset$.

By construction, for any two nodes $N, N' \in \mathcal{N}$, we have $(N \cap N' \neq \emptyset) \Rightarrow (N \subseteq N' \vee N' \subseteq N)$, i.e. intersection between nodes implies inclusion.

Finally, it is important to notice that, except for degenerate cases (when $|\Omega| < 3$), a BPT is not unique for a given $(\Omega, \frown)$. The construction of a BPT is then an important issue.

Graph-based structural description

Graph and valuation function

The way to describe the construction of a BPT is generally considered from spatial (the way regions are built) and descriptive (the way regions are characterized and how they can be considered similar) points of view. Indeed, the classical description of the BPT construction considers as input: the image I (i.e., the geometrical embedding of Ω, and the value associated to each point of Ω); a region model, that "describes" the nodes; and a merging criterion, that quantifies the homogeneousness of nodes before and after fusion. These information are important from an applicative point of view.

However, beneath these image and knowledge-based notions, the construction of a BPT is intrinsically a graph collapsing process. From an algorithmic point of view, the region models and merging criteria define a valuation on the edges that allows us to *choose* which nodes to fuse at any given step. In the sequel, we will then consider, without loss of generality, that a BPT is fully defined by only two input information (see left part of Figure 1.3):

1. a graph $\mathfrak{G}_{\mathcal{L}} = (\mathcal{L}, \frown_{\mathcal{L}})$ that models the initial partition of the image;
2. a valuation function (i.e., metric) $W : (2^{\Omega})^2 \times \mathbb{V}^{\Omega} \to \mathbb{R}$ for choosing the pair of nodes to be merged.

Structural description of the algorithm

Now, let us consider an initial partition $\mathcal{L}$ of Ω. Each node $L \subseteq \Omega$ of $\mathcal{L}$ is assumed to be connected with respect to $\frown$. This partition $\mathcal{L}$ defines the

set of the BPT leaves we are going to build. For instance, $\mathcal{L}$ can be the set of the image flat zones. It is equipped by the adjacency $\frown_{\mathcal{L}}$ inherited from $\frown$. This leads to a graph $\mathfrak{G}_{\mathcal{L}} = (\mathcal{L}, \frown_{\mathcal{L}})$ that models the structure of the partition of the image I.

The BPT is the data-structure that describes the progressive collapse of $\mathfrak{G}_{\mathcal{L}}$ onto the trivial graph $(\Omega, \emptyset)$. This process consists of defining a sequence $(\mathfrak{G}_i = (\mathcal{N}_i, \frown_{\mathcal{N}_i}))_{i=0}^{n}$ (with $n = |\mathcal{L}| - 1$) as follows. First, we set $\mathfrak{G}_0 = \mathfrak{G}_{\mathcal{L}}$. Then, for each i from 1 to n, we choose the two nodes N_1 and N_2 of $\mathfrak{G}_{i-1}$ linked by the edge $\frown_{\mathcal{N}_{i-1}}$ such that $N_1 \frown_{\mathcal{N}_{i-1}} N_2$, which minimizes the valuation function W. We define $\mathfrak{G}_i$ such that $\mathcal{N}_i = \mathcal{N}_{i-1} \setminus \{N_1, N_2\} \cup \{N\}$, with $N = N_1 \cup N_2$; in other words, we replace these two nodes by their union. The adjacency $\frown_{\mathcal{N}_i}$ is defined accordingly from $\frown_{\mathcal{N}_{i-1}}$. We remove the edge (N_1, N_2), and we replace each edge (N_1, N_3) and / or (N_2, N_3) by an edge (N, N_3). In particular, two former edges may be fused into a single.

From a structural point of view, the BPT $\mathfrak{T}$ is built in parallel to this progressive collapse from $\mathfrak{G}_0$ to $\mathfrak{G}_n$. We define a sequence $(\mathfrak{T}_i)_{i=0}^{n}$ as follows. We set $\mathfrak{T}_0 = (\mathcal{N}_0, \emptyset) = (\mathcal{L}, \emptyset)$. Then, for each i from 1 to n, we build $\mathfrak{T}_i$ from $\mathfrak{T}_{i-1}$ by adding the new node $N = N_1 \cup N_2$, and the two edges $N \searrow N_1$ and $N \searrow N_2$. The BPT $\mathfrak{T}$ is then defined as $\mathfrak{T}_n$.

Remark In [55], a graph-based definition of the BPT construction is also proposed, that relies on a minimum spanning tree paradigm. However, this formalization is valid only if the merging order is associated to a valuation of the edges that is fixed *a priori* on the initial partition. This is not the case, in general.

Data-structures

The above description of the BPT construction algorithm implies to define and update the following data-structures:

- the graph $\mathfrak{G}$ which stores nodes that remain to be merged, and their adjacency links; and
- the tree $\mathfrak{T}$ that is progressively built.

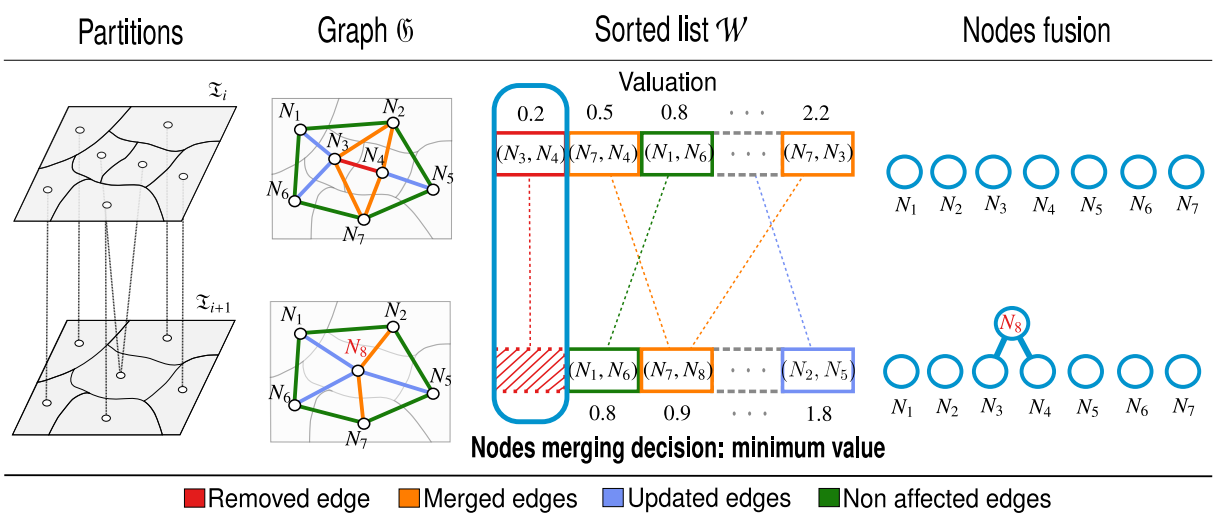


Figure 2.1: One step of the BPT construction from one image. From the left to the right: First: the partition of Ω before and after the fusion of two nodes. Second: the associated graph $\mathfrak{G}$, before and after the fusion of N_3 and N_4, forming the new node N_8. The red edge is removed. The blue and orange edges are updated, e.g., (N_1, N_3) becomes (N_1, N_8). The orange edges are merged by pairs, e.g., (N_7, N_3) and (N_7, N_4) become (N_7, N_8). The green edges are not affected. Third: the sorted list that gathers the scalar valuations of each remaining edge of $\mathfrak{G}$. The red cells are removed, as the edge (N_3, N_4) is suppressed (this edge had been chosen due to its highest position in the list). The scores of blue and orange cells are updated with respect to N_8. The orange cells are merged by pairs. The positions of the blue and orange cells are updated with respect to their new scores. The scores of the green cells are not affected. Last: a new part of the BPT $\mathfrak{T}$ is created by adding the new node N_8, and linking it to its two children nodes N_3 and N_4.

In order to efficiently compute the valuation W, it is also important to associate each node of $\mathfrak{G}$ to the corresponding part of the image I (e.g., via a mapping between $\mathfrak{G}$ and Ω).

The last required data-structure is a sorted list $\mathcal{W}$ that gathers the scalar valuations of each remaining edge of $\mathfrak{G}$. This list contains the information for choosing the couple of nodes to be merged relatively to a given metric. One iteration of the BPT construction algorithm is illustrated in Figure 2.1. This choice is made in constant time $\mathcal{O}(1)$, since $\mathcal{W}$ is sorted. After the merging operation, $\mathcal{W}$ has to be updated: (1) to remove the edge between the two nodes; (2) to update the edges affected by the merging operation; and (3) to re-order these updated edges. Operation (1) is carried out in constant time $\mathcal{O}(1)$. Operation (2) is carried out in $\mathcal{O}(\alpha.T_W)$, where T_W is the cost of the computation of W for an edge, and α is the number of neighbours of the merged nodes (α is generally bounded by a low constant value). Operation (3) is carried out in $\mathcal{O}(\alpha.\log_2|\mathcal{W}|)$.

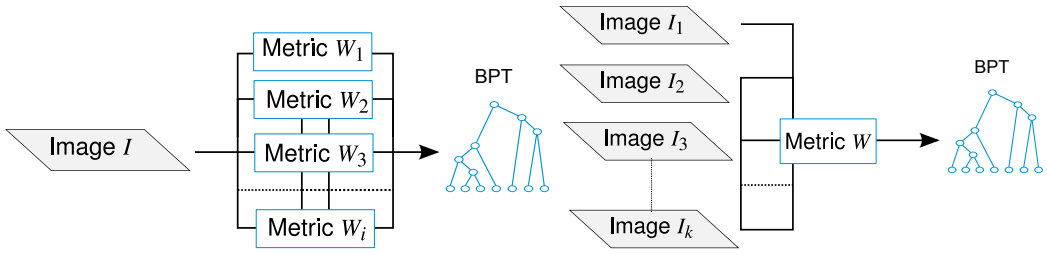


Figure 2.2: Left: one BPT built from several metrics associated to a same image. Right: one BPT built from a same metric associated to several images.

Based on the graph-based formulation presented in this section, the next section (Section 2.2) proposes an extension of the BPT construction in a multi-feature way.

2.2 Multi-feature generalization of the BPT construction

In this section, we investigate a multi-feature generalization of the BPT construction algorithm (one of the major contributions of this thesis). We still consider that this construction, viewed as a graph collapsing problem, takes as input the graph $\mathfrak{G}_{\mathcal{L}} = (\mathcal{L}, \frown_{\mathcal{L}})$. However, we now use *several* valuation functions[1] (i.e., multi-metric) $W_\star : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$. Their role remains to choose, at each step of the process, the pair of nodes to be merged. We discuss the structural and algorithmic side effects of using several valuation functions "at the same time", instead of only one. Our purpose is still to build *one* BPT from these input information. Introducing several valuation functions allows us to embed several features in an independent way in the construction process. These features can represent several metrics associated to a same image (see left part of Figure 2.2); a same metric associated to several images of a same scene (see right part of Figure 2.2); or various metrics on various images of a same scene (see half top of Figure 2.3). In other words, this generalized paradigm opens the way to a versatile handling of multi-image (e.g., multimodal, multi-time, etc.) and / or multi-criteria definition of consensual BPTs, without the necessity of defining beforehand a unified metric.

[1]In this manuscript, the term "valuation function" refers to metric for measuring region similarities.

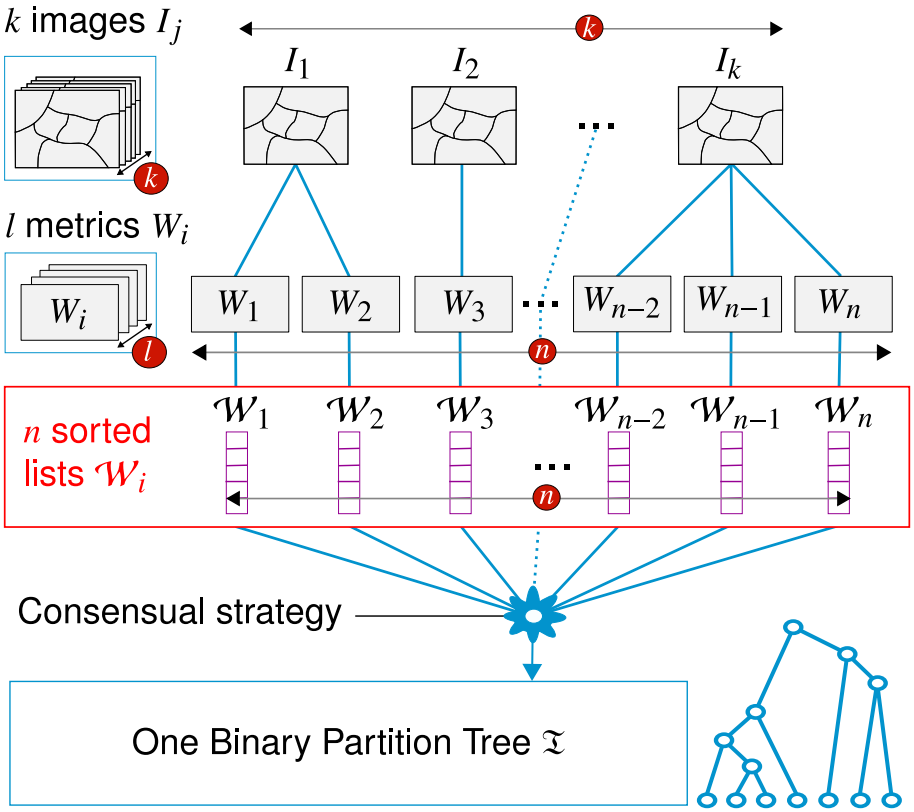


Figure 2.3: General structure of a BPT construction involving n sorted lists. Each list $\mathcal{W}_i$ gathers edges information from one valuation function associated to a given metric on a given image. A BPT is obtained by applying a consensual strategy between the information carried by these n lists (see Section 2.2).

Structural evolutions

The proposed generalization deals with the "feature" part of the construction. As stated in Section 2.1, we need a graph that models the initial partition $\mathcal{L}$ of the image(s). Our purpose remains to collapse this unique graph associated to a unique spatial scene. Practically, it implies that the (potentially multiple) images involved in the BPT construction are defined in the same spatial reference (i.e., the same support Ω). This constraint is not a real issue. Indeed, this spatial coherence assumption is generally a standard requirement in image processing (e.g., in medical imaging, via registration; in remote sensing via georeferencing; etc.). A graph $\mathfrak{G}_{\mathcal{L}}$, which is isomorphic to $(\Omega, \frown)$, can be obtained easily, either by subdividing Ω into one-point singleton sets or by considering flat zones.

Then, the "graph" part of the BPT construction process remains unchanged. In terms of data-structures, the generalized BPT construction

will still handle one graph $\mathfrak{G}$, that will be progressively collapsed; and one tree $\mathfrak{T}$ that will be built to provide the BPT. A unique mapping between $\mathcal{N}$ and Ω will still allow us to have access to the values of a node for the different images.

Let us now consider the "feature" part of the data-structure. In the initial BPT construction approach, the valuation function $W : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$ was explicitly modelled by a sorted list $\mathcal{W}$ of the values of all graph edges. This list was updated during the progressive collapsing of $\mathfrak{G}$, by removing elements from the list; updating the values of some edges; and re-sorting edges with respect to their updated values.

We now consider $n \geq 1$ valuation functions $W_\star : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$, which means that each edge is associated to n values, one for each function. By considering k distinct images and l distinct metrics, we may have up to $n = k.l$ such valuation functions. This leads to define no longer one, but n sorted lists $\mathcal{W}_i$ $(1 \leq i \leq n)$. Each list $\mathcal{W}_i$ is associated with a specific valuation function $W_i : (2^\Omega)^2 \times \mathbb{V}_i^\Omega \to \mathbb{R}$, defined with respect to a value set $\mathbb{V}_i$ (see Figure 2.3). The handling of these sorted lists remains the same in terms of removal, value updating / resorting, as for one list.

Our purpose is now to build a BPT from these n lists by generalizing the algorithm described in Chapter 1 (Section 1.3), which initially depended on one list $\mathcal{W}$.

Algorithmic evolutions

From an algorithmic point of view, each iteration of the construction preserves the same organization. An edge is chosen and the two incident nodes of the graph are merged. This operation leads to update the nodes and edges of $\mathfrak{G}$, and adds a new node plus two edges in $\mathfrak{T}$. The main differences are that:

- $n \geq 1$ sorted lists are updated;
- and the choice of the optimal edge is made with respect to the information carried by these n sorted lists instead of only one.

At each iteration of the algorithm (see Figure 2.4), the choice of the optimal edge to remove, leading to the node fusion, depends on a decision. This decision is indeed a consensus, made with respect to the information

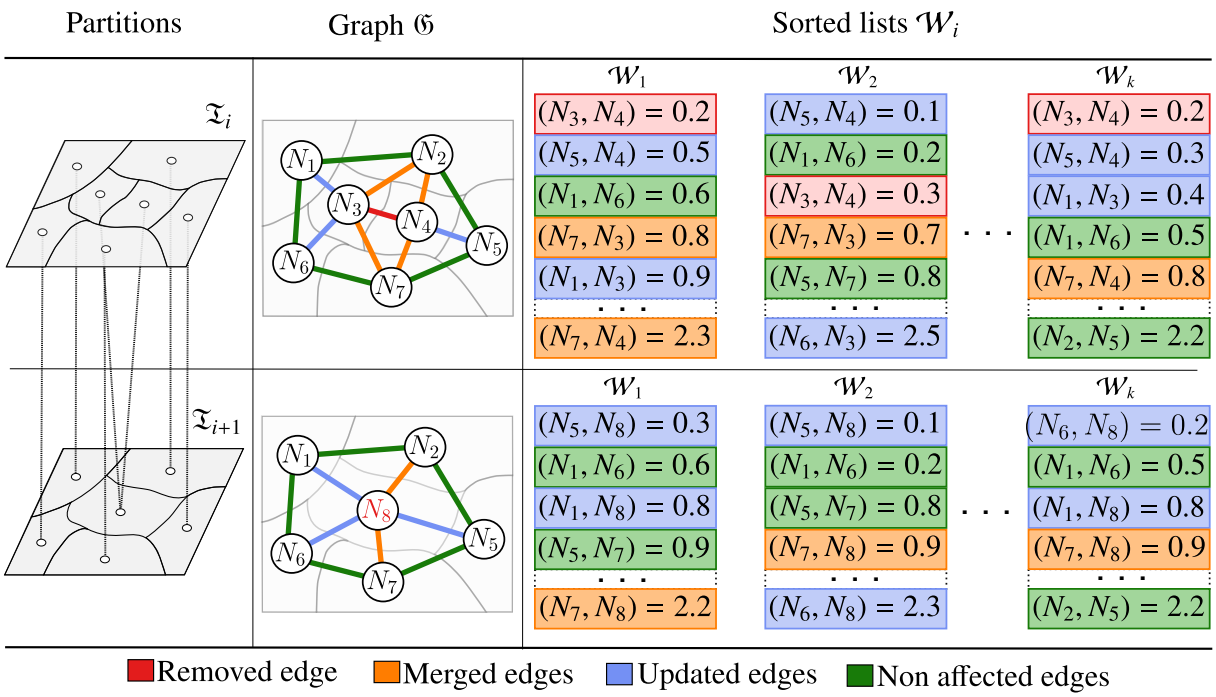


Figure 2.4: One step of a BPT construction involving n lists $\mathcal{W}_i$. Left: the partition of Ω before and after the fusion of two nodes. Center: the associated graph $\mathfrak{G}$, before and after the fusion of N_3 and N_4, forming the new node N_8. The red edge is removed. The blue and orange edges are updated, e.g., (N_1, N_3) becomes (N_1, N_8). The orange edges are merged by pairs, e.g., (N_7, N_3) and (N_7, N_4) become (N_7, N_8). The green edges are not affected. Right: the n lists $\mathcal{W}_i$, each corresponding to a valuation function W_i. The red cells are removed, as the edge (N_3, N_4) is suppressed (this edge had been chosen due to its "optimal" position and / or valuation in the n lists). The scores of blue and orange cells are updated with respect to N_8. The orange cells are merged by pairs. The positions of the blue and orange cells are updated with respect to their new scores. The scores of the green cells are not affected.

provided by these n lists. On the one hand, information are carried by the sorted lists $\mathcal{W}_i$, that give a *relative* information on edges, induced by their ordering with respect to W_i. On the other hand, information are carried by the valuation functions $W_i : (2^\Omega)^2 \times \mathbb{V}_i^\Omega \to \mathbb{R}$ that give an *absolute* value to each edge. These information are of distinct natures. We study their relevance according to various kinds of consensus policies. In particular, we identify three main families of consensus strategies.

Absolute information consensus

Let us consider, for instance, that the consensus policy consists of choosing the edge of lowest mean valuation among the n lists $\mathcal{W}_i$, or the edge of minimal valuation among all lists. The first consensus (namely *min of*

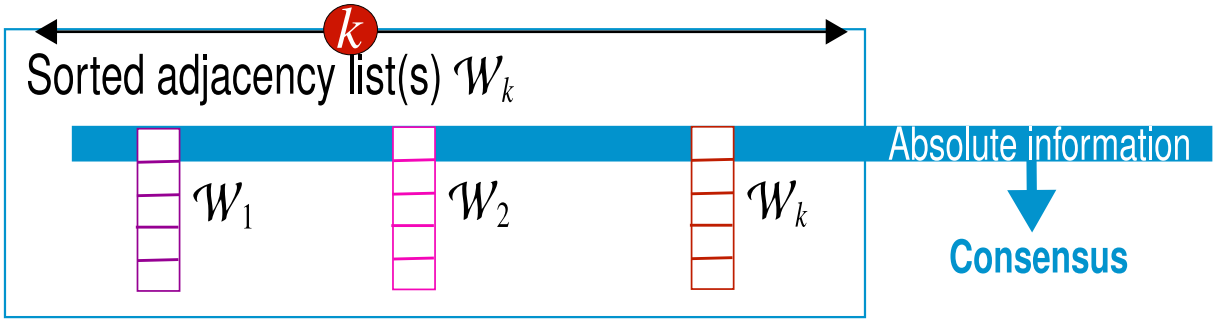


Figure 2.5: Consensus from absolute information contained in all lists.

mean) is defined by a linear formulation:

$$\arg_{(N,N')\in\mathcal{N}} \min \sum_{i=1}^{n} W_i((N,N')) \tag{2.6}$$

The second (namely *min of min*) is defined by a non-linear formulation:

$$\arg_{(N,N')\in\mathcal{N}} \min \min_{i=1}^{n} W_i((N,N')) \tag{2.7}$$

However, in both cases, the decision is made by considering *absolute* information carried by the edges (see Figure 2.5). More generally, when the information carried by the values of each edge is a sufficient knowledge, n sorted lists $\mathcal{W}_i$ are not necessary. In such a case, a single sorted list $\mathcal{W}$ that contains the information of these, linear or non-linear, formulations is indeed sufficient. The BPT construction involving n lists is then equivalent to that from one list.

The main difficulty raised by this policy derives from the potential heterogeneity of the values carried by the different $W_\star$ valuation functions. Indeed, when the features are not of same nature, their values should be normalized and / or weighted to allow for the definition of adequate fusion / comparison operators. This issue is mainly the same that occurs in most optimization schemes where a given metric is built from several terms of varying semantics. Although this remains tractable, at the cost of a certain expertise, it argues in favour of considering the next two proposed policies as relevant alternatives.

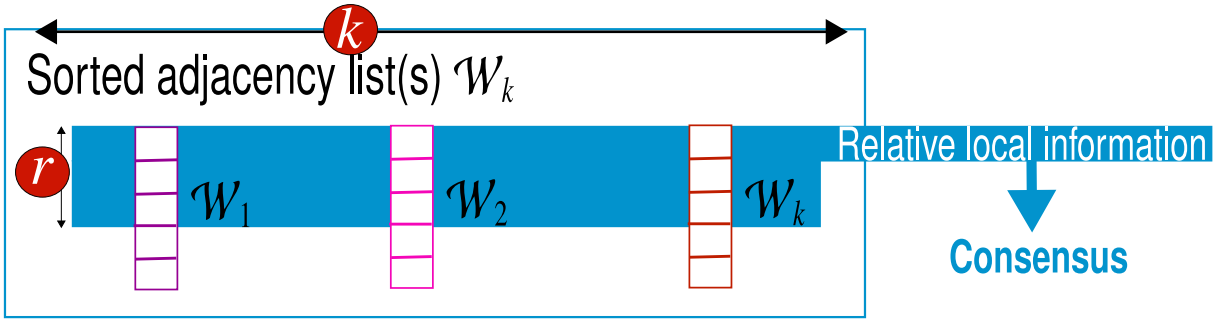


Figure 2.6: Consensus from relative information contained in a top defined range r of all lists.

Relative local information consensus

Let us now consider, for instance, that the consensus policy consists of choosing the edge that is the most often in first position in the n sorted lists $\mathcal{W}_i$, or the most frequently present in the $r \ll |\mathcal{W}_i|$ first positions in the n lists $\mathcal{W}_i$. These consensus (namely, *majority vote* and *most frequent*, potentially weighted) policies do not act on the absolute valuations of the edges, but on their relative positions in the lists (see Figure 2.6). Another strategy can also consist of choosing the edge with the lowest *mean of ranks*, according to the position of the edges within the lists. In such cases, it is then mandatory to maintain n sorted lists. However, the decision process does not require to access the whole lists. It can be restricted to the first (or the first r) element(s) of each, leading to a *local* decision process.

Relative global information consensus

Let us finally consider that the consensus policy consists of choosing the edge that has the best global ranking among the n sorted lists $\mathcal{W}_i$. As previously, such consensus (e.g., *best average* or *best median ranking*) policy also acts on the relative positions of the edges in the lists, and need not consider the absolute values of the edges. However, in contrast with the above case, the decision process requires to explicitly access the whole content of all these lists (see Figure 2.7), leading to a *global* decision process of higher computational cost.

In the next section (Section 2.3), we detail the proposed algorithm and its complexity analysis.

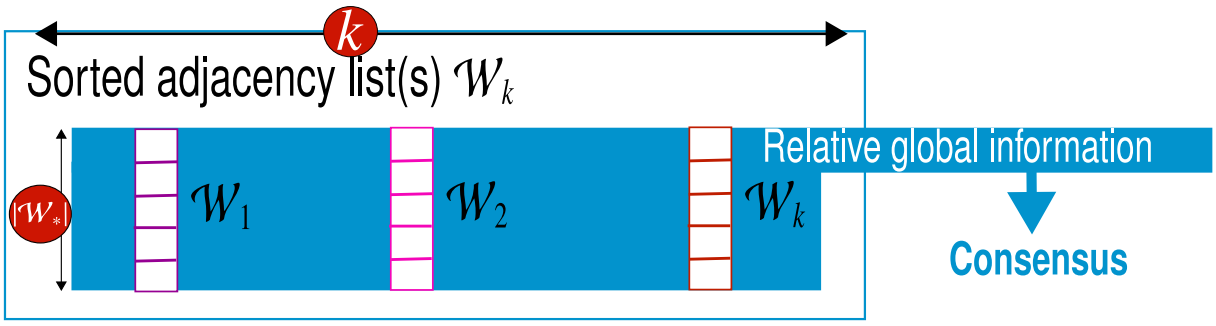


Figure 2.7: Consensus from relative information contained in all lists.

2.3 Technical discussion

Details of the proposed algorithm

To build a multi-feature BPT, namely MBPT, the main structure of the classical BPT construction remains unchanged. Indeed, the MBPT is still built from its leaves to its root, by choosing and merging adjacent regions that minimize a merging criterion. However, instead of dealing with a classical merging criterion on one image and one metric, we may now consider many images, different metrics, and a consensus strategy as input parameters. The core of the algorithm is divided as follows:

1. preparation of the lists $\mathcal{W}_i$ based on each image and metric couple;
2. preparation of the graph $\mathfrak{G}$ modelled as a Region Adjacency Graph (RAG);
3. creation of the MBPT $\mathfrak{T}$ by merging pairs of nodes.

The general structure of the proposed algorithm is presented in Algorithm 1. For the sake of reproducibility, implementation details of the proposed algorithms and data-structures are provided in Appendix A and Appendix B.

Preparation of the lists $\mathcal{W}_i$ and the RAG

As already mentioned in Section 2.2, by considering k distinct images and l distinct metrics, we may have up to $n = k.l$ valuation functions. This leads us to prepare no longer one, but n sorted lists $\mathcal{W}_i$. Each list $\mathcal{W}_i$ contains all edges of the graph $\mathfrak{G}$ and is maintained ordered during whole process (see

Algorithm 1: MBPT construction

```
Data: Image I[ ]; Metric W[ ]; Consensus strategy Consensus;
Result: A Multi-feature Binary Partition Tree MBPT;
/* Variables                                                          */
Region Adjacency Graph 𝔊;                                  // the RAG
List 𝒲[ ];                                   // the list of lists 𝒲_i
Image preSeg;                          // initial partition of the image
/* Initialization                                                     */
prepareLists(I, W);          // associates metrics, images & adjacency lists
prepareRAG(𝒲, preSeg);                  // builds initial leaves & adjacencies
fillLists(𝒲, 𝔊);              // fills the lists with the initial adjacencies
/* Node fusion                                                        */
while 𝔊.tabAdja is not empty do
    Adjacency chosenAdja = Consensus.apply(𝒲);
    Node n1 = chosenAdja.region1;
    Node n2 = chosenAdja.region2;
    Node newNode = fusionOf(n1, n2);
    𝔊.tabNode.add(newNode);
    𝔊.tabAdja.update();                  // adds, removes and updates values
    /* Updates the content of the lists                               */
    for i = 0; i < 𝒲.size(); i++ do
        𝒲[i].update();                   // adds, removes and updates values
MBPT.tabNode = 𝔊.tabNode;
return MBPT
```

right part of Figure 2.4). Its data-structure and maintenance are detailed in Appendix B. An object[2] representing the metric is instantiated by the algorithm and linked to each list $\mathcal{W}_i$. At this stage, these n lists are still empty.

When all lists $\mathcal{W}_i$ are prepared and linked to the right couple of image and metric, a graph $\mathfrak{G}$ (implemented as a RAG) has to be built. The initial nodes of the graph are created from the pixels or from an initial partition of the image support. They constitute the set of leaves of the MBPT. Once the leaves are prepared, the edges of the graph are created and their values are computed by invoking particular methods of the corresponding metric. These edges represent the adjacencies between neighbouring regions of the image. For each edge, n computed metric values are associated. In parallel, we use these metric values to fill the n lists $\mathcal{W}_i$ that will be further maintained sorted.

[2]The term "object" refers here to the objects used for oriented object software development.

Creation of the multi-feature BPT by merging nodes

Once the lists $\mathcal{W}_i$ and the graph $\mathfrak{G}$ are ready, an iterative fusion of the nodes is operated to build the MBPT. This iterative process stops when the lists $\mathcal{W}_i$ are empty. The root of the tree is then reached. Each iteration is mainly composed of:

1. the choice of the next two adjacent nodes of $\mathfrak{G}$ to be merged, based on the consensus strategy;
2. the handling of the n lists $\mathcal{W}_i$ by removing, in each list, the adjacency linking the two chosen nodes, and updating the values of all adjacencies that link them to their neighbours;
3. the resorting of the lists.

Complexity analysis

In the multi-feature paradigm, choosing the edge to remove is no longer a constant time operation. Indeed, it depends on the way information are used and compared. Afterwards, operations (1–3) described in the standard BPT construction algorithm, for the sorted list maintenance, have to be duplicated for each list. These operations are then carried out in $\mathcal{O}(n)$, $\mathcal{O}(n.\alpha.T_{W_\star})$ and $\mathcal{O}(n.\alpha.\log_2|\mathcal{W}_\star|)$, respectively, where T_{W_*} is the cost for computing W_* for a given edge, while α is an upper bound for the nodes degree within the graph $\mathfrak{G}$.

This analysis can be refined with respect to the policies considered for choosing an edge, and the information carried by the valuation functions W_i and / or the sorted lists $\mathcal{W}_i$. Besides, the choice of a consensus strategy is strongly application-dependent. As a consequence, it is important to consider a trade-off between the structural and computational cost of the approach versus the benefits in terms of results accuracy. In particular, these costs are summarized in Table 2.3.

This table provides the cost of an elementary step of the BPT construction. The number of these steps is equal to the size of the initial graph $\mathfrak{G}_\mathcal{L}$, namely $|\mathcal{L}|$. More precisely, it is equal to the number of vertices in this initial graph (minus one). Indeed, each step merges two of these vertices, until obtaining a graph formed by exactly one node. At each step, at least

Table 2.1: Cost of the BPT construction for various families of consensus. For readability purpose, α and $T_{W_\star}$, practically bounded by low constant values, have been omitted.

	$\#\ \mathcal{W}_\star$	Edge choice	Edge removal	Edges update	Edges sorting
Standard BPT [22]	1	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(\log_2 \lvert\mathcal{W}\rvert)$
Absolute information	1	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(\log_2 \lvert\mathcal{W}\rvert)$
Relative local information	n	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n.\log_2 \lvert\mathcal{W}_\star\rvert)$
Relative global information	n	$\mathcal{O}(n.\lvert\mathcal{W}_\star\rvert)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n.\log_2 \lvert\mathcal{W}_\star\rvert)$

one edge is removed from the graph. The number of remaining edges is then equal to the size of the list(s) $|\mathcal{W}_i|$. We can assume that the number of edges is bounded by the number of vertices of the graph, up to a multiplicative constant α, generally low for images defined on discrete grids. Based on these assumptions, an upper bound for the overall computational cost of the standard BPT construction is

$$\sum_{i=1}^{|\mathcal{L}|}(\mathcal{O}(1)+\mathcal{O}(\log_2|\mathcal{W}|)) = \sum_{i=1}^{|\mathcal{L}|}\mathcal{O}(1)+\sum_{i=1}^{|\mathcal{L}|}\mathcal{O}(\log_2(\alpha.i)) \tag{2.8}$$

$$= \mathcal{O}(|\mathcal{L}|)+\log\alpha.\mathcal{O}(\log_2(|\mathcal{L}|!)) \tag{2.9}$$

$$= \mathcal{O}(|\mathcal{L}|\log_2|\mathcal{L}|) \tag{2.10}$$

and so is the cost for the first consensus policy (absolute information consensus). Following the same kind of calculation, the cost for the second consensus policy (relative local information consensus) is $\mathcal{O}(n.|\mathcal{L}|\log_2|\mathcal{L}|)$. That of the third (relative global information consensus) is $\mathcal{O}(n.|\mathcal{L}|^2)$.

In the next chapter (Chapter 3), we discuss about some perspectives for scaling up the multi-feature BPT construction.

2.4 Résumé : Cadre de construction d'un BPT multicritère

Dans ce chapitre, notre première contribution correspond à la représentation du BPT d'un point de vue graphe. Ensuite, nous avons proposé une approche généralisant la construction classique d'un BPT menant à la notion de BPT multicritère (MBPT). Nous avons aussi discuté des détails techniques de l'algorithme que nous proposons, suivis d'une étude de

complexité.

Arbre binaire de partitions : un point de vue basé graphe Dans cette thèse, nous montrons que le BPT peut être observé d'un point de vue graphe. Considérons un graphe $\mathfrak{G} = (\Omega, \frown)$ qui satisfait les propriétés de la notion d'image : Ω est un ensemble fini non vide, et $\frown$ est une relation irréflexive et symétrique. Un tel graphe $(\Omega, \frown)$ peut être celui d'une image, mais peut aussi être obtenu à partir d'une partition initiale (e.g., partition initiale obtenue à partir des superpixels). Un BPT associé à $(\Omega, \frown)$ est un arbre $\mathfrak{T} = (\mathcal{N}, \searrow)$ (i.e., graphe orienté, connecté, acyclique). En particulier, l'ensemble $\mathcal{N}$ est partitionné en trois sous-ensembles composés par la racine $\mathcal{R}$, les nœuds $\mathcal{B}$ et les feuilles $\mathcal{L}$. Par construction, considérant deux nœuds $N, N' \in \mathcal{N}$, nous avons $(N \cap N' \neq \emptyset) \Rightarrow (N \subseteq N' \vee N' \subseteq N)$, i.e. une intersection entre nœuds implique une inclusion. La description classique de la construction d'un BPT considère en entrée l'image I, d'un modèle de région $M_r(N_i)$ et d'un critère de fusion $O_r(N_i, N_j)$. D'un point de vue algorithmique, le modèle de région et le critère de fusion définissent des valeurs associées aux arêtes du graphe permettant le choix des nœuds à fusionner. Nous considérons donc que le BPT peut être totalement défini par un graphe $\mathfrak{G}_\mathcal{L} = (\mathcal{L}, \frown_\mathcal{L})$ (qui modélise la partition initiale de l'image) et d'une métrique $W : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$ (pour le choix des nœuds à fusionner) en entrée. Pour une partition initiale $\mathcal{L}$ de Ω, $\mathfrak{G}_\mathcal{L} = (\mathcal{L}, \frown_\mathcal{L})$ est un graphe qui modélise la structure de cette partition de I. La construction d'un BPT est un processus intrinsèque de réduction du graphe $\mathfrak{G}_\mathcal{L}$ en un graphe trivial $(\Omega, \emptyset)$. Ce processus consiste en la définition d'une séquence $(\mathfrak{G}_i = (\mathcal{N}_i, \frown_{\mathcal{N}_i}))_{i=0}^n$ (avec $n = |\mathcal{L}| - 1$) qui symbolise la réduction progressive de $\mathfrak{G}_0$ à $\mathfrak{G}_n$. La séquence $\mathfrak{G}_i$ est associée à une séquence $(\mathfrak{T}_i)_{i=0}^n$ qui représente l'arbre. A la fin du processus, le BPT $\mathfrak{T}$ est alors défini par $\mathfrak{T}_n$. Cette description de la construction d'un BPT implique la mise à jour de structures de données comme le graphe $\mathfrak{G}$, l'arbre $\mathfrak{T}$ et une liste triée $\mathcal{W}$ qui contient les valeurs de chaque arête de $\mathfrak{G}$ calculées par la métrique choisie. La valeur minimale contenue dans la liste $\mathcal{W}$ permet le choix du couple de nœuds à fusionner pour chaque étape de la création du BPT $\mathfrak{T}$. Ce choix est effectué en temps constant $\mathcal{O}(1)$, puisque $\mathcal{W}$ est triée. Après la fusion, $\mathcal{W}$ doit être mise à jour : (1) enlever les arêtes entre les deux nœuds ;

(2) mettre à jour les arêtes affectées par la fusion ; et (3) ré-ordonner les arêtes. L'opération (1) est effectuée en temps constant $\mathcal{O}(1)$. L'opération (2) est effectuée en $\mathcal{O}(\alpha.T_W)$, où T_W est le coût de calcul de W pour une arête, et α est le nombre de régions des nœuds fusionnés. L'opération (3) est effectuée en $\mathcal{O}(\alpha.\log_2|\mathcal{W}|)$.

Généralisation multicritère de la construction d'un BPT L'une de nos contributions majeures dans cette thèse est la généralisation multicritère de la construction d'un BPT. D'un point de vue graphe, nous considérons toujours que cette construction prend en entrée une image I et un graphe $\mathfrak{G}_\mathcal{L} = (\mathcal{L}, \frown_\mathcal{L})$. Cependant, nous utilisons maintenant plusieurs métriques $W_\star : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$ qui contiennent des informations permettant le choix des couples de nœuds à fusionner lors de chaque étape de la construction. Notre but est toujours l'obtention d'un unique BPT à partir de ces informations. Dans notre cas, l'utilisation de plusieurs métriques permet de prendre en compte plusieurs caractéristiques d'une façon plus indépendante durant le processus de construction. Ces caractéristiques peuvent représenter soit plusieurs métriques associées à une image, soit une métrique associée à plusieurs images d'une même scène. En d'autres termes, notre paradigme de généralisation permet de gérer plusieurs images et plusieurs critères d'une façon plus versatile, tout en n'ayant pas à définir une super-métrique unifiant plusieurs à partir d'une combinaison linéaire. En termes de structure de données, la construction généralisée gérera toujours un unique graphe $\mathfrak{G}$ qui sera réduit progressivement pour former un arbre $\mathfrak{T}$ qui sera le BPT. Contrairement à la construction classique, nous considérons maintenant $n \geq 1$ métriques $W_\star : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$, ce qui implique l'association de chaque arête du graphe avec n valeurs. En considérant k images distinctes et l attributs distincts, nous pouvons avoir $n = k.l$ listes $\mathcal{W}_i$ $(1 \leq i \leq n)$ triées associées à une métrique $W_i : (2^\Omega)^2 \times \mathbb{V}_j^\Omega \to \mathbb{R}$. La mise à jour de ces listes restent la même que pour une seule liste dans la procédure classique, en termes de suppression et de mise à jour des valeurs. Notre but est maintenant de construire un BPT à partir de ces n listes $\mathcal{W}_i$. D'un point de vue algorithmique, l'organisation de chaque itération de la construction reste la même. La différence principale est qu'il y a maintenant $n \geq 1$ listes à trier et à mettre à jour, et le choix de

l'arête optimale pour le choix du couple de nœuds à fusionner dépend des information contenues dans ces listes. Pour chaque décision de fusion, la pierre angulaire de notre approche se base sur une stratégie de consensus prenant en compte les informations fournies par les n listes. Nous avons identifié trois familles principales de stratégies de consensus. La première regroupe les stratégies de consensus des informations absolues des listes (e.g., *min of mean*, *min of min*). La deuxième regroupe les stratégies de consensus des informations relatives locales (i.e., contenues par les r premiers éléments des listes) des listes (e.g., *majority vote*, *most frequent*). Et la troisième regroupe les stratégies de consensus des informations relatives globales (i.e., tout le contenu des listes) des listes (e.g., *best average*, *best median ranking*).

Discussion technique Trois étapes importantes de l'algorithme de construction des MBPTs correspondent à la préparation des n listes $\mathcal{W}_i$ basée sur chaque couple image-métrique; la préparation du graphe d'adjacence des régions $\mathfrak{G}$; et la création du MBPT $\mathfrak{T}$ par la fusion des nœuds. Les listes $\mathcal{W}_i$ sont maintenues triées pendant tout le processus de construction et leur préparation consiste à la mise en liaison entre les couples images-métriques. Dès que ces listes sont prêtes, le graphe $\mathfrak{G}$ est construit à partir de la partition initiale $\mathcal{L}$ qui correspond aux feuilles du MBPT $\mathfrak{T}$. Une fois les sommets du graphe $\mathfrak{G}$ préparés, les arêtes qui symbolisent les adjacences entre les régions voisines sont créées et associées à n valeurs contenues dans les listes $\mathcal{W}_i$ et calculées par les métriques W_i correspondantes. À partir de ces valeurs, la stratégie de consensus choisie permet de prendre des décisions menant à des fusions progressives des couples de nœuds similaires voisins. À la fin de chaque itération, les listes sont mises à jour, maintenues triées et un nouveau nœud N du MBPT $\mathfrak{T}$ est créé. Le processus itératif s'arrête lorsque les listes sont vides. Cela signifie que la racine R du MBPT $\mathfrak{T}$ a été générée. Avec le paradigme multicritère, le choix des arêtes à supprimer ne se fait plus en temps constant. En effet, cela dépend de la façon dont les informations ont été utilisées et comparées. Les étapes (1–3) décrites pour l'algorithme de construction d'un BPT classique se retrouvent dupliquées pour chaque liste. Ces opérations sont respectivement réalisées en $\mathcal{O}(n)$, $\mathcal{O}(n.\alpha.T_{W_\star})$ et $\mathcal{O}(n.\alpha.\log_2|\mathcal{W}_\star|)$ où T_{W_i} est le coût

de calcul de W_i pour chaque arête, et α le nombre de régions des nœuds fusionnés. Cette analyse peut être raffinée en respectant les politiques de consensus considérées pour le choix des arêtes et les informations apportées par les métriques W_i et / ou des listes triées $\mathcal{W}_i$. De plus, le choix de la stratégie de consensus dépend fortement de l'application.

Chapter 3

Perspectives for scaling up the multi-feature BPT construction

Contents

The complexity analysis presented in Section 2.3 emphasises three bottlenecks for BPT construction in general, and in particular for MBPTs:

(1) the size of the edge lists, linearly dependent on the size $|\mathcal{L}|$ of the initial graph; this leads to the $\log_2 |\mathcal{L}|$ part of the cost (and a $|\mathcal{L}|$ part for the third policy);

(2) the number of steps of the construction process, linearly dependent on the size $|\mathcal{L}|$ of the initial graph; this leads to the $|\mathcal{L}|$ part of the cost;

(3) the number of valuation functions; this leads to the n part of the cost.

We propose in this chapter some sequential (Section 3.1) and distributed (Section 3.2) strategies devoted to avoid or decrease the effects of these bottlenecks. The induced algorithmic evolutions may lead to similar results as the exhaustive algorithm, but with a lower complexity.

3.1 Heuristics for sequential algorithmics

The bottleneck induced by the valuation functions can be hardly avoided when considering a sequential algorithmics. Then, optimizations mainly rely on heuristic strategies, designed for standard mono-feature BPT construction. These optimizations also allow us to decrease that of MBPTs.

The first way to reduce the computational complexity consists of initiating the process from a partition $\mathcal{L}$ with fewer regions. Instead of using singleton sets —which is equivalent to setting $\mathcal{L}$ isomorphic to Ω—, flat zones or superpixels [56] can be considered. The gain of complexity derives from the reduction of $|\mathcal{L}|$, with the counterpart of an additional cost for superpixel / flat zones computation.

Another solution is to reduce the cost of each elementary step. The resorting of edges within the list(s), after node fusions, can be carried out only after a given number ρ of iterations. This optimization remains marginal. Indeed, these edges still have to be resorted, and the gain concerns the lower size of the list at the resorting time. This optimization influences the $\log_2 |\mathcal{L}|$ part of the computational complexity. In addition, the risks of choosing non-optimal edges for the merging increase linearly with ρ.

An alternative consists of working with lists that do not contain all the edges [57]. This heuristic relies on decreasing the number of edges by only adding in the list(s) those of lower values during the update. This leads to work whith lists of lower sizes, acting on the $|\mathcal{L}| \log_2 |\mathcal{L}|$ term of the cost. This algorithm does not guarantee that the BPT structure will be equivalent to that of a BPT built using the original algorithm.

3.2 Distributed algorithmics

The sequential optimizations of the mono-feature BPT construction affect the result, by providing an approximate BPT compared to the original algorithm. In addition, such heuristics do not tackle the issue of efficiently handling multiple lists, in the case of MBPT. Then, we explore strategies based on distributed algorithmics to build approximate, but close, BPTs.

The time cost of the MBPT construction is mainly due to the space cost of the data-structures, i.e., the number (n) and size ($|\mathcal{L}|$) of the lists

$\mathcal{W}_i$. To reduce this time cost via distributed algorithmics, we have to split this space cost by partitioning the amount of information to be processed. Two alternatives can be considered: either distributing each list on a given computing core, or splitting the graph $\mathfrak{G}$ and distributing each subgraph on these cores.

List-based distribution

By assuming that n cores are used, one list $\mathcal{W}_i$ can be assigned to each core. Then, each core is able to carry out some list-dependent operations, in parallel and without interaction. Such operations are: the removal of an edge from the list; the update of edges impacted by the fusion of the two vertices; the re-sorting of the lists. The speed-up for these operations is linear, and we preserve the same cost as for standard BPT for these parts of the process.

However, the choice of the edge to be removed still requires to share information between the n lists / cores. Then, the distribution of the lists has no speed-up effects on this part of the process.

The properties, advantages, and drawbacks of this distribution strategy are the following:

- the number of algorithmic cores is constant and determined a priori as the number n of lists;
- only the lists are distributed on the cores, while the graph and image structures are shared;
- the distributed algorithm is equivalent to the sequential in terms of result;
- the time complexity of the overall process remains bounded by a n factor, due to the communications between lists for choosing the edge;
- the time cost is however reduced due to the linear speed-up of the other operations on lists.

Graph-based distribution

The dual solution to distributing one list on each core, is to split the image space into p sets of nodes and distribute them onto p cores. Then, each core

still has to handle n lists $\mathcal{W}_i$. The idea is to use the cores to compute little parts of the structure simultaneously and merge them at the end. Such an approach is not obvious even if solutions for similar issues are proposed in [58, 59] for the case of max-trees used in various multi-scale analysis tasks such as in remote sensing [60]. Indeed, for the BPTs, each list can be restricted to the edges of the associated sub-image / sub-graph. The spatial coherence of each split sub-image / sub-graph is crucial. Indeed, at each step, the removal of an edge implies to update and resort the edges in a direct neighbourhood. As a consequence, a regular subdivision of the image into connected regions is mandatory.

This space-partitioning strategy implies to cautiously handle edges shared by two sub-images. This handling is directly linked to the policy for managing the evolution of the partition during the iterative construction of the BPT. Indeed, each core processes its own edges, within its own subgraph. In other words, each core builds a BPT for its sub-image, independently from the other cores. This modus operandi is a fair approximation, in terms of results, of a sequential algorithm. This assertion is notably justified when the distribution of the W_i values of edges is homogeneous over the image. This ensures that each sub-image contains similar values at a same step.

In particular, this is verified for sub-images correctly designed with respect to the evolution of the BPT construction. It is then important to progressively fuse the sub-images and their partial BPTs, and to carry out these mergings just in time. In this context, two approaches can be considered. The first is deterministic and parameterised beforehand. It consists of experimentally assessing the number of iterations required before each merging of a regular hierarchical partitioning. The purpose is to approximate at best a sequential BPT construction by a distributed one. The second is non-deterministic. It consists of merging two subgraphs whenever an edge between two sub-graphs presents a W_i value that is lower than any edge in both sub-graphs.

The properties, advantages, and drawbacks of this distribution strategy are the following:

- the number of algorithmic cores is scalable and requires to be sufficiently high to handle a hierarchical decomposition into sub-images;

- each of the p cores contains n lists, but their size is p times lower than in the sequential case;
- each core contains the subgraph it handles;
- the distributed algorithm is not equivalent to the sequential one in terms of results, as some merging operations are carried out in parallel and without communication;
- however the result can be designed to be similar to that obtained with a sequential approach;
- the speed-up of the process is directly correlated with the parallelization degree over the hierarchy of subgraphs.

In the next chapter (Chapter 4), a first set of validations on remote sensing images emphasizing the interest of the multi-feature generalization of the BPT construction.

3.3 Résumé : Perspectives pour le passage à l'échelle de la construction d'un BPT

L'analyse de complexité présentée précédemment met en évidence trois verrous pour la construction d'un BPT en général, et en particulier pour les MBPTs : (1) la taille des listes, linéairement dépendant de la taille $|\mathcal{L}|$ du graphe initial, entraîne une partie $\log_2 |\mathcal{L}|$ du coût (et une partie $|\mathcal{L}|$ pour la troisième politique); (2) le nombre d'étapes du processus de construction, aussi linéairement dépendant de la taille $|\mathcal{L}|$ du graphe initial, entraîne une partie $|\mathcal{L}|$ du coût; (3) le nombre n de métriques entraîne une partie n du coût. Nous proposons dans ce chapitre des stratégies séquentielles et distribuées visant à réduire les effets de ces verrous. L'évolution algorithmique induite par ces stratégies pourrait mener à l'obtention de résultats similaires à ceux de l'algorithme exhaustive, mais avec une complexité moins élevée.

Heuristiques pour des algorithmes séquentiels Le verrou induit par les métriques est difficile à éviter lorsqu'il s'agit d'algorithmes séquentiels.

Ainsi, les optimisations se basent principalement sur des stratégies heuristiques conçues pour la construction classique d'un BPT. La première façon de réduire la complexité de calcul consiste en l'initialisation du processus à partir d'une partition initiale $\mathcal{L}$ contenant moins de régions. Une autre solution est de réduire le coût élémentaire de chaque étape en ne triant, par exemple, les listes qu'après chaque ρ itérations. Cette optimisation reste marginale mais influence la partie $\log_2 |\mathcal{L}|$ du calcul de complexité. De plus, un tel procédé induit des risques de choix non optimal de couples de nœuds à fusionner. Une alternative consiste à travailler avec des listes qui ne contiennent pas toutes les valeurs correspondant à toutes les arêtes du graphe $\mathfrak{G}$. Cette heuristique diminue le nombre d'éléments à traiter en ne gardant dans les listes que ceux qui ont des valeurs optimales (i.e., faible valeur) pendant la mise à jour. Ces heuristiques ne garantissent pas l'obtention d'un BPT ayant une structure similaire à celle d'un BPT construit à partir de l'algorithme originale.

Algorithmes distribués Les heuristiques séquentielles évoquées ci-avant ne permettent pas de gérer efficacement plusieurs listes, comme dans le cas des MBPTs. Nous explorons alors des stratégies basées sur des algorithmes distribués pour créer des BPTs approchés. Le coût en termes de temps de la construction d'un MBPT est principalement dû au coût spatial de la structure de données (i.e., le nombre n et la taille $|\mathcal{L}|$ des listes $\mathcal{W}_i$). Le partitionnement de la quantité d'information à traiter en divisant l'espace peut réduire ce coût de façon pertinente. La première alternative à considérer consiste à distribuer chaque liste sur un cœur donné du processeur. Chaque cœur peut alors s'occuper de chaque liste parallèlement pour des tâches indépendantes comme la suppression des arêtes, les mises à jours des valeurs et le tri. L'accélération induite par cette optimisation est linéaire et permet d'avoir un coût similaire au BPT classique pour ces tâches. Cependant, le choix des couples de nœuds à fusionner nécessite toujours le partage d'informations entre les n listes / cœurs. Ainsi, la distribution des listes n'a pas d'effets pour cette étape. De plus, seules les listes $\mathcal{W}_i$ sont distribuées sur les cœurs, mais le graphe $\mathfrak{G}$ et la structure de l'image I doivent être partagés. La deuxième alternative consiste à diviser le graphe $\mathfrak{G}$ et distribuer chaque sous-graphe sur un cœur du processeur. En d'autres ter-

mes, l'idée est de diviser l'image I en p imagettes distribuées sur p cœurs. De cette façon, chaque cœur gère n petites listes $\mathcal{W}_i$ correspondant aux imagettes. La difficulté principale d'une telle approche est le regroupement des résultats obtenus sur chaque cœur en un seul bon résultat. Ce genre de procédé n'est pas évident car les listes se retrouvent totalement isolées les unes des autres et les informations sur les bords des imagettes ne sont pas partagées. La cohérence spatiale de chaque imagette / sous-graphe est cruciale. En effet, des mises à jours, nécessitant des informations du voisinage direct, sont effectuées à chaque itération dans l'algorithme. En conséquence, une division régulière de l'image en respectant la connexion entre régions est obligatoire. Cette stratégie de partitionnement de l'espace implique de la vigilance pour la gestion des arêtes qui sont partagées par deux imagettes. Chaque cœur construit, indépendamment des autres cœurs, un petit BPT pour l'imagette correspondante. Ce mode opératoire est une approximation de l'algorithme séquentiel en termes de résultats. Cette supposition est probablement justifiée quand la distribution des valeurs W_i de chaque arête est homogène à travers l'image. Cela assure le fait que chaque imagette contienne des valeurs similaires pour chaque étape. Il est alors important de fusionner progressivement les informations des imagettes et de leur BPT partiel. Une première approche consiste à procéder à cette fusion après un nombre donné d'itérations. Une seconde consiste à fusionner deux sous-graphes dès qu'ils contiennent une arête commune dont la valeur W_i est optimale pour une fusion de nœuds potentielle. Il est évident que le résultat d'une telle heuristique n'est pas similaire à celui d'un algorithme classique mais pourrait être conçu pour être proche.

Chapter 4

Experimental studies

Contents

To illustrate our framework, two application cases have been considered in the domain of remote sensing image analysis. Our purpose is to highlight the versatility of the multi-feature framework by demonstrating how it can be used to build BPTs from either multiple images and / or multiple metrics computed through the image content. In Section 4.1, we show how a multi-feature BPT can be built from a single complex satellite image by considering simultaneously, and in a consensual manner, various metrics. In Section 4.2, we show how a multi-feature BPT can be built from multiple satellite images sensed over the same geographical area by considering simultaneously information provided by different image contents.

The BPT construction and segmentation approaches were voluntarily chosen as very simple, in order to avoid any bias related to these choices, thus better focusing on the actual structural effects of multi-feature BPT versus standard BPT. These experiments then have to be considered as illustrative examples, since neither quantitative validation nor fine parameter tuning were carried out. Our purpose is mainly to give the intuition

of potential uses of such BPTs in complex imaging domains.

In this preliminary study, we present only a qualitative evaluation of potential segmentation results obtained from MBPTs. The next part of this manuscript will focus on supervised Binary Partition Tree quality evaluation and will present the quantitative evaluation of MBPTs.

4.1 Illustrative example 1: Multi-criteria segmentation

The segmentation of very-high spatial resolution (VHSR) satellite images is a challenging task since the latest generation of images presents high spectral and spatial resolution properties, leading to huge volumes of data. In this context, the segmentation of satellite images using classical monometric BPTs has already been widely studied [43, 44, 45, 47, 48]. This motivates in particular the experimentation of multi-criteria segmentation procedures on such images.

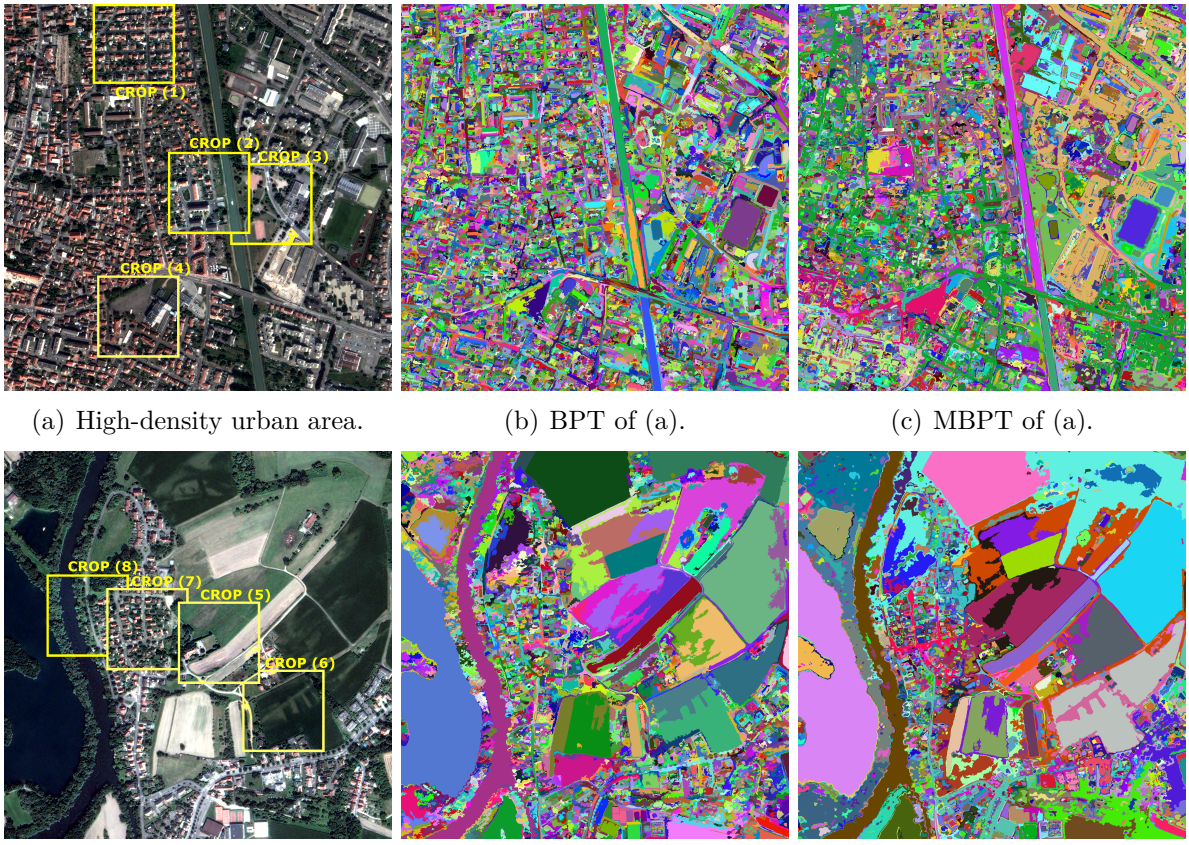
Data

The dataset used here (courtesy LIVE, UMR CNRS 7263) was sensed over the town of Strasbourg (France) by the PLÉIADES satellite in 2012. The first sample is a VHSR image ($2\,000 \times 2\,000$ pixels) representing a complex high-density urban area (Figure 4.1(a)) composed of different urban objects (e.g., individual houses, industrial buildings, parking lots, roads, shadows, water canals). The second sample is also a VHSR image ($2\,000 \times 2\,000$ pixels) but that represents a typical low-density urban area (Figure 4.1(d)) composed of different geographical objects (e.g., crop fields, forests, bare soils, rivers). These two satellite images are pansharpened multispectral images at a spatial resolution of 60 cm with four spectral bands (R, G, B, NIR)[1].

Method

To reduce the spatial complexity of this approach, the BPTs are built from an initial partition $\mathcal{L}$ composed of 200 000 regions obtained from a cut

[1] R, G, B, NIR stand respectively for the spectral reflectance measurements acquired in the red, green, blue and mear-infrared.



(a) High-density urban area. (b) BPT of (a). (c) MBPT of (a).

(d) Low-density urban area. (e) BPT of (d). (f) MBPT of (d).

Figure 4.1: (a, d) Illustration of two very-high spatial resolution satellite images (2 000 × 2 000 pixels) at a spatial resolution of 60 cm sensed by the PLÉIADES satellite and covering different geographical areas. (b, resp. e) Segmentation result from a cut containing 23 500 regions (resp. 5 000 regions) performed on a "standard" BPT of (a, resp. d). (c, resp. f) Segmentation result from a cut containing 23 500 regions (resp. 5 000 regions) performed on a MBPT of (a, resp. d) using four features: the colour intensity, the region elongation, NDVI and NDWI values $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$.

performed on a "standard" BPT. Note that any other super-pixel approach may also have been considered for such initialization.

We considered here four complementary valuation functions $W_\star : (2^\Omega)^2 \times \mathbb{V}^\Omega \to \mathbb{R}$ that model radiometrical or geometrical information related to region dissimilarities.

The first valuation function W_{colour} is defined as the increase of the ranges of the pixel intensity values for each radiometric band, potentially induced by the fusion of incident regions (i.e., a virtual merge). Practically, W_{colour} can be computed as follows. Let $N_i, N_j \in \mathcal{P}$ be two distinct and adjacent regions / nodes. A multispectral image is modelled as a function $\mathcal{I} : \Omega \to \mathbb{V}$ which associates, to each point $x \in \Omega$, a s-uple (with $s > 1$ the

number of spectral bands) of spectral intensities $\mathcal{I}(x) = \prod_{b=1}^{s} \mathcal{I}_b(x)$. The W_{colour} valuation function is then computed as

$$\frac{1}{s} \sum_{b=1}^{s} \max\{v_b^+(N_i), v_b^+(N_j)\} - \min\{v_b^-(N_i), v_b^-(N_j)\} \tag{4.1}$$

where $v_b^\star$ provides the extremal values for the b-th spectral band in $\mathcal{I}$ (i.e., in $\mathcal{I}_b$).

The second valuation function W_{ndvi} quantifies the difference of NDVI between two adjacent regions. The NDVI (Normalized Difference Vegetation Index) of a pixel is a simple indicator that can be used to analyze remote sensing data and assess whether the target being observed contains live green vegetation or not. It is simply computed as a function that associates to each point $x \in \Omega$ the ratio $(I_{NIR}(x) - I_R(x))/(I_{NIR}(x) + I_R(x))$ involving both the R and NIR channels. Practically, the W_{ndvi} valuation function is computed for two adjacent nodes N_i, N_j as the absolute difference between the averages of NDVI associated to the pixels of the two regions.

The third valuation function W_{ndwi} computes the difference of NDWI (Normalized Difference Water Index) between two adjacent regions. The NDWI of a pixel is an indicator to assess whether the target being observed contains presence of water or not. The W_{ndwi} valuation function is computed similarly to the W_{ndvi} valuation function by replacing the R channel by the G one.

The fourth valuation function W_{elong} is defined as the change of the geometrical elongation values, potentially induced by the fusion of two incident regions. The geometrical elongation of a node N is computed as the ratio of the height and width of its bounding box and is denoted as $elong(N)$. Practically the W_{elong} valuation function is computed for two adjacent nodes N_i, N_j as

$$|elong(N_i \cup N_j) - \frac{elong(N_i) + elong(N_j)}{2}| \tag{4.2}$$

In the case of multi-criteria BPTs, the relative local information consensus policy *mean-of-ranks*, according to the position of the edges within the lists is applied for the first 15% of the lists $\mathcal{W}_\star$.

As a baseline, the "standard" BPTs of the satellite images of Fig-

ure 4.1(a, d) are constructed using the colour intensity value criterion W_{colour}. These trees are then segmented by considering a user-defined horizontal cut to produce a partition whose the region scales are adapted to segment the various objects contained in the sensed geographical areas (e.g., individual houses, buildings, roads, forests), see Figure 4.1(b, e). For visualisation purpose, the segmentation results are depicted here in random false colours.

In order to evaluate the impact of the different valuation functions $W_\star$ on the segmentation results, the multi-criteria BPTs are then built by considering various combinations of valuation functions (e.g., $\{W_{colour}, W_{ndvi}\}$, $\{W_{colour}, W_{elong}, W_{ndvi}\}$, $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$). The produced multi-criteria trees are then segmented in the same way as for the "standard" BPTs, leading to the same number of regions. For illustration purpose on the entire images, Figure 4.1(c, f) shows some examples of the best results obtained from our multi-criteria BPTs.

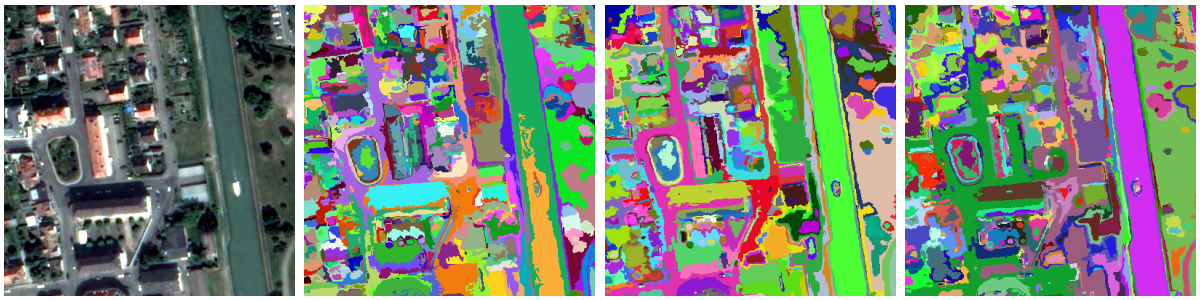
Results

From the "standard" BPT results (see Figure 4.1(b, e)), we observe that the obtained regions are quite radiometrically homogeneous and are adapted to extract simple urban objects (e.g., small road segments, bare soils). Concerning more complex objects strongly structured by their geometrical shapes (e.g., house roofs, rivers, large vegetation areas and roads with elongated structures), the regions produced with the "standard" BPT are not always relevant since the considered urban objects are often composed of several small sub-regions (see the river, road sections and vegetation areas in Figure 4.1(b, e)). In comparison, the cut extracted from the MBPT enables to directly gather, in a same partition, regions corresponding both to simple urban objects and to complex ones (see the previous mentioned geographical objects in Figure 4.1(c, f)).

To facilitate the qualitative analysis of the obtained results, we focus hereinafter on eight illustrative image crops (from Figure 4.2 to Figure 4.9) extracted from the two satellite images of the dataset. Note that the BPTs and MBPTs are still computed from the entire image contents (Figure 4.1(a, d)) and are still segmented in the same way to provide cuts with an equal number of regions. Figure 4.2(a) presents a typical exam-

(a) Image crop (1). (b) BPT. (c) MBPT-1. (d) MBPT-2.

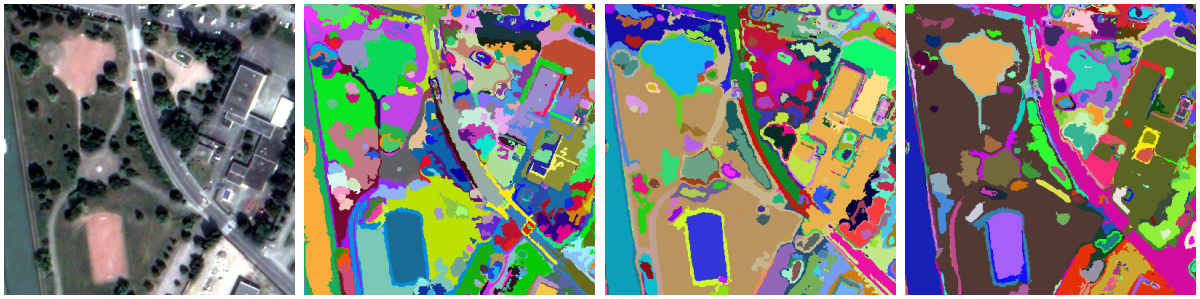
Figure 4.2: Segmentation results from the BPTs and the MBPTs, centered on crop (1) of the image presented in Figure 4.1(a). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndvi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$.



(a) Image crop (2). (b) BPT. (c) MBPT-1. (d) MBPT-2.

Figure 4.3: Segmentation results from the BPTs and the MBPTs, centered on crop (2) of the image presented in Figure 4.1(a). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndwi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$.

ple of an urban area composed of structured complex urban blocks. In this example, we observe that the segmentation result provided by the "standard" BPT (Figure 4.2(b)) does not enable to reconstruct these urban blocks since they often appear composed of several small sub-regions (see, in particular, the road segments). By adding the valuation function W_{ndvi} in the creation process of our multi-criteria BPT (Figure 4.2(c)), we observe that the extracted cut permits to regroup connected regions having similar NDVI values (objects containing vegetation pixels or not). In this result, the garden areas between houses (see centre bottom of the crop (1)) are gathered in larger segments and the road sections are better connected than before. This last phenomenon is also easily observable in Figure 4.2(d) where the cut was performed on a multi-criteria BPT involving four valuation functions $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$. In this example, the elongation criterion permitted to produce compact and large



(a) Image crop (3). (b) BPT. (c) MBPT-1. (d) MBPT-2.

Figure 4.4: Segmentation results from the BPTs and the MBPTs, centred on crop (3) of the image presented in Figure 4.1(a). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndvi}, W_{ndwi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}\}$.

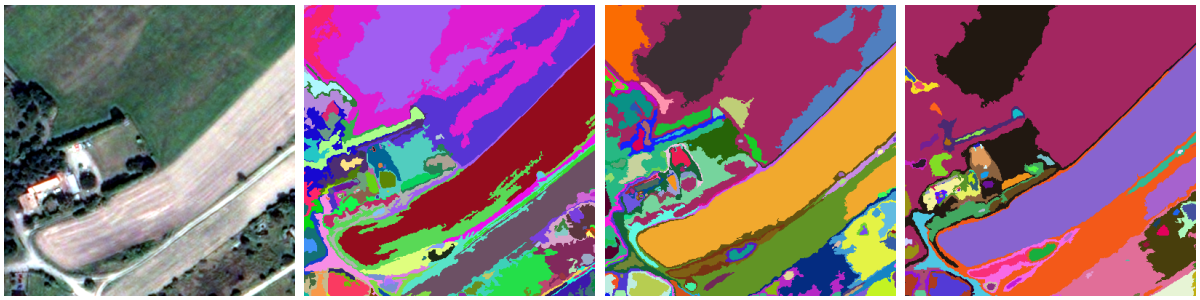
road sections.

In Figure 4.3, it can be noticed that the river, at the right side of the crop (2), is divided by three segments in the result obtained from the "standard" BPT Figure 4.3(b). By considering the W_{ndwi} valuation function during the MBPT creation process (see segmentation result in Figure 4.3(c)), the river corresponds to only one segment. Indeed, this valuation function enables to quantify the similarity between neighbouring regions containing water or not. We also observe that the segments of road are also less disconnected since they do not contain water but they present similar (low) NDWI values. The best result for the segmentation of both simple and complex objects is shown in Figure 4.3(d) where four valuation functions are involved $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$. Here, the road sections are well delineated, the river belongs to one segment and the vegetation area (at the right side of the crop (2)) looks more homogeneous. Our assumption is that complex urban objects appearing in the image content as either homogeneous and elongated can be extracted from a cut of the multi-criteria BPT thanks to the consensus made between the different criteria during the MBPT creation.

In comparison to what we observe in Figure 4.4(b), which is a result from the "standard" BPT, the use of the W_{ndvi} and the W_{ndwi} valuation functions together in Figure 4.4(c) enhances the segmentation of both the vegetation areas and the bare soils in an urban area. Adding the W_{elong} metric in Figure 4.4(d) also helped the creation of the curved road and the small elongated group of trees at the right side of the crop (3). However, some

(a) Image crop (4). (b) BPT. (c) MBPT-1. (d) MBPT-2.

Figure 4.5: Segmentation results from the BPTs and the MBPTs, centred on crop (4) of the image presented in Figure 4.1(a). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndvi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}\}$.



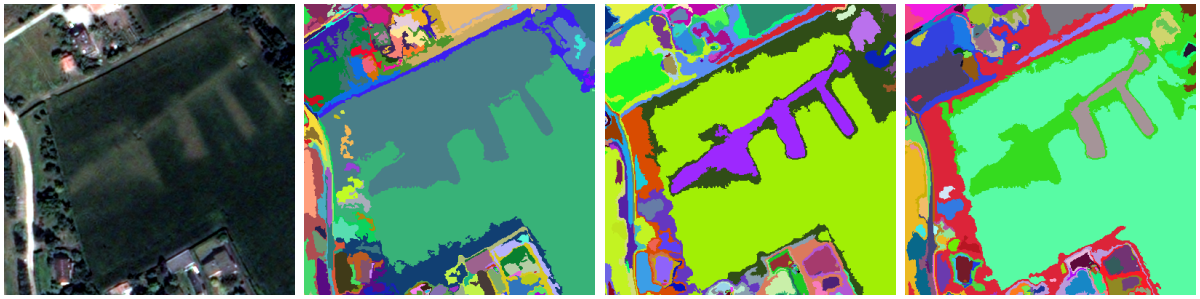
(a) Image crop (5). (b) BPT. (c) MBPT-1. (d) MBPT-2.

Figure 4.6: Segmentation results from the BPTs and the MBPTs, centred on crop (5) of the image presented in Figure 4.1(d). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndvi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$.

geographical objects having certainly similar NDVI and / or NDWI values are mixed up together although they have different semantics. Indeed, the building roofs at the right side of the crop (3) are merged with the parking lot.

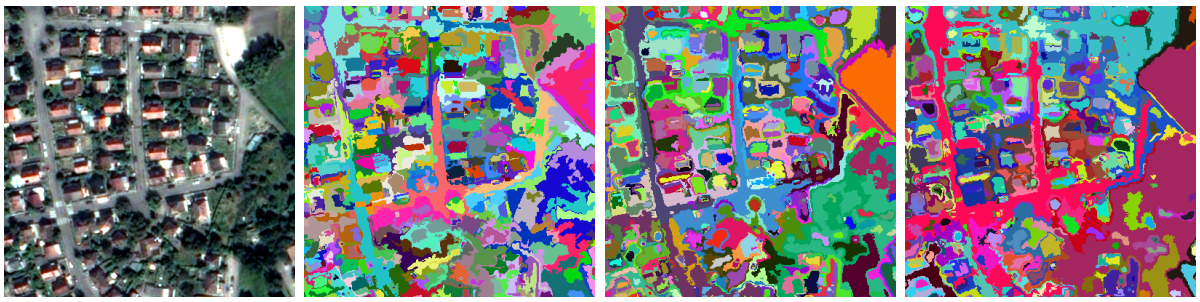
The same phenomenon is observed in Figure 4.5(c, d) where the triangular bare soil, the road sections and the vegetation areas are visually better segmented with the MBPTs than with the "standard" BPT (see Figure 4.5(b)).

The image crops extracted from the low-density urban area image (Figure 4.1(d)) represent illustrative examples of vegetation areas, rivers and scattered urban objects. Figure 4.6(c) and Figure 4.7(c) highlight the ability of the W_{ndvi} valuation function to regroup objects containing live green vegetations. In these examples the wide vegetation areas are gathered in larger segments while they were divided in several sub-regions in the results



(a) Image crop (6). (b) BPT. (c) MBPT-1. (d) MBPT-2.

Figure 4.7: Segmentation results from the BPTs and the MBPTs, centred on crop (6) of the image presented in Figure 4.1(d). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndvi}\}$; MBPT-2: $\{W_{colour}, W_{ndvi}, W_{ndwi}\}$.

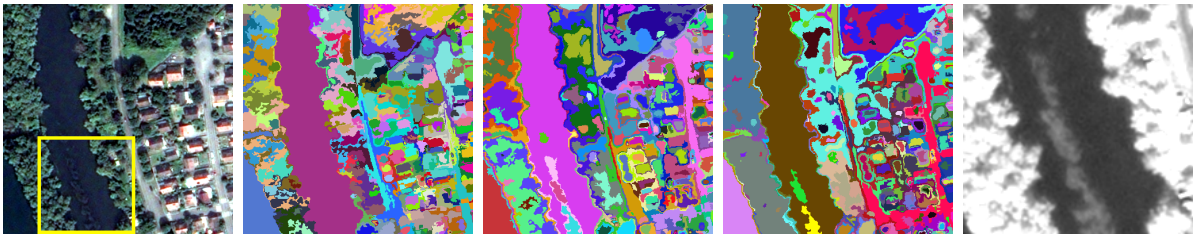


(a) Image crop (7). (b) BPT. (c) MBPT-1. (d) MBPT-2.

Figure 4.8: Segmentation results from the BPTs and the MBPTs, centred on crop (7) of the image presented in Figure 4.1(d). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndvi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$.

obtained from the "standard" BPTs (see Figure 4.6(b) and Figure 4.7(b)). The consensual results obtained from the use of more valuation functions in the MBPTs (see Figure 4.6(d) and Figure 4.7(d)) enable to deal with the colour heterogeneity between the different crop fields.

Figure 4.8(a) represents an area composed of urban objects mixed up with some complex vegetation zones. The road section are visible in the result from the "standard" BPT (see Figure 4.8(b)) but the vegetation areas at the right side of the crop (7) are divided in numerous small regions. From the results of the multi-feature BPT involving both the W_{colour} and the W_{ndvi} valuation functions (Figure 4.8(c)), we observe that the vegetation areas and the road sections form now better connected regions. The result obtained from adding the W_{elong} metric to the MBPT (Figure 4.8(d)) presents a balanced consensual segmentation where road sections can be easily recognized and the vegetation areas are more homogeneous.



(a) Img. crop (8). (b) BPT. (c) MBPT-1. (d) MBPT-2. (e) NIR band.

Figure 4.9: Segmentation results from the BPTs and the MBPTs, centred on crop (8) of the image presented in Figure 4.1(d). BPT: $\{W_{colour}\}$; MBPT-1: $\{W_{colour}, W_{ndwi}\}$; MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$; NIR band: crop of the river captured by the NIR band, and showing non-obvious information.

Finally, the result obtained from the "standard" BPT, in Figure 4.9(b), shows that the vegetation areas at the left part of the crop (8) are divided in several small sub-regions, the river seems to be perfect and the urban zone is over-segmented. In comparison, the result obtained from a multi-feature BPT using the W_{ndwi} valuation function offers an unexpected observation. The river, that seems to be totally homogeneous in Figure 4.9(a), is segmented differently in Figure 4.9(c, d). In fact, the result obtained for the multi-feature BPTs helped to highlight some vegetation structures that cannot be easily interpreted by human observation. Indeed, the river is composed of small sandbanks covered by vegetation tissues that could be clearly observed when we focus on the near infrared band (see the zoom in the Figure 4.9(e)). By combining four valuation functions $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$ in Figure 4.9(d), a consensual segmentation result shows us a pertinent balance between the vegetation areas that are quite homogeneous, the urban area where the road sections are well structured, and the complex content of the river that we observed earlier.

4.2 Illustrative example 2: Multi-image segmentation

We now illustrate the interest of multi-criteria BPT for multi-image segmentation in the context of satellite imaging.

Data

The dataset used here is a time serie of images ($1\,000 \times 1\,000$ pixels) sensed over an area located near Toulouse (France). This area is a typical agricultural zone composed of different types of crop fields and vegetation. Images were acquired by the FORMOSAT-2 satellite over the 2007 cultural year. They were ortho-rectified and have a spatial resolution of 8 m, with four spectral bands (R, G, B, NIR). From this dataset, we selected four images (Figure 4.10) all acquired late August and September to reduce as much as we can the temporal evolution effects. Note that the second image is partially affected by the presence of clouds. The main purpose of this experiment is actually to assess the ability of our multi-feature BPT framework to capture time-independent and redundant information from the contents of multiple images representing the same scene.

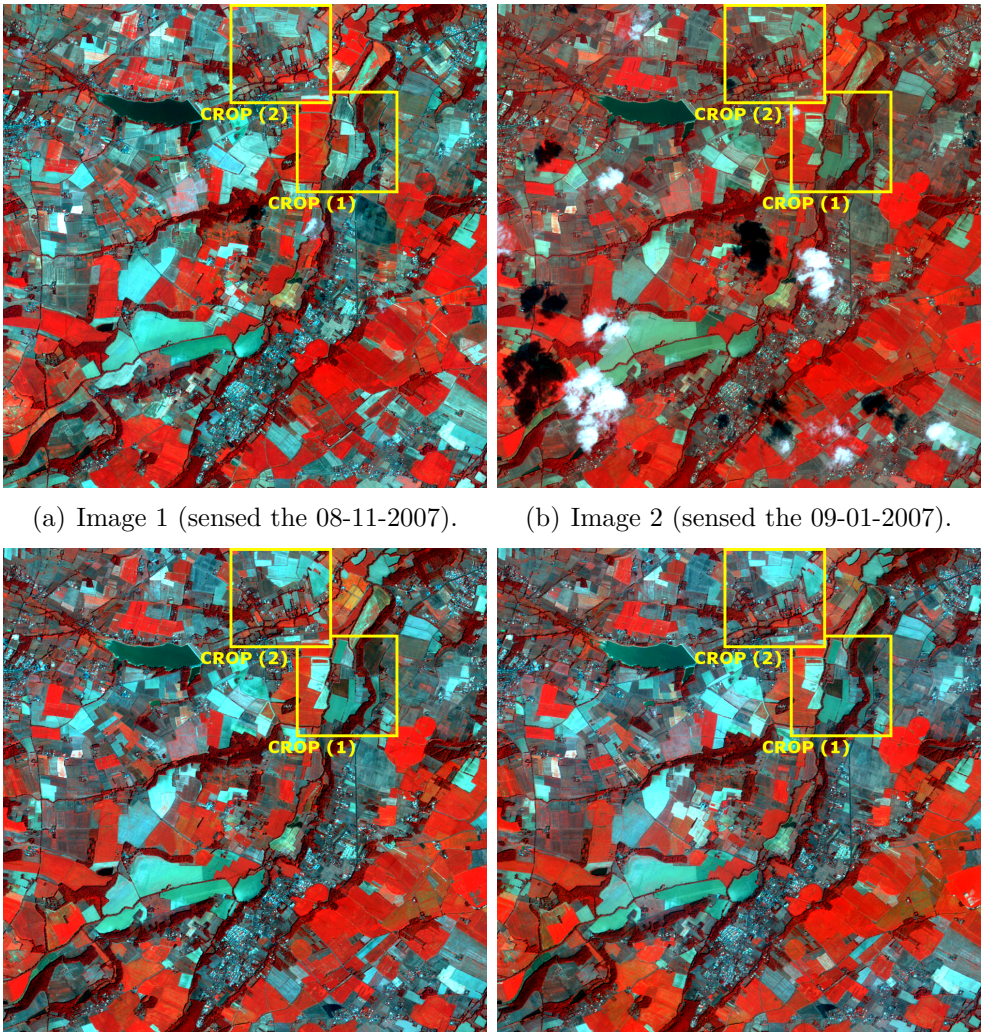
Method and results

To reduce the spatial complexity of this approach, the BPTs are built from an initial partition $\mathcal{L}$ composed of 200 000 regions obtained from a cut performed on a "standard" BPT. This initial partition was produced from the Image 1 (Figure 4.10(a)) that is not affected by the presence of cloud.

In the case of multi-image BPTs, the relative local information consensus policy *mean-of-ranks*, according to the position of the edges within the lists is applied for the first 15% of the lists $\mathcal{W}_\star$.

As a baseline, a "standard" BPT is constructed for each satellite image presented in Figure 4.10 by considering the intensity value criterion W_{colour} (denoted hereinafter as W_{colour}^{img1}, W_{colour}^{img2}, W_{colour}^{img3}, W_{colour}^{img4} depending on the image where this criterion is computed). These binary trees are then segmented by using a user-defined horizontal cut to produce a partition whose the region scales are adapted to segment the various geographical objects covering the sensed areas (e.g., agricultural crop fields, wide forest areas), see Figure 4.11(f, g, h, i) and Figure 4.12(f, g, h, i). For visualisation purpose, the segmentation results are depicted here in random false colours on illustrative image crops.

To evaluate the impact of the different valuation functions on the segmentation results, the multi-image BPTs are built by considering simul-



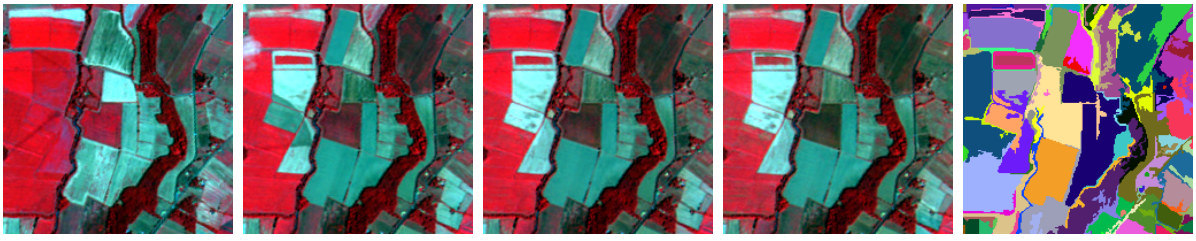
(a) Image 1 (sensed the 08-11-2007). (b) Image 2 (sensed the 09-01-2007).

(c) Image 3 (sensed the 09-08-2007). (d) Image 4 (sensed the 09-15-2007).

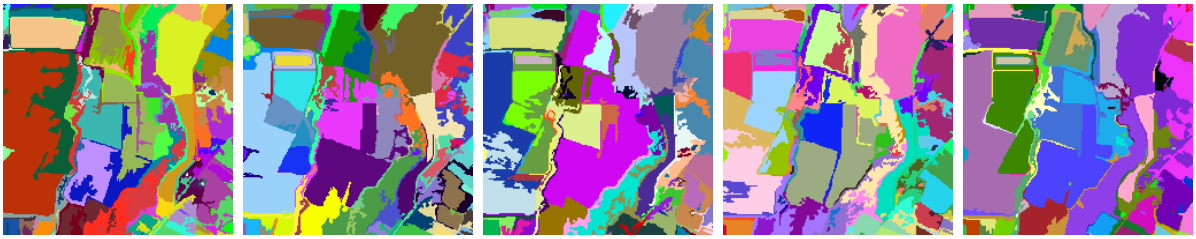
Figure 4.10: Illustration of four satellite images (1000×1000 pixels) at a spatial resolution of 8 m sensed by the FORMOSAT-2 satellite covering the same geographical area.

taneously information extracted from the contents of the four satellite images (e.g., $\{W_{colour}^{img1}, \ldots, W_{colour}^{img3}\}$, $\{W_{colour}^{img1}, W_{colour}^{img2}, W_{ndvi}^{img1}, W_{ndvi}^{img2}\}$). The produced multi-image trees are then segmented in the same way as for the "standard" BPTs, leading to the same number of regions.

For a better visualization, we focus in the following on two image crops representing different parts of the geographical scene. In the crop (1) represented by Figure 4.11(a, b, c, d), we observe that, although some variations are due to the time evolutions, the main structures of the crop fields do not evolve. After a visual interpretation of these images, it can be noticed that the majority of observed geographical areas preserve the same crop field partition. However, each segmentation obtained from the "standard"



(a) Image 1. (b) Image 2. (c) Image 3. (d) Image 4. (e) MBPT-1.



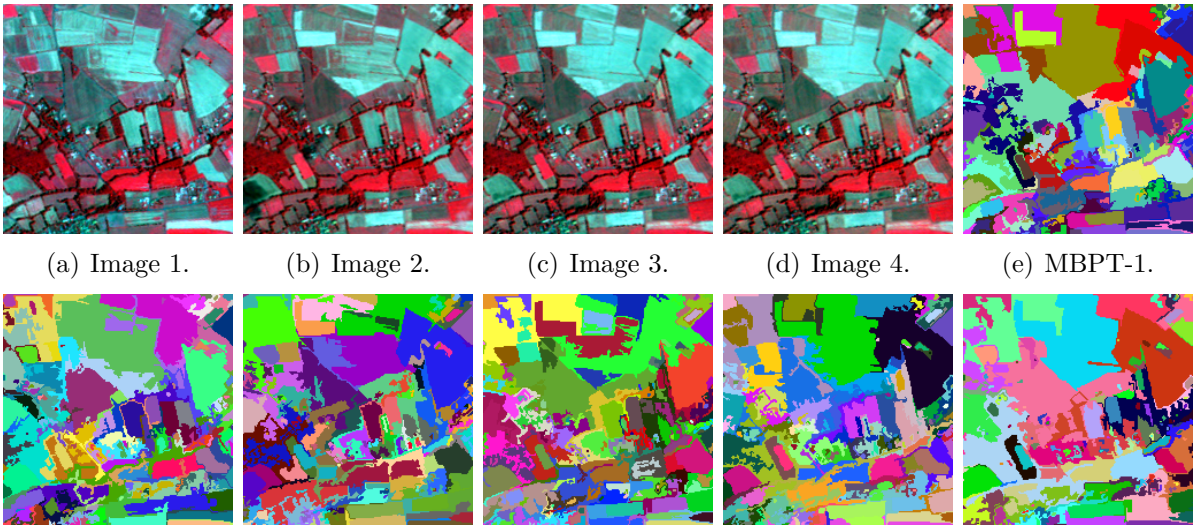
(f) BPT-1. (g) BPT-2. (h) BPT-3. (i) BPT-4. (j) MBPT-2.

Figure 4.11: Segmentation results from the mono-image BPTs and the multi-images/metrics MBPTs. (a) Crop 1 from Image 1 (Figure 4.10(a)). (b) Crop 1 from Image 2 (Figure 4.10(b)). (c) Crop 1 from Image 3 (Figure 4.10(c)). (d) Crop 1 from Image 4 (Figure 4.10(d)). (f) Segmentation result from a BPT of Image 1. (g) Segmentation result from a BPT of Image 2. (h) Segmentation result from a BPT of Image 3. (i) Segmentation result from a BPT of Image 4. (e) Segmentation result from a MBPT with 4-images/1-metric $\{W^{img1}_{colour}, W^{img2}_{colour}, W^{img3}_{colour}, W^{img4}_{colour}\}$. (j) Segmentation result from a MBPT with 3-images/2-metrics $\{W^{img1}_{colour}, \ldots, W^{img3}_{colour}, W^{img1}_{ndvi}, \ldots, W^{img3}_{ndvi}\}$.

BPT of each image presents totally different partitions (see Figure 4.11(f, g, h, i)). Indeed, some particular crop fields (centre of crop (1)) are always divided in various numbers of regions. The same phenomenon can be observed with the elongated vegetation area that is sometimes merged with the neighbouring crop fields. In comparison, the results obtained from our multi-image BPT (see Figure 4.11(e)) are more stable since they arise from a consensual discussion between the results of the "standard" BPT results.

The same behaviour can be observed in the second crop (Figure 4.12) where the results obtained from the "standard" BPT of each image (see Figure 4.12(f, g, h, i)) are almost totally different but most important information are gathered in the result of our multi-image BPT (see Figure 4.12(e)).

The best results are shown in Figure 4.11(j) and Figure 4.12(j) where we used three of the four images with two metrics $\{W^{img1}_{colour}, \ldots, W^{img3}_{colour}, W^{img1}_{ndvi}, \ldots, W^{img3}_{ndvi}\}$. The regions corresponding to crop fields are more homogeneous and the number of sub-regions



(a) Image 1. (b) Image 2. (c) Image 3. (d) Image 4. (e) MBPT-1.

(f) BPT-1. (g) BPT-2. (h) BPT-3. (i) BPT-4. (j) MBPT-2.

Figure 4.12: Segmentation results from the mono-image BPTs and the multi-images/metrics MBPTs. (a) Crop 2 from Image 1 (Figure 4.10(a)). (b) Crop 2 from Image 2 (Figure 4.10(b)). (c) Crop 2 from Image 3 (Figure 4.10(c)). (d) Crop 2 from Image 4 (Figure 4.10(d)). (f) Segmentation result from a BPT of Image 1. (g) Segmentation result from a BPT of Image 2. (h) Segmentation result from a BPT of Image 3. (i) Segmentation result from a BPT of Image 4. (e) Segmentation result from a MBPT with 4-images/1-metric $\{W_{colour}^{img1}, W_{colour}^{img2}, W_{colour}^{img3}, W_{colour}^{img4}\}$. (j) Segmentation result from a MBPT with 3-images/2-metrics $\{W_{colour}^{img1}, \ldots, W_{colour}^{img3}, W_{ndvi}^{img1}, \ldots, W_{ndvi}^{img3}\}$.

of each area is well adapted for a future object-based recognition or classification step.

In this chapter, we proposed first qualitative validations of the proposed framework for MBPT. Indeed, we do not aim at developping a whole segmentation method but a new framework for the segmentation search space reduction (i.e., a new way to build a BPT in a multi-feature way). Hence, it does not seem to be fair to compare our framework directly with other image segmentation methods. For a quantitative validation, a comparison between comparable structures would be more interesting (i.e., BPT vs. MBPT). However, an important question to ask is about the way to proceed such comparison. In the next part, we propose a supervised BPT quality evaluation that will allow us to compare our framework to the one for standard BPT.

4.3 Résumé : Études expérimentales

Pour illustrer notre approche de construction de BPT multicritère, nous considérons deux cas d'application dans le cadre de l'analyse d'images de télédétection. Notre but est de mettre en évidence la versatilité de l'approche multicritère en montrant son utilité pour la construction de MBPTs à partir d'une / plusieurs image(s) et d'une / plusieurs métrique(s). Dans un premier temps, nous montrons comment construire des BPTs multicritères à partir d'une image satellite complexe en considérant simultanément, et de façon consensuelle, plusieurs métriques. Dans un deuxième temps, nous montrons comment construire des BPTs multicritères à partir de plusieurs images satellites représentant une même scène, en considérant simultanément les informations contenues dans chaque image.

La construction des BPTs et la méthode de segmentation ont été volontairement choisies d'une façon simple, afin d'éviter des biais liés à ces choix. Ainsi, nous voulons nous concentrer sur les effets structurels des BPT multicritères face aux BPTs classiques. Ces premières expérimentations doivent être considérées comme des exemples illustratifs, puisque ni une validation quantitative ni une étude sur le réglage des paramètres n'ont été réalisées. Notre but est de donner principalement des intuitions sur l'utilité potentielle des BPT multicritères pour l'analyse et le traitement d'images complexes.

Dans cette étude préliminaire, nous présentons une évaluation qualitative des résultats potentiels que l'on peut obtenir à partir des MBPTs. La deuxième partie de ce manuscrit se concentrera sur l'évaluation supervisée de la qualité de structures hiérarchiques comme les BPTs et les MBPTs, aidant à l'évaluation quantitative de notre approche multicritère de construction.

Exemple illustratif 1 : segmentation multicritère La segmentation d'images satellites à très haute résoluation présente un défi depuis que les nouvelles générations d'images ont des résolutions spectrales et spatiales élevées menant à des données volumineuses. Dans ce contexte, la segmentation d'images satellites utilisant des BPTs mono-métriques a déjà été largement étudiée. Cela nous motive à effectuer des expérimentations de procédures de segmentation multicritère sur de telles images. L'ensemble de

données utilisé ici (fournies grâcieusement par le LIVE, UMR CNRS 7263) contiennent des images ($2\,000 \times 2\,000$ pixels) de la ville de Strasbourg (France) captées par les satellites PLÉIADES en 2012. Le premier échantillon d'image représente une région à forte densité urbaine composée de différents objets urbains comme des maisons individuelles, des bâtiments industriels, des parkings, des routes, des ombres et des canaux. Le second échantillon représente une région typique à faible densité urbaine composée de différents objets géographiques comme des champs, des forêts, des sol nus et des rivières. Afin de réduire la complexité de l'approche, les (M)BPTs sont construits à partir d'une partition initiale $\mathcal{L}$ composée de 200 000 régions. Nous considérons dans ces expérimentations quatre métriques. La métrique W_{colour} est basée sur l'intensité radiométrique des pixels pour chaque bande spectrale. La métrique W_{ndvi} quantifie la différence des valeurs de l'indice NDVI de deux régions adjacentes vis-à-vis de la présence de végétation. La métrique W_{ndwi} quantifie la différence des valeurs de l'indice NDWI de deux régions adjacentes vis-à-vis de la présence d'eau. Pour la construction de BPTs multicritères, la politique de consensus d'information relative locale *mean-of-ranks* est appliquée pour les 15% des têtes de listes. Comme base, un BPT classique est construit en utilisant la métrique W_{colour} à partir de chaque image. Un processus de segmentation des images a ensuite été effectué en appliquant une coupe horizontale à chaque BPT afin d'obtenir des partitions contenant des régions ayant des tailles raisonnables par rapport aux objets d'intérêt comme les maisons, les bâtiments, les routes et les forêts. Afin d'évaluer l'impact des métriques sur les résultats de segmentations, les MBPTs ont été construits à partir de plusieurs combinaisons (e.g., $\{W_{colour}, W_{ndvi}\}$, $\{W_{colour}, W_{elong}, W_{ndvi}\}$, $\{W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}\}$). Par la suite, la même coupe effectuée sur les BPTs classiques a été appliquée sur les BPT pour disposer d'un nombre de régions similaire. D'un point de vue qualitatif, les MBPTs permettent d'extraire des objets d'intérêt hétérogènes et complexes, contrairement aux BPTs qui peinent à fournir une partition où cohabitent ces types d'objets. En effet, les MBPTs sont capables de fournir une partition mettant en évidence divers objets suggérés par chaque métrique utilisée dans le processus de construction. En d'autres termes, si la métrique W_{colour} détecte facilement les bâtiments, si la métrique W_{elong} favorise les routes et si la

métrique W_{ndvi} met en évidence les végétaux, le MBPT construit à partir de la combinaison $\{W_{colour}, W_{elong}, W_{ndvi}\}$ est capable de fournir une partition au regard de ces types d'objets d'intérêt. Par ailleurs, la structure hiérarchique des MBPTs dépend fortement du choix de ces métriques ainsi que du choix de la stratégie de consensus. Cela dépend également de l'étude que l'utilisateur souhaite mener.

Exemple illustratif 2 : segmentation multi-images Des illustrations dans le contexte de l'imagerie satellite montrent aussi l'intérêt de l'approche de construction de BPTs multicritères à partir de plusieurs images. L'ensemble de données utilisée ici correspond à une série temporelle d'images ($1\,000 \times 1\,000$ pixels) captée aux alentours de la région de Toulouse (France) par le satellite FORMOSAT-2 au cours de l'année agricole de 2007. Nous avons sélectionné quatre images capturées vers la fin du mois d'août et le début du mois de septembre afin de reduire autant que possible les effets de l'évolution temporelle. Notre but n'est pas d'effectuer une analyse temporelle mais de montrer l'utilité de notre approche multicritère pour la construction de MBPTs à capturer des informations redondantes et indépendantes du temps, présentes dans plusieurs images d'une même scène. Afin de réduire la complexité spatiale de l'approche, les (M)BPT sont construits à partir d'une partition initiale $\mathcal{L}$ composée de 200 000 régions. Pour la construction des MBPTs à partir de plusieurs images, la stratégie de consensus choisie correspond à la stratégie *mean-of-ranks* appliquée sur les 15% des têtes des listes. Comme base, un BPT classique est construit à partir de chaque image en utilisant la métrique W_{colour}. Un processus de segmentation par application d'une coupe horizontale sur ces BPTs a été effectué par la suite, fournissant des partitions dans lesquelles les tailles des régions sont adaptées à des objets géographiques tels que les champs, les forêts et les rivières. Ensuite, nous avons construit des MBPTs en utilisant la métrique W_{colour} et en variant le nombre d'images (e.g., $\{W^{img1}_{colour}, \ldots, W^{img3}_{colour}\}$, $\{W^{img1}_{colour}, W^{img2}_{colour}, W^{img1}_{ndvi}, W^{img2}_{ndvi}\}$). Afin d'évaluer l'impact de la construction de multi-image, une coupe similaire à celle effectuée sur les BPTs classiques est appliquée sur ces MBPTs. Visuellement, les quatre images présentent des structures similaires. Cependant, les partitions obtenues à partir des BPTs classiques de chaque image présentent des structures to-

talement différentes pour un même nombre de régions. Les résultats de segmentation obtenus à partir des MBPTs semblent plus stables, plus homogènes et mettent, visuellement, en évidence la structure que l'on observe sur les images. La variation du nombre d'images pour une même métrique montre l'intérêt de l'adéquation des informations redondantes de chaque image pour l'amélioration des résultats de segmentation. Le meilleur résultat obtenu correspond à une partition provenant d'un MBPT issu de trois images et de deux métriques. Cela montre la pertinence de l'utilisation de l'approche multicritère proposée pour la création de MBPTs.

Part II

Supervised Binary Partition Tree quality evaluation

Afin d'évaluer l'intérêt de l'approche proposée pour la construction de BPT multicritère, il est nécessaire de comparer les MBPTs aux BPTs classiques. Ce besoin induit de nouvelles questions concernant l'évaluation de la qualité d'une représentation hiérarchique d'images (e.g., BPT, MBPT) à fournir des résultats de segmentation pertinents. Dans cette partie, nous introduisons d'abord l'impact de la qualité des BPTs sur la segmentation d'image, les travaux connexes et nos motivations pour cette recherche. Ensuite, nous proposons une approche supervisée évaluant la qualité d'une structure hiérarchique dans le contexte de la segmentation d'images. Nous développons deux parties dans cette approche. Nous proposons à la fois une analyse intrinsèque et une analyse extrinsèque de la structure hiérarchique du BPT, s'appuyant sur des exemples de vérité-terrain définis par l'expert du domaine (i.e., l'utilisateur). Dans le contexte de segmentation d'image, des expérimentations utilisant des images de télédétection montrent l'utilité de notre approche pour l'évaluation de la qualité des (M)BPTs.

In order to evaluate the interest of the proposed framework for building multi-feature BPTs, it is required to compare the MBPT with the standard BPT. This need opens new questions on how to evaluate the quality of a hierarchical image representation (e.g., BPT, MBPT) to provide further a relevant segmentation result. In this part, we first introduce the impact of the quality of a BPT in the context of image segmentation, the related works and our motivations for this work. Then, we propose a supervised framework assessing the ability of hierarchical structures in the context of segmentation. We develop two sides within this framework. We propose both an intrinsic and an extrinsic analysis of the hierarchical structure of a BPT with respect to some ground-truth examples defined by the domain expert (i.e., the user). In the context of image segmentation, experiments

using remote sensing images emphasize the usefulness of our framework for evaluating the quality of (M-)BPTs.

Chapter 5

Problem statement

Contents

In this chapter, we first discuss about the impact of the quality of a BPT for further image segmentation (Section 5.1). Then, we present the closest works related to the evaluation of hierarchical image representations such as BPTs (Section 5.2). Finally, we introduce our second contribution in this thesis leading to a new framework for BPT quality evaluation (Section 5.3). Our methods are able to assess both BPTs and MBPTs and can be extended to other hierarchical image representations.

5.1 Impact of the quality of a BPT for image segmentation

A BPT constitutes a relevant data-structure for image segmentation. In particular, any of its (partial) cuts providing a (partial) partition, it constitutes a search space containing potential segmentation results, associated to a given image. At first sight, BPTs appear as very good image representations for image segmentation, allowing us to obtain a cut / partition / segmentation delineating objects of interest in a fast and efficient way. Actually, BPTs are indeed very good data-structures for that purpose, *provided they are correctly built*.

BPTs can be considered as mixed hierarchies since they rely not only on the information directly embedded in the image (signal, topological structure), but also on extrinsic information, namely a priori knowledge related to the structures of interest (present within the image content), modelled via a metric (see Section 1.3). Then, both image and metric are involved in the construction of the BPT data-structure.

When dealing with an intrinsic image model (e.g., component-trees), the quality of the segmentation result only depends on the efficiency of the segmentation algorithm (i.e., the strategy for defining a relevant cut within the tree). By contrast, when dealing with a BPT (and, more generally, a mixed image model), the quality of the segmentation result depends not only on the efficiency of the segmentation algorithm, but also on the quality of the hierarchy of partitions built beforehand. This notion of BPT quality is directly related to the way the BPT is built, given a particular metric impacting the node merging order and a particular application. In other words, when considering BPT-based image processing, *a good segmentation algorithm has to be applied on a good image tree data-structure.*

Indeed, BPTs are low-size data-structures, that give access to possible partitions of Ω. However, the space of partitions of a set Ω is huge, and grows at a super-exponential rate with respect to its size. In other words, a BPT provides a (very) small subset of possible partitions of Ω. As a direct consequence, the "quality" of a BPT has a direct impact on the efficiency of any subsequent algorithm that will use the hierarchical structure for image segmentation.

Being able to evaluate the BPT quality is then of paramount importance. Providing such an evaluation framework is our main purpose in the second part of this manuscript.

5.2 Related works to BPT quality evaluation

As mentioned in Section 1.3, a large literature has been devoted to BPT-based segmentation, in particular in the challenging context of remote sensing. Various metrics were investigated (spectral, spatial, geometric, etc.) for building BPTs. The design of these metrics (choice and combination) strongly influences the resulting BPTs (i.e., the reduced search space for

further segmentation), and thus the quality of the subsequent segmentation result.

But, surprisingly, there exist very few works devoted to evaluate the quality of BPTs. In other words, there does not exist any framework allowing us to determine if a BPT is defined in a relevant way —i.e. with the adapted construction metric(s)— with respect to a given segmentation application on a specific image. To the best of our knowledge, the only framework for assessing the quality of a hierarchy of partitions for segmentation purpose was proposed in [61]. It consists of selecting in the tree a set of segments matching an ideal partition; such partition is forced to be in the hierarchy and its selection is considered as a linear fractional combinatorial optimization problem.

More generally, in the case of supervised segmentation evaluation, the results are compared to a ground-truth, composed of reference segments; it is then possible to rely on standard quality indices, e.g. Jaccard index [62], Dice coefficient (a.k.a F-measure) [63, 64], based on spatial overlapping information. From such indices, many frameworks for segmentation quality measures were developed [65, 66, 67, 68, 69, 70], based from example on region or contour-based strategies.

However, in the case of BPTs —and, more generally, hierarchies of partitions— these frameworks can not be directly considered, since several segmentation results can be obtained from a single BPT.

5.3 Towards a supervised BPT quality evaluation framework

In this part, we propose a framework for BPT quality evaluation, in the context of image segmentation. *It is fundamental to notice that this framework does not deal with segmentation evaluation, but with the evaluation of a hierarchical data-structure, namely the search space of potential segmentation results. More precisely, our goal is to evaluate the ability of a BPT to build nodes matching with potential objects of interest.* Such data-structure participates, of course, to the quality of a subsequent segmentation. However, we aim at helping the user to choose the right BPT, but not to use it the right way. Indeed, this latter issue is equivalent to

segmentation evaluation, and was already deeply investigated in the literature.

This framework is necessarily supervised. Indeed, the notion of segmentation quality depends not only on the chosen application, but also on the user expectations. In particular it has to rely on his / her expertise in terms of expected results. This is done by providing the framework with user-defined ground-truth examples. By side effect, this framework also offers the possibility to consider relevant elements in the context of expert / human-based evaluation, namely semantic enrichment and uncertainty handling.

We develop two sides within this framework. First, we propose an intrinsic quality analysis of a BPT. This analysis relies on the structural coherence of the BPT with respect to ground-truth examples. More precisely, we evaluate in which extent the hierarchical structure of a BPT is well-matching such examples, in a set / combinatorial fashion. This contribution is described in Chapter 6.

Second, we propose an extrinsic quality analysis of a BPT. This is done by allowing the user to assess the quality of a BPT based on chosen quality metrics that correspond to the desired properties of a subsequent segmentation. More precisely, we evaluate the ability of a BPT to provide nodes optimizing such quality metrics according to a set of ground-truth segments examples. This contribution is described in Chapter 7.

In Chapter 8, we present some experiments emphasizing the interest of our evaluation methods, allowing us to assess quantitatively our multi-feature framework by comparing BPTs and MBPTs.

5.4 Résumé : Besoins et position du travail

Dans ce chapitre, nous avons discuté de l'impact de la qualité d'un BPT sur la segmentation d'image. Ensuite, nous avons présenté les travaux les plus proches concernant l'évaluation des représentations d'images hiérarchiques comme les BPTs. Finalement, nous avons introduit notre seconde contribution qui correspond à de nouvelles approches pour l'évaluation de la qualité des BPTs. Nos méthodes sont à la fois capables d'évaluer les BPTs et les MBPTs et peuvent être étendues pour l'évaluation d'autres

représentations d'images hiérarchiques.

Impact de la qualité d'un BPT pour la segmentation d'image Lors de l'utilisation d'un modèle intrinsèque d'image (e.g., arbres de composantes), la qualité du résultat de segmentation dépend seulement de la méthode de segmentation utilisée pour extraire les objets d'intérêt à partir du modèle (e.g., coupe). Contrairement à un tel modèle intrinsèque d'image, le BPT peut être considéré comme une structure hiérarchique mixte puisque sa construction se base à la fois sur le contenu de l'image I mais aussi de la métrique W extrinsèque issue de connaissances a priori de l'utilisateur. Pour le cas des BPTs (et, généralement, des modèles mixtes), la qualité du résultat de segmentation ne dépend pas seulement de l'efficacité de l'algorithme de segmentation, mais aussi de la qualité de la hiérarchie de partitions construite préalablement. La notion de qualité de BPT est directement reliée à sa construction suivant une métrique précise pour une application donnée. En d'autres termes, une bonne méthode de segmentation devrait être appliquée sur un bon BPT. En effet, le BPT est une structure de données de taille raisonnable proposant quelques partitions possibles de Ω, comparativement à un espace de recherche de taille plus élevée. En conséquence, la qualité du BPT a un impact direct sur l'efficacité de l'algorithme de segmentation subséquent qui utilise sa structure hiérarchique. Pouvoir évaluer la qualité d'un BPT est donc d'une importance capitale. Proposer une approche d'évaluation d'une telle structure hiérarchique est notre deuxième principal objectif dans cette thèse.

Travaux actuels relatifs à l'évaluation de la qualité d'un BPT De nombreux travaux ont été dédiés aux méthodes de segmentation utilisant des BPTs, en particulier dans le domaine de la télédétection. Plusieurs métriques de construction ont été proposées et leurs combinaisons linéaires ont été étudiées dans le but d'améliorer les qualités de segmentations que l'on puisse obtenir à partir des BPTs. Cependant, peu de travaux s'intéressent à l'évaluation des BPTs. En d'autres termes, il n'existe pas d'approche nous permettant de déterminer si un BPT a été construit d'une façon pertinente pour une application spécifique en segmentation d'image. À notre connaissance, seule une approche a été proposée pour évaluer la qualité

d'une partition hiérarchique pour la segmentation d'image. Elle consiste en la sélection d'un ensemble de segments, dans l'arbre, qui correspondent au mieux à une partition idéale dont la présence dans la hiérarchie est requise. Généralement, pour le cas de l'évaluation supervisée d'une méthode de segmentation, les résultats sont comparés à une vérité-terrain composée de segments de référence. Il est alors possible d'utiliser des indices de qualité standards tels que les indices de Jaccard ou Dice. Ces indices se basent sur les informations de chevauchement des régions. Cependant, dans le cas des BPTs —et, plus généralement, des partitions hiérarchiques— ces approches ne peuvent pas être considérées directement car plusieurs résultats de segmentations possibles peuvent être obtenus à partir d'un seul BPT.

De nouvelles approches pour l'évaluation de la qualité d'un BPT Dans cette partie, nous proposons une approche pour l'évaluation de la qualité des BPTs, dans un contexte de segmentation d'images. Il est important de remarquer que cette approche ne traite pas l'évaluation d'une méthode de segmentation, mais se concentre sur l'évaluation d'une structure de données hiérarchique utilisée comme espace de recherche des résultats potentiels de segmentation. De telles structures participent, effectivement, à la qualité de la segmentation subséquente. Notre but est d'aider l'utilisateur à choisir le bon BPT, mais pas la bonne manière d'utiliser un BPT. La notion de qualité de segmentation ne dépend pas seulement de l'application, mais aussi des attentes de l'utilisateur. Ainsi, notre approche d'évaluation de la qualité des BPTs est alors supervisée. Afin de matérialiser ses attentes, l'utilisateur doit fournir des exemples de vérité-terrain. Par effet de bord, cette approche offre également la possibilité de considérer des éléments pertinents comme des informations sémantiques et la prise en compte de l'incertitude. Nous développons deux parties dans le cadre de cette approche. Premièrement, nous proposons une analyse intrinsèque permettant une évaluation d'un BPT à partir de la cohérence de sa structure hiérarchique par rapport à un segment de référence de vérité-terrain. Deuxièmement, nous proposons une analyse extrinsèque permettant à l'utilisateur de choisir une métrique d'évaluation basée sur les propriétés désirées de la segmentation subséquente.

Chapter 6

Intrinsic quality analysis of a BPT

Contents

Our purpose is to evaluate the quality of a given BPT. More precisely, we aim to evaluate the quality of the construction process of a BPT. Indeed, as stated in Section 1.3, the building process —and the resulting BPT— mainly depends on the way to iteratively *choose* the successive couples of adjacent nodes to be merged, induced by a chosen construction metric. As a consequence, assessing the quality of a BPT for a given image is actually a way of assessing the relevance of this choice strategy, that generally depends on specific features for node description and specific metrics for node comparison in the induced feature space.

Let us consider a given image $I : \Omega \to \mathbb{V}$, and let us suppose that a BPT $\mathfrak{T} = (\mathcal{N}, \searrow)$ has been built from I. The quality assessment of $\mathfrak{T}$ is also a supervised process, that requires to provide user-defined ground-truth examples. Such examples can be simply defined as subsets of Ω. For the sake of readability, we will assume that such examples are connected sets of pixels (with respect to the adjacency relation $\frown$ on Ω). We will also discuss, in the sequel, the case where one (non-empty) ground-truth example $G \subseteq \Omega$ is provided; the case of $k > 1$ examples can then be trivially derived.

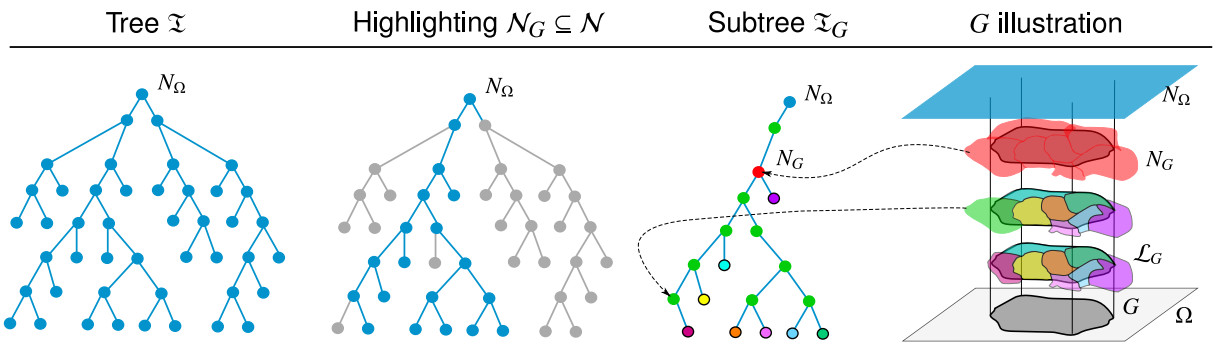


Figure 6.1: Illustration of the subtree extraction where $\mathfrak{T}$ is the whole tree, N_Ω is its root, $\mathcal{L}_G$ is the set of leaves intersecting the ground-truth example G. The tree $\mathfrak{T}_G$ is the subtree of $\mathfrak{T}$ that is for interest here and N_G is its preferable root.

In this chapter, our contribution is to show that a first quality analysis can be carried out by directly observing the BPT $\mathfrak{T}$ with respect to the ground-truth example G. In other words, the BPT provides —by its inner hierarchical structure and its spatial embedding in Ω— some information about its relevance and its ability to extract particular objects of interest. We will discuss, in Chapter 7 about some complementary analysis that involve also additional quality metrics[1] provided by the user.

6.1 Example-based subtree definition

Our purpose, in this first quality analysis, is to evaluate how well a BPT $\mathfrak{T}$ is adapted to provide nodes matching with a ground-truth example G. The nodes $\mathcal{N}$ of $\mathfrak{T}$ can be subdivided into two families: nodes that intersect G, and nodes that do not. Practically, the second family is indeed useless for this analysis. Then, we focus only on the nodes $N \in \mathcal{N}$ such that $N \cap G \neq \emptyset$.

The subset of such nodes can be easily computed, in a bottom-up fashion, by first selecting all the leaves $L \in \mathcal{L}$ that intersect G, and then preserving iteratively all the parent nodes connected to any such leaf by the $\searrow$ relation, until the root N_Ω, that necessarily satisfies this property (see Figure 6.1).

The obtained set of nodes $\mathcal{N}_G \subseteq \mathcal{N}$ induces a subtree $\mathfrak{T}_G$ of $\mathfrak{T}$, of root

[1] In the next chapters of this thesis, two different types of metrics will be encountered, the "construction metric" that is used for the BPT construction, and the "quality metric" that is measuring the similarity / quality between a region and a reference segment.

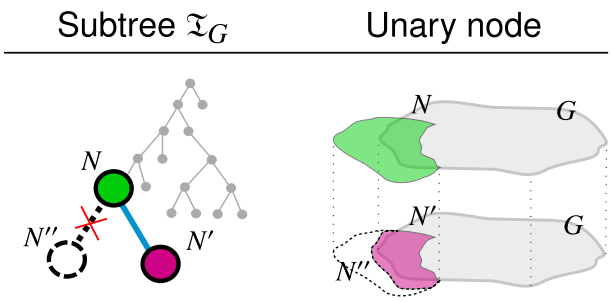


Figure 6.2: Example of a unary node N where $N \searrow N'$ and $N \cap G = N' \cap G$, while $N' \subset N$. The only other node $N'' \subset N$ such that $N \searrow N''$ does not intersect G.

N_Ω and of leaves $\mathcal{L}_G = \{L \in \mathcal{L} \mid L \cap G \neq \emptyset\}$. By construction, this subtree $\mathfrak{T}_G$ *may not be binary*, in general. This means that there may exist nodes $N \in \mathcal{N}_G$ that have exactly one child $N' \in \mathcal{N}_G$ (i.e., $N \searrow N'$). This happens when the only other node $N'' \subset N$ such that $N \searrow N''$ does not intersect G. Such nodes N are called *unary* nodes (see Figure 6.2), by contrast with the other *binary* nodes that have two children nodes in $\mathcal{N}_G$.

An important property of a unary node N is that its intersection with G is exactly the same as for its only child N'. In other words, for a unary node N, if $N \searrow N'$, then we have $N \cap G = N' \cap G$, while $N' \subset N$. This means that a unary node "increases" the amount of false positive material with respect to G, compared to its descendents.

This is highlighted in the upper part of the tree $\mathfrak{T}_G$, in particular between the root N_Ω and the first binary node N_G of $\mathfrak{T}_G$. Indeed, except in the case where N_Ω is itself a binary node, there exists, within $\mathcal{N}_G$, a sequence of successive nodes $N_\Omega = N_0 \searrow N_1 \searrow \ldots \searrow N_k = N_G$ $(k \geq 1)$ such that all N_i $(0 \leq i \leq k-1)$ are unary. In such case, we can relevantly remove from $\mathfrak{T}_G$ all these nodes N_i, to only preserve, as root, the first binary node N_G (see the red node N_G of Figure 6.1). By construction, N_G is the smallest node of $\mathcal{N}$ that includes G.

The final subtree (still noted $\mathfrak{T}_G$, by abuse of notation) contains all the nodes of interest[2] of $\mathfrak{T}$, with respect to the ground-truth example G. Its combinatorial analysis then allows us to obtain preliminary information about the quality of the BPT $\mathfrak{T}$ according to the provided ground-truth

[2]In order to reduce the space cost of $\mathfrak{T}_G$ in a lossless way, it is also possible to remove the children of the (binary) nodes N that are included in G. Indeed, such nodes have all their children binary and also included in G. The information to be stored in N is the number l of leaves of the subtree of root N that also gives access to the number of removed nodes (i.e., $2.l - 2$).

example.

6.2 Relevance of the intrinsic quality analysis

The analysis of the subtree $\mathfrak{T}_G$ provides information about the quality, with respect to the ground-truth segment G, of the way of choosing successively the couple of nodes to be merged.

However, the initial partition of Ω, that defines the set of leaves $\mathcal{L}$, and thus, the subset $\mathcal{L}_G$ of leaves of the subtree, is an input of the BPT construction process. This partition is non-correlated to the node-merging strategy. Nevertheless, this partition has an important impact on the relevance of the proposed approach for assessing the quality of the BPT. In particular, two properties of this partition are crucial in this context.

Granularity First, the *granularity* of $\mathcal{L}_G$ is of importance. The granularity index γ is defined as the ratio between the size of G (number of points) and the size of $\mathcal{L}_G$ (number of leaves). It is defined as

$$\gamma = \frac{|\mathcal{L}_G|}{|G|} \tag{6.1}$$

and lies in $]0, 1]$. *The higher the granularity index, the most relevant the intrinsic quality analysis carried out on* $\mathfrak{T}_G$. To illustrate this fact, we consider the two extremal cases. Let us suppose that $\gamma = 1$. This means that the number of leaves that intersect G is equal to the number of points of G. In this context, each leaf contains exactly one point of G, and the ability of the BPT to allow for the definition of a cut that correctly fit G highly depends on the way to progressively merge the nodes. Now, let us suppose that $\gamma = 1/|G|$. This means that one leaf already includes G (i.e., $\mathfrak{T}_G$ is a trivial tree composed of one node). In this context, the initial partition already determined the adequacy between the BPT and G, and the way of further building $\mathfrak{T}_G$ has no actual influence.

Discordance Second, the *discordance* of $\mathcal{L}_G$ is of importance. The discordance δ is defined as the relative quantitative error on the size of G induced by $\mathcal{L}_G$, and more precisely by the nodes that partially intersect G. It is

defined as

$$\delta = \frac{1}{|G|} \sum_{L \in \mathcal{L}_G, L \not\subseteq G} \min\{|L \setminus G|, |L \cap G|\} \tag{6.2}$$

and lies in $[0, 1[$. *The lowest the discordance, the most relevant the intrinsic quality analysis carried out on* $\mathfrak{T}_G$. To illustrate this fact, we consider the two extremal cases. Let us suppose that $\delta = 0$. This means that the initial partition provides a set of leaves that perfectly fits G. In this context, the quality of the BPT directly depends on the ability to progressively merge these nodes in order to retrieve G. Now, let us assume that $\delta \approx 1$. This means that many leaves of $\mathcal{L}$ partially intersect both G and the remainder of Ω. In this context, the ability to finally obtain a "good" segmentation of G is quite low and weakly depends on the ability to merge the nodes when building $\mathfrak{T}$.

It is important to note that both granularity and discordance have to provide good values, if we want to relevantly assess the structural quality of a BPT. For G sufficiently large, a high value of granularity will imply a low value of discordance. However, the counterpart is not true: a low value of discordance can be obtained for a low value of granularity (e.g., if a leaf L already fits G). As a consequence, it is important to carefully compute these two scores before any such study.

Of course, such precaution becomes useless if the initial partition $\mathcal{L}$ is isomorphic with Ω (i.e., when the initial partition is composed of leaves that are points of the image). Indeed, in such case, we necessarily have $\gamma = 1$ and $\delta = 0$; in other words, the quality of the final BPT only depends on the node merging process, since the latter is not biased by the initial grouping of points into primitive regions.

6.3 Intrinsic quality analysis

From now on, we assume that both the γ and δ scores are sufficiently good to carry out an intrinsic quality analysis. In other words, we assume that the partition of the leaves of the BPT $\mathfrak{T}$ is sufficiently fine, and globally well-fitting the ground-truth segment G. Otherwise, it is plain that the user has to reconsider his / her way to define the initial partition, in order to adapt it to the targeted structures of interest.

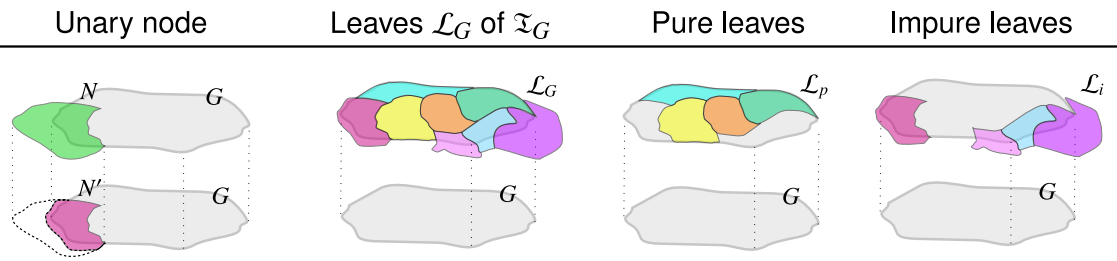


Figure 6.3: Example showing a unary node (impure) and sets of pure and impure leaves.

Combinatorial analysis

The subtree $\mathfrak{T}_G$ is composed of n nodes (including 1 root and l leaves) and $n-1$ edges (i.e., $\searrow$ relation between parent–child nodes). Observing the status of these nodes and carrying out a combinatorial analysis on the different populations provide us with quality clues of the BPT.

Let us first introduce the following terminology:

- a node $N \in \mathcal{N}_G$ is *pure* if $N \subseteq G$;
- a node $N \in \mathcal{N}_G$ is *impure* if $N \not\subseteq G$.

The classification of the leaves of $\mathcal{L}_G$ into pure / impure can be done easily by observing their support, compared to G (see Figure 6.3). We set l_p and l_i the number of pure and impure leaves, respectively.

In particular, we have $l_p + l_i = l$, with l —number of leaves— close to $|G|$, due to high granularity hypothesis.

The purity / impurity of the other nodes can be computed quite easily, by considering that:

- if $N \in \mathcal{N}_G$ is a unary node, then it is impure;
- if $N \in \mathcal{N}_G$ is a binary node, with $N \searrow N', N''$, then
  - if N' or N'' is impure, then N is impure;
  - if N' and N'' are pure, then N is pure.

A "good" BPT construction should preserve as much as possible the purity of nodes, and avoid (1) to merge pure and impure nodes, and (2) to increase the size of impure nodes. Based on these assertions:

- the merging leading to a unary node in $\mathfrak{T}_G$ is a bad operation, as it creates from impure —and sometimes pure— node, an impure node with a greater amount of points out of G;
- the merging leading to a binary node from two pure nodes in $\mathfrak{T}_G$ is a good operation, as it allows one to converge towards G;
- the merging leading to a binary node from two impure nodes in $\mathfrak{T}_G$ is a good operation, as it avoids to deteriorate pure areas and to increase the amount of points out of G;
- the merging leading to a binary node from a pure and an impure node in $\mathfrak{T}_G$ is a bad operation, as it makes the result diverging from G.

By simply observing the subtree $\mathfrak{T}_G$, it is possible to count the number of each kind of nodes: u_i, b_{pp}, b_{ii} and b_{pi}, for the unary nodes, and binary nodes built from pure-pure, impure-impure, pure-impure couples, respectively (see Table 6.1).

Table 6.1: Definitions of u_i, b_{pp}, b_{ii} and b_{pi}.

u_i	number of unary nodes
b_{pp}	number of binary nodes built from pure-pure couples
b_{ii}	number of binary nodes built from impure-impure couples
b_{pi}	number of binary nodes built from pure-impure couples

We have $u_i + b_{pp} + b_{ii} + b_{pi} + l_p + l_i = n$, and a good BPT should minimize u_i and b_{pi}, while maximizing b_{pp} and b_{ii}. In particular, a "perfect" BPT, with respect to these criteria, should satisfy:

$$u_i = 0 \tag{6.3}$$
$$b_{pp} = l_p - 1 \tag{6.4}$$
$$b_{ii} = l_i - 1 \tag{6.5}$$
$$b_{pi} = 1 \tag{6.6}$$

(except when $l_i = 0$, where b_{pp} should be equal to $l - 1$, and all others to 0).

From this classification of nodes, and the combinatorial analysis of their population in the best case, it is possible to build a wide range of structural

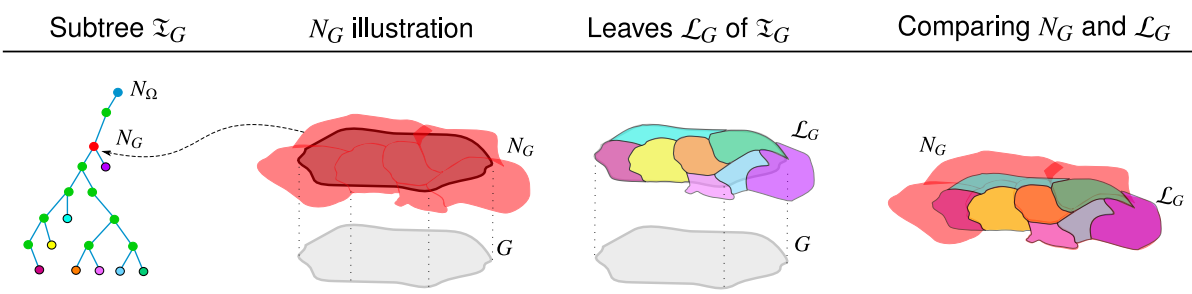


Figure 6.4: Comparison between the amount of points contained in the root $N_{\mathfrak{G}}$ of $\mathfrak{T}_G$ and the amount of points that could be theoretically obtained from the initial partition of the leaves $\mathcal{L}_G$.

measures that quantify the difference of quality between BPTs, such as described in Eqs. (6.3–6.6). For instance, $(b_{pp} + b_{ii})/(u_i + b_{pi})$ and b_{pp}/b_{pi} are illustrative examples of such measures.

Quantitative analysis

A more quantitative assessment of the quality of the BPT can also be carried out by observing the lowest set including G and the greatest set included in G which can be built from $\mathfrak{T}_G$.

Let us first focus on the lowest set including G. By construction, this set is indeed the root of $\mathfrak{T}_G$, namely N_G. The interesting information carried by N_G is the amount of points that are outside of G (see Figure 6.4).

More precisely, this amount $|N_G \setminus G| = |N_G| - |G|$ has to be compared to the amount that could be theoretically obtained from the initial partition of the leaves $\mathcal{L}_G$, namely $|(\cup_{L \in \mathcal{L}_G} L) \setminus G| = (\Sigma_{L \in \mathcal{L}_G} |L|) - |G|$. Computing the difference, or the ratio, between these two values allows us to assess the quantitative error related to the existence of unary nodes in $\mathfrak{T}_G$ (i.e., the addition of non-relevant zones to the expected exhaustive segmentation of G). The lower this value, the better the ability of the BPT to use at best the potential adequacy of the initial partition to the ground-truth segment.

Second, let us focus on the greatest set included in G (see Figure 6.5). By definition, this set is defined as the union of all the pure nodes of $\mathcal{N}_G$ which are maximal with respect to the $\searrow$ relation. In other words, we consider all the pure nodes whose parent nodes are impure; we note this set $\mathcal{N}_p$. Here, two values are important. Similarly to the above study on N_G, the first value of interest is the amount of points of G which are

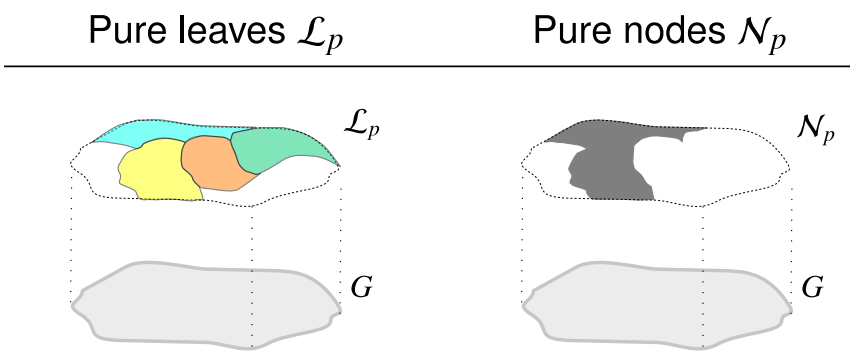


Figure 6.5: Illustration of the set $\mathcal{L}_p$ of *pure* leaves and the set $\mathcal{N}_p$ of *pure* nodes that are maximal with respect to the $\searrow$ relation. The dashed white regions represent $|G| - (\sum_{N \in \mathcal{L}_p} |N|)$, on the left, and $|G| - (\sum_{N \in \mathcal{N}_p} |N|)$, on the right.

outside this pure set. This amount is defined as $|G| - (\sum_{N \in \mathcal{N}_p} |N|)$. It can be compared to the amount of G that is outside of the best theoretical pure subset that could be obtained from the leaves (i.e., $|G| - (\sum_{N \in \mathcal{L}_p} |N|)$, where $\mathcal{L}_p = \{L \in \mathcal{L} \mid L \subseteq G\}$). The lower this value, the better the ability of the BPT to use at best the potential adequacy of the initial partition to the ground-truth segment. A second value is also of interest; it is the size of $\mathcal{N}_p$ (i.e., the number of nodes which are required to form the best segmentation lower than G). The lower this value, the better the ability of the BPT to avoid over-segmentation of the structure of interest (see Figure 6.5).

Of course, these two values have to be considered in association. Indeed, a low value of $|\mathcal{N}_p|$ is not meaningful if the union of the nodes of $\mathcal{N}_p$ is highly degraded with respect to the initial partition. Similarly, an optimal union of the nodes of $\mathcal{N}_p$ is meaningless if these nodes are numerous (i.e., could not be merged together by the BPT, and were instead fused into impure greater nodes).

In the next chapter (Chapter 7), we present an extrinsic quality analysis of a BPT.

6.4 Résumé : Analyse de la qualité intrinsèque d'un BPT

Notre but est d'évaluer la qualité d'un BPT donné. Plus précisément, nous cherchons à évaluer la qualité du processus de construction d'un BPT. En

effet, le processus de construction, ainsi que le BPT résultant, dépendent des choix successifs des couples de nœuds adjacents à fusionner itérativement, induits par une métrique de construction choisie. En conséquence, évaluer la qualité d'un BPT d'une image donnée est actuellement un moyen pour évaluer la pertinence du choix de stratégie, qui dépend généralement d'attributs spécifiques pour la description des nœuds et de métriques pour leur comparaison.

Considérons une image donnée $I : \Omega \to \mathbb{V}$, et supposons qu'un BPT $\mathfrak{T} = (\mathcal{N}, \searrow)$ a été créé à partir de I. L'évaluation de la qualité de $\mathfrak{T}$ est un processus supervisé, qui nécessite des exemples de segments de vérité-terrain fournis par l'utilisateur. De tels exemples peuvent être définis comme sous-ensembles de Ω. Pour une bonne lecture, nous supposons que de tels exemples sont des ensembles de pixels connectés (en respectant la relation d'adjacence $\frown$ sur Ω). Nous discutons également, par la suite, du cas où un exemple de segment de vérité-terrain $G \subseteq \Omega$ est fourni; le cas de $k > 1$ exemples peut ensuite en dériver.

Dans ce chapitre, notre contribution consiste à montrer qu'une première analyse de qualité peut être effectuée en observant directement le BPT $\mathfrak{T}$ vis-à-vis du segment de vérité-terrain G. En d'autres termes, le BPT fournit —à travers sa structure hiérarchique interne et son intégration spatiale dans Ω— des informations concernant sa pertinence et sa capacité pour l'extraction d'un objet d'intérêt particulier. Nous discutons de l'analyse extrinsèque impliquant des métriques de qualité additionnelles choisies par l'utilisateur.

Extraction de sous-arbres « basés-exemple » Dans cette première analyse de la qualité, notre but est d'évaluer si le BPT $\mathfrak{T}$ est adapté ou non à fournir des nœuds correspondant au segment de référence G. Les nœuds de $\mathfrak{T}$ peuvent être divisés en deux familles : ceux qui intersectent G, et ceux qui ne l'intersectent pas. En pratique, cette deuxième famille est inutile pour l'analyse. Ainsi, nous nous concentrons seulement sur les nœuds $N \in \mathcal{N}$ tels que $N \cap G \neq \emptyset$. Le sous-ensemble de tels nœuds peut être facilement calculé, d'une façon ascendante, en sélectionnant premièrement toutes les feuilles $L \in \mathcal{L}$ qui intersectent G, et en ne conservant par la suite que les nœuds parents connectés à ces feuilles par la relation $\searrow$, jusqu'à la racine

N_Ω. L'ensemble de nœuds $\mathcal{N}_G \subseteq \mathcal{N}$ obtenu induit un sous-arbre $\mathfrak{T}_G$ de $\mathfrak{T}$, de racine N_Ω et de feuilles $\mathcal{L}_G = \{L \in \mathcal{L} \mid L \cap G \neq \emptyset\}$. Par construction, ce sous-arbre $\mathfrak{T}_G$ peut ne pas être binaire, en général. Cela signifie qu'il y a des nœuds $N \in \mathcal{N}_G$ ayant seulement un seul fils $N' \in \mathcal{N}_G$ (i.e., $N \searrow N'$). Cela se produit quand l'autre fils $N'' \subset N$ tel que $N \searrow N''$ n'intersecte pas G. De tels nœuds sont appelés *unaires* contrairement aux nœuds *binaires* qui ont deux fils dans $\mathcal{N}_G$. Une propriété importante d'un nœud unaire est son intersection avec G qui est exactement la même que pour son unique fils N'. En d'autres termes, pour un noœud unaire, si $N \searrow N'$, alors nous avons $N \cap G = N' \cap G$, tandis que $N' \subset N$. Cela signifie que les nœuds unaires augmentent le nombre de faux positifs potentiels relativement à G. Cela est mis en évidence dans la partie haute de l'arbre $\mathfrak{T}_G$, en particulier entre la racine N_Ω et le premier nœud binaire N_G de $\mathfrak{T}_G$. En effet, outre le cas où N_Ω est un nœud binaire, il existe dans $\mathcal{N}_G$ une séquence de nœuds successifs $N_\Omega = N_0 \searrow N_1 \searrow \ldots \searrow N_k = N_G$ ($k \geq 1$) telle que tous les N_i ($0 \leq i \leq k-1$) sont unaires. Dans ce cas, nous pouvons faire abstraction de ces nœuds N_i de $\mathfrak{T}_G$, pour ne préserver, comme racine, que le premier nœud binaire N_G. Par construction, N_G est le plus petit nœud de $\mathcal{N}$ incluant totalement G. Le sous-arbre final (que nous appelons toujours $\mathfrak{T}_G$, par abus de langage) contient tous les nœuds d'intérêt de $\mathfrak{T}$ tout en respectant le segment de référence G. L'analyse combinatoire de ce sous-arbre $\mathfrak{T}_G$ nous permet d'obtenir des informations préliminaires concernant la capacité du BPT $\mathfrak{T}$ à fournir des nœuds correspondant au mieux au segment de référence G.

Pertinence de l'analyse de la qualité intrinsèque L'analyse du sous-arbre $\mathfrak{T}_G$ fournit des informations à propos de la qualité des décisions de fusions de nœuds du BPT $\mathfrak{T}$ suivant le segment de référence G. Cependant, la partition initiale de Ω, qui définit l'ensemble de feuilles $\mathcal{L}$, et ainsi le sous-ensemble $\mathcal{L}_G$ des feuilles du sous-arbre, est un paramètre du processus de construction de BPT. Cette partition est généralement non-corrélée à la stratégie de fusion des nœuds. Néanmoins, cette partition a un impact important sur la pertinence de l'approche proposée pour évaluer la qualité des BPTs. En particulier, deux propriétés de cette partition sont cruciales dans ce contexte. Premièrement, la granularité γ de $\mathcal{L}_G$ est importante, et

est définie par le rapport entre la taille de G et celle de $\mathcal{L}_G$ (i.e., $\gamma = \frac{|\mathcal{L}_G|}{|G|}$). Sa valeur est bornée dans $]0, 1]$. Plus la valeur de cet indice de granularité est élevée, plus l'analyse intrinsèque effectuée sur $\mathfrak{T}_G$ est pertinente. Pour illustrer ce fait, nous considérons les cas extrêmes. Pour $\gamma = 1$, i.e., le nombre de feuilles qui intersectent G est égal au nombre de points de G, chaque feuille contient exactement un point de G, et la capacité du BPT à autoriser une coupe correspondant correctement à G dépend de la manière dont les nœuds ont été progressivement fusionnés. Pour $\gamma = 1/|G|$, une feuille inclut déjà G (i.e., $\mathfrak{T}_G$ est un arbre trivial ne comportant qu'un seul nœud). Dans ce contexte, la partition initiale détermine déjà l'adéquation entre le BPT $\mathfrak{T}$ et G, et la manière de construire $\mathfrak{T}_G$ n'a pas d'influence. Deuxièmement, la discordance δ de $\mathcal{L}_G$ est aussi importante, et est définie par l'erreur quantitative relativement sur la taille de G induite par $\mathcal{L}_G$, et plus précisément par les nœuds qui intersectent partiellement G (i.e., $\delta = \frac{1}{|G|} \sum_{L \in \mathcal{L}_G, L \not\subseteq G} \min\{|L \setminus G|, |L \cap G|\}$). Sa valeur est bornée dans $[0, 1[$. Plus la discordance est faible, plus l'analyse de la qualité intrinsèque effectuée sur $\mathfrak{T}_G$ est pertinente. Pour illustrer ce fait, nous considérons deux cas extrêmes. Pour $\delta = 0$, la partition initiale fournit un ensemble de feuilles qui correspondent exactement à G. Dans ce contexte, la qualité du BPT dépend directement de sa capacité à fusionner progressivement des nœuds dans le but de retrouver G. Pour $\delta \approx 1$, beaucoup de feuilles de $\mathcal{L}$ intersectent partiellement à la fois G et le reste du support Ω. Dans ce contexte, la capacité à finalement obtenir une bonne segmentation de G est minime et dépend faiblement des fusions de nœuds lors de la construction de $\mathfrak{T}$. Il est important de remarquer que la granularité et la discordance doivent avoir des valeurs optimales, si nous souhaitons évaluer de façon pertinente la qualité de la structure d'un BPT. Une telle précaution devient inutile si la partition initiale $\mathcal{L}$ est isomorphe à Ω (i.e., quand la partition initiale est composée de feuilles qui sont des points de l'image). En effet, dans ce cas, nous avons $\gamma = 1$ et $\delta = 0$. En d'autres termes, la qualité finale du BPT dépend seulement du processus de fusion des nœuds.

Analyse de la qualité intrinsèque À partir de ce point, nous supposons que les valeurs de γ et de δ sont suffisamment bonnes pour permettre l'analyse intrinsèque de la qualité des BPTs. En d'autres termes, nous supposons

que la partition initiale $\mathcal{L}$ de feuilles du BPT $\mathfrak{T}$ est suffisamment correcte et correspond globalement au segment de vérité-terrain G. Autrement, il est évident que l'utilisateur doit reconsidérer la manière de définir la partition initiale, afin de l'adapter à la structure d'intérêt ciblée. Deux types d'analyses sont présentées. Premièrement, une analyse combinatoire permet d'avoir une première intuition sur la qualité de la structure interne du BPT $\mathfrak{T}$ suivant un segment de vérité terrain G. Le sous-arbre $\mathfrak{T}_G$ est composé de n nœuds (incluant 1 racine et l feuilles) et $n-1$ arêtes (i.e., des relations $\searrow$ entre nœuds père–fils). L'observation du statut de ces nœuds et une analyse combinatoire sur les différentes populations nous fournit des pistes sur la qualité du BPT. Tout d'abord, il est important de déterminer les nœuds *purs* (i.e., $N \subseteq G$) et les nœuds *impurs* (i.e., $N \not\subseteq G$). La classification des feuilles de $\mathcal{L}_G$ en pure / impure peut est faite facilement en observant leur support par rapport à G. Le nombre de feuilles pures et le nombre de feuilles impures sont notés l_p et l_i, respectivement. En particulier, nous avons $l_p + l_i = l$, avec le nombre de feuilles l proche de $|G|$, en respectant l'hypothèse de granularité. La pureté / impureté des autres nœuds peut être facilement déterminée en considérant qu'un nœud unaire $N \in \mathcal{N}_G$ est impur, un nœud obtenu par fusion de deux nœuds impaires est impaire et un nœud obtenu par fusion de deux nœuds paires est paire. Une bonne construction de BPT devrait préserver autant que possible la pureté des nœuds, et éviter de (1) fusionner des nœuds purs avec des nœuds impurs, et (2) augmenter la taille des nœuds impurs. Ainsi, la fusion ajoutant un nœud unaire dans $\mathfrak{T}_G$ est une mauvaise opération parce que cela crée des nœuds impurs avec un nombre élevé de points à l'extérieur de G. La fusion emmenant un nœud binaire à partir de deux nœuds purs dans $\mathfrak{T}_G$ est une bonne opération parce que cela permet de converger vers G. La fusion de deux nœuds impurs dans $\mathfrak{T}_G$ est une bonne opération parce que cela évite de détériorer les zones pures et d'augmenter la quantité de points à l'extérieur de G. La fusion d'un nœud pur et d'un nœud impur dans $\mathfrak{T}_G$ est une mauvaise opération parce que le résultat diverge de G. En observant le sous-arbre $\mathfrak{T}_G$, il est possible de compter le nombre u_i de nœuds de type unaire, le nombre b_{pp} de type pur-pur, le nombre b_{ii} de type impur-impur et le nombre b_{pi} de type pur-impur. Nous avons alors $u_i + b_{pp} + b_{ii} + b_{pi} + l_p + l_i = n$, et un bon BPT devrait minimiser u_i et

b_{pi} tout en maximisant b_{pp} et b_{ii}. En particulier, un BPT parfait devrait avoir $u_i = 0$, $b_{pp} = l_p - 1$, $b_{ii} = l_i - 1$ et $b_{pi} = 1$ sauf pour le cas où $l_i = 0$, où b_{pp} devrait être égal à $l - 1$, et le reste à 0). À partir de cette classification de nœuds, et l'analyse combinatoire de leur population dans les meilleurs cas, il est possible de concevoir des mesures structurelles qui quantifient la qualité d'un BPT et permettent ainsi de les comparer. Les valeurs $(b_{pp} + b_{ii})/(u_i + b_{pi})$ et de b_{pp}/b_{pi} sont des exemples illustratifs. Deuxièmement, une analyse quantitative évaluant la qualité d'un BPT peut aussi être effectuée en observant le plus petit ensemble incluant G et le plus grand ensemble inclus dans G que l'on puisse obtenir à partir de $\mathfrak{T}_G$. Par construction, le plus petit ensemble incluant G est effectivement la racine N_G de $\mathfrak{T}_G$. L'information intéressante apportée par N_G est la quantité de points qui sont à l'extérieur de G. Plus précisément, cette quantité $|N_G \setminus G| = |N_G| - |G|$ devrait être comparée à la quantité de points $|(\cup_{L \in \mathcal{L}_G} L) \setminus G| = (\sum_{L \in \mathcal{L}_G} |L|) - |G|$ qui devrait être théoriquement obtenue à partir de la partition initiale des feuilles $\mathcal{L}_G$. Le calcul de différence, ou le ratio, entre ces deux valeurs permet d'évaluer les erreurs quantitatives liées à l'existence des nœuds unaires dans $\mathfrak{T}_G$. Plus cette valeur est petite, meilleure est la capacité du BPT à utiliser l'adéquation potentielle de la partition initiale au segment de vérité-terrain. Par définition, le plus grand ensemble inclus dans G est l'union de tous les nœuds purs de $\mathcal{N}_G$ qui sont maximaux en respectant la relation $\searrow$. En d'autres termes, nous considérons tous les nœuds purs $\mathcal{N}_p$ dont les parents sont impurs. Une valeur importante dans l'analyse est la quantité de points de G qui sont à l'extérieur de cet ensemble pur qui est défini par $|G| - (\sum_{N \in \mathcal{N}_p} |N|)$. Elle peut être comparée à la quantité de points de G se trouvant à l'extérieur du meilleur sous-ensemble pur $|G| - (\sum_{N \in \mathcal{L}_p} |N|)$, où $\mathcal{L}_p = \{L \in \mathcal{L} \mid L \subseteq G\}$ qui pourrait être théoriquement obtenu à partir des feuilles. Plus cette valeur est petite, mieux est la capacité du BPT à utiliser l'adéquation potentielle de la partition initiale au segment de vérité-terrain. Une autre valeur intéressante est le nombre de nœuds requis pour une meilleure segmentation plus petit que G (i.e., taille de $\mathcal{N}_p$). Plus cette valeur est petite, mieux lest la capacité du BPT à éviter le cas de la sur-segmentation de la structure d'intérêt. Évidemment, les deux valeurs que nous avons énoncées doivent être considérées ensembles. En effet, une valeur optimale de $|\mathcal{N}_p|$ n'apporte

rien si l'union des nœuds de $\mathcal{N}_p$ est hautement dégradée par rapport à la partition initiale. De façon similaire, l'union optimale des nœuds de $\mathcal{N}_p$ n'a pas beaucoup de sens si la quantité de ces nœuds est élevée.

Chapter 7

Extrinsic quality analysis of a BPT

Contents

In this second chapter, we propose an approach to evaluate the quality of a given BPT by not only relying on its intrinsic structure —as proposed in Chapter 6— but also by taking advantage on some extrinsic information. We propose a new supervised framework for BPT quality evaluation. Our contribution is twofold. On the one hand, our purpose is not to match an ideal partition of the image, but to evaluate the ability of a BPT to construct nodes that match at best with a subset of expert-defined segments; in other words, our ground-truth is not required to be global but partial. On the other hand, since the ground-truth examples are defined by human experts, we assume that they can be imperfect, and we integrate the induced uncertainty in the evaluation process. In addition, these ground-truth segments can be labelled by the expert (see Figure 7.1), then allowing us to embed semantic criteria in the evaluation framework, and to improve the robustness of the global quality score. In Section 7.1, we present first the workflow of our contribution that reminds also our motivations. Then, the three next sections (i.e., Section 7.2, Section 7.3 and Section 7.4) detail each step mentioned in this workflow.

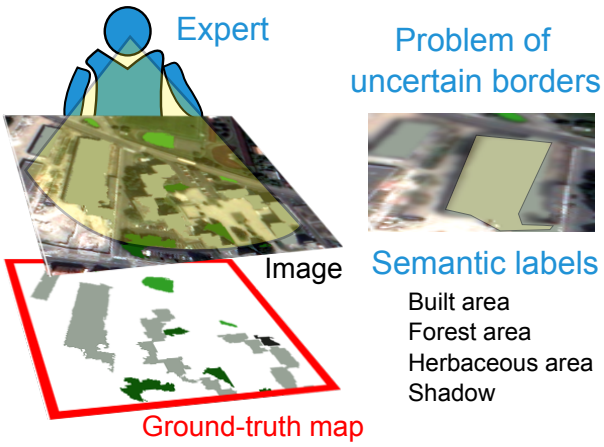


Figure 7.1: The ground-truth map is chosen by the expert and composed by k segments G_i (i.e., ground-truth examples). These segments can be labelled semantically by the expert and may present some border uncertainty.

7.1 Workflow

We define hereinafter the workflow of the proposed extrinsic quality analysis of a BPT. Once the ground-truth map is defined by the expert, our framework for supervised assessment of BPT quality relies on three main components: (i) a local metric that quantitatively evaluates the matching degree between a ground-truth example G and a node N of the BPT $\mathfrak{T}$ (Section 7.2); (ii) the determination of nodes within the BPT that locally maximise matching with the different ground-truth examples (Section 7.3); and (iii) a global quality measure that merges the local scores induced by this optimal cut for each of the ground-truth examples (Section 7.4). The handling of uncertainty of the ground-truth is mainly considered in the local metric of Step (i), while the semantic labelling of the ground-truth examples is mainly used in the definition of the global quality measure of Step (iii). The global process of our extrinsic supervised BPT evaluation framework is summarized in Figure 7.2.

7.2 Node / segment matching metrics and uncertainty

Comparing a node N and a ground-truth example that is a segment G consists of quantifying the degree of similarity between them, via the setting

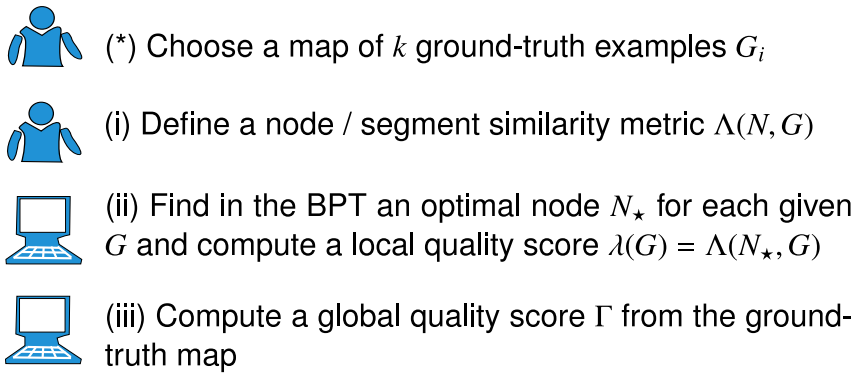


Figure 7.2: Global process of our extrinsic supervised BPT evaluation framework.

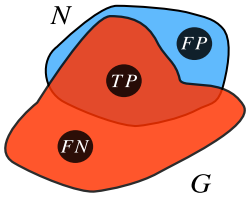


Figure 7.3: Region-based overlapping information where TP: true positives, FP: false positives and FN: false negatives.

of a local score $\Lambda(N, G) \in \mathbb{R}^+$, where Λ is a metric[1] or pseudo-metric that evolves monotonically with respect to this similarity.

Basically, such similarity assessment can be made in a region-based fashion, by considering spatial overlapping information (i.e., by computing the true positives (TP), false positives (FP) and false negatives (FN) between N and G, given by $TP = |N \cap G|$, $FP = |N \backslash G|$, and $FN = |G \backslash N|$, respectively). The Figure 7.3 illustrates an example for such region-based overlapping information.

The most common quality indices are based on a combination of these three criteria. For instance, the Jaccard index J' [62] is defined as:

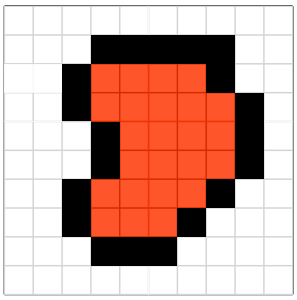
$$J'(N, G) = \frac{|N \cap G|}{|N \cup G|} = \frac{TP}{TP + FP + FN} \tag{7.1}$$

while the Dice coefficient D [63] is expressed as:

$$D(N, G) = \frac{2.|N \cap G|}{|N| + |G|} = \frac{2.TP}{2.TP + FP + FN} \tag{7.2}$$

[1]Let us remind that the term "metric" here is related to the quality of a BPT and is different to the "metric of construction" described earlier in this document.

Ground-truth segment G



Distance map

2	2	1	1	1	1	1	1	1	2
2	1	1	0	0	0	0	0	1	2
2	1	0	−1	−1	−1	−1	0	1	1
2	1	0	−1	−1	−2	−1	−1	0	1
2	1	1	0	−1	−2	−2	−1	0	1
2	1	1	0	−1	−2	−1	−1	0	1
2	1	0	−1	−1	−1	−1	0	1	1
2	1	0	−1	−1	−1	0	1	1	2
2	1	1	0	0	0	1	1	2	2
2	2	1	1	1	1	1	2	2	3

Membership map where $\alpha = 1.2$

0.08	0.08	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.08
0.08	0.23	0.23	0.5	0.5	0.5	0.5	0.5	0.23	0.08
0.08	0.23	0.5	0.77	0.77	0.77	0.77	0.5	0.23	0.23
0.08	0.23	0.5	0.77	0.77	0.92	0.77	0.77	0.5	0.23
0.08	0.23	0.23	0.5	0.77	0.92	0.92	0.77	0.5	0.23
0.08	0.23	0.23	0.5	0.77	0.92	0.77	0.77	0.5	0.23
0.08	0.23	0.5	0.77	0.77	0.77	0.77	0.5	0.23	0.23
0.08	0.23	0.5	0.77	0.77	0.77	0.5	0.23	0.23	0.08
0.08	0.23	0.23	0.5	0.5	0.5	0.23	0.23	0.08	0.08
0.08	0.08	0.23	0.23	0.23	0.23	0.23	0.08	0.08	0.02

Figure 7.4: Illustration of a ground-truth segment example G, the corresponding distance map computed from $\sigma_G(x)$ and the corresponding membership map computed from $\mu_\alpha(d)$ where $\alpha = 1.2$.

The Jaccard and Dice indices J' and D, but also TP, FP and FN (or, possibly, their normalized ratio w.r.t. $|N|$ or $|G|$) are examples of functions Λ. They essentially rely on a region (i.e., combinatorial) paradigm, that consists of "counting" pixels.

Since a ground-truth segment example G is assumed to be provided by a human expert, it may be imperfect. This uncertainty mainly comes from its delineation (generally operated by photo-interpretation of the image), i.e. the definition of the contour of the segment. In particular, *the closer a pixel from a segment contour, the less probable its actual correctness.* This statement motivates the computation, for each segment G, of a (signed) distance map [71] to the border of the segment. More precisely, a function $\sigma_G : \Omega \to \mathbb{R}$ is computed, and provides, for each pixel $x \in \Omega$, the distance between x and the border of G. In particular, we set $\sigma_G(x) > 0$ (resp. < 0) outside (resp. inside) G in order to differentiate external and internal pixels (see the bottom left part of Figure 7.4).

The degree of uncertainty on G can then be expressed by associating, to each distance value, a probability of correct membership of x to G. Such a membership function is defined as $\mu : \mathbb{R} \to [0, 1]$, with the constraint of

being decreasing, and verifying $\lim_{-\infty} \mu = 1$, $\lim_{+\infty} \mu = 0$, and $\mu(0) = 0.5$. The membership function μ considered here is defined as a sigmoid function $\mu_\alpha(d) = 1/(1+e^{\alpha d})$, where $\alpha > 0$ allows us to control the degree of fuzziness (see the right part of Figure 7.4).

Based on this uncertain framework, the notions of true positives, false positives and false negatives can be reformulated[2] as:

$$TP_\alpha(N, G) = \int_N \mu_\alpha(\sigma_G(x)).dx \tag{7.3}$$

$$FP_\alpha(N, G) = \int_N (1 - \mu_\alpha(\sigma_G(x))).dx \tag{7.4}$$

$$FN_\alpha(N, G) = \int_{\Omega \setminus N} \mu_\alpha(\sigma_G(x)).dx \tag{7.5}$$

Practically, the true and false positives can be easily computed as $TP_\alpha(N, G) = \sum_{x \in N} \mu_\alpha(\sigma_G(x))$ and $FP_\alpha(N, G) = |N| - \sum_{x \in N} \mu_\alpha(\sigma_G(x))$. The evaluation of the false negatives requires, in theory, a whole computation over Ω, which is not tractable in practice. By assuming that μ rapidly converges onto 0, it is however sufficient to compute μ in a neighbourhood superset $\Delta G \supseteq G$ of G, and $FN_\alpha(N, G)$ can then be fairly approximated as $\sum_{x \in \Delta G \setminus N} \mu_\alpha(\sigma_G(x))$.

Some uncertain versions of classical scores Λ such as the Jaccard index J' and the Dice index D can be simply obtained by embedding Eqs. (7.3–7.5) into Eqs. (7.1–7.2). More generally, the use of uncertain notions of true / false positives and false negatives allows us, on the one hand, to take into account the low confidence of the border of the segments and, on the other hand, to introduce some contour-based information into standard region-based metrics, by penalizing mismatching errors far from the ground-truth segment example contours.

7.3 Finding matching nodes in a Binary Partition Tree

Once defined a similarity metric for nodes / ground-truth segments, the algorithmic question of determining, for each ground-truth segment, what is the most similar node within the BPT, is raised. Given k ground-truth segments G_i ($i \in [\![1, k]\!]$), and a BPT $\mathfrak{T}$ composed of n nodes N_j ($j \in [\![1, n]\!]$),

[2] For $\alpha \to +\infty$, we retrieve the standard notions of TP, FP and FN.

we then have to solve k times the following optimization problem

$$N_\star^i = \arg\max_{j \in [\![1,n]\!]} \Lambda(N_j, G_i) \tag{7.6}$$

A brute-force approach would be to browse, for each ground-truth segment G_i, the whole BPT. This would lead to compute $\Lambda(N_j, G_i)$ for all $(i,j) \in [\![1,k]\!] \times [\![1,n]\!]$ with, at least, a time cost $\mathcal{O}(k.n)$ (the cost of Λ computation also has to be considered). For instance, with a 2D image of size $1\,000 \times 1\,000$, with an initial set of leaves where each leaf represents a single pixel, the number of nodes n would be $2.1000^2 - 1 \simeq 2.10^6$, leading to a total cost of $2k.10^6$ Λ computations. It is then crucial to reduce this cost, by avoiding useless node and ground-truth segment comparisons.

To this end, let us recall that our objective is to optimize the matching between a ground-truth segment G and a node N. On the one hand, we want to maximize the intersection between G and N (i.e., $|G \cap N|$, and then get a high value of true positives TP). On the other hand, we want to minimize the mismatching between G and N (i.e., $|S \setminus N|$ and $|N \setminus S|$), and then get low values of false positives FP and false negatives FN. These two goals are often antagonistic, and classical indices (e.g., Eqs. (7.1–7.2)), handle the trade-off between them.

Nevertheless, this provides us useful information for setting spatial constraints and quantitative heuristics that optimize the matching nodes finding process. By using specific vertical and horizontal selections, the search space reduction is almost similar to the subtree $\mathfrak{T}_G$ extraction proposed in Section 6.1. We can use the subtree $\mathfrak{T}_G$ in this process but using the spatial constraints and the quantitative heuristics that we present in this section allows to find the matching nodes directly during the process without performing a supplementary subtree extraction.

Spatial constraints: vertical selection The intersection between G and N is mandatory to guarantee a non-null value of TP, itself required to avoid null values of standard quality indices as J' or D. A node N that does not intersect G is built, in the BPT, from the union of leaves that present the same property. By contraposition, a node that intersects G is composed by at least one leaf that also intersects G. In particular, for finding an optimal node $N_\star$ it is then sufficient, for a given ground-truth segment G, to restrict

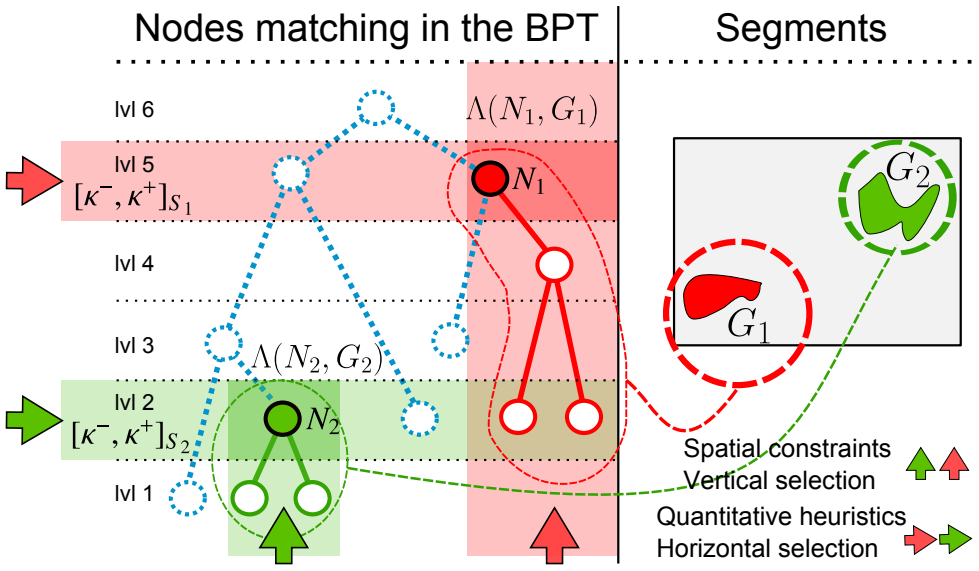


Figure 7.5: Illustration of the spatial constraints (vertical selection) and quantitative heuristics (horizontal selection) employed to optimize the search of matching nodes in the BPT.

our study to the BPT branches whose ending leaves also intersect G. This provides a way to "vertically" restrict the search space within the BPT, by only considering the branches with such leaves (see vertical arrows in Figure 7.5).

Quantitative heuristics: horizontal selection To correctly match a ground-truth segment G, a node N has to be of comparable size. Indeed, if we have $|N| \ll |G|$, the TP value will be low and the FN value high with respect to the size of G. On the contrary, if we have $|N| \gg |G|$, the FP value will be high. Consequently, having $|N| \simeq |G|$ is a necessary (yet non-sufficient) condition for obtaining satisfactory matching scores between nodes and segments. Practically, it is then relevant to restrict the actual computation of the Λ values to nodes N with size within a confidence interval $[\kappa^-, \kappa^+] \subseteq \mathbb{N}$, with $\kappa^- < |G| < \kappa^+$. This is indeed a heuristic reduction of the search space, that consists of selecting an "horizontal" set of nodes within the BPT (see horizontal arrows in Figure 7.5). In particular, the choice of the κ values has to be wisely made, to handle the trade-off between time cost reduction and near-optimal matching node determination.

These two strategies for selecting candidate nodes to solve Eq. (7.6) al-

low for reducing the computational time cost. In particular, we consider a bottom-up approach, initialized with the set of leaves intersecting G, and directly climbing the branches up to the first nodes of size κ^-. At this stage, the scores Λ are explicitly computed up to the nodes of size κ^+; then the climbing stops and the score $\Lambda(N_\star, G)$ of the best node is kept as the quality score $\lambda(G)$ of $\mathfrak{T}$ for the ground-truth segment G. In certain cases, the properties of the chosen metric Λ could lead to algorithmic optimizations; for instance, separable [26, 27] and, more generally, hierarchically increasing metrics [72] avoid an exhaustive computation of the Λ scores over the whole set of nodes.

7.4 Global quality score

By performing this optimization process for each ground-truth segment G_i, we obtain a set of k nodes $N_\star^i$, associated with a (near-)optimal quality score $\lambda(G_i) = \Lambda(G_i, N_\star^i)$. The issue is then to gather the information provided by these local quality scores to finally define a global quality metric Γ that will express the quality of the BPT (i.e., its ability to fit at best the set of ground-truth segments).

The very first idea for defining Γ is indeed to compute an average value of all the local quality scores (i.e., $\Gamma = 1/k.\sum_{i=1}^{k}\lambda(G_i)$). However, since we assume that each ground-truth segment is endowed with a given label (or concept), it is important to consider this semantic information in order to avoid potential bias effects. For instance, if 25% of the ground-truth segments have a label A, while the other 75% have a label B, a simple mean value Γ for a BPT with an average quality value of 0.1 for the class A and 0.9 for the class B will lead to a global quality score of $\frac{1}{4}\times 0.1 + \frac{3}{4}\times 0.9 = 0.7$, which is non-relevant if both classes have comparable importance, that should lead to a 0.5 global score. It is then crucial to model in the global measure the relative importance of each label / class of segment. In addition, it may be also relevant to model the relative importance of each segment within each class.

To this end, we define a weighted formulation of the global quality score Γ as

$$\Gamma = \sum_{\ell \in L} w_\ell \sum_{G_i \in \mathcal{C}_\ell} w_i.\lambda(G_i) \tag{7.7}$$

with $\sum_{\ell \in L} w_\ell = 1$, $\sum_{G_i \in C_\ell} w_i = 1$ for all l, and $w_\star \geq 0$, where L is the label set and $\mathcal{C}_\ell$ are the different semantic classes of ground-truth segments.

The weights w_ℓ can be used to assess the relative importance of each semantic class $\mathcal{C}_\ell$. In particular, $w_\ell = 1/|L|$ if each class has the same importance. The weights w_i can be used for normalizing the local quality metric (e.g., in the case of extensive metrics as TP), and / or to discriminate the importance of each ground-truth segment (i.e., the necessity to correctly segment it). They can also be used for quantifying the relevance of a ground-truth segment; for instance, if these segments are obtained from a crowd-sourcing campaign, the w_i weight may be proportional to the confidence assigned to G_i.

It is worth mentioning that the use of semantic information for designing the global quality metric also argues in favour of the possible design of non-linear definitions of Γ. Two specific formulations can be proposed: $\Gamma_{\min} = \min_{\ell \in L} \sum_{G_i \in C_\ell} w_i.\lambda(G_i)$ and $\Gamma_{\max} = \max_{\ell \in L} \sum_{G_i \in C_\ell} w_i.\lambda(G_i)$. The first allows one to characterize BPTs that are able to efficiently characterize *all* the classes of objects provided as ground-truth; the second provides a way to discriminate BPTs that detect (at least) one among a set of given classes.

In the next chapter (Chapter 8), experiments on remote sensing images emphasize the usefulness of our supervised BPT quality evaluation framework. This framework also enables to quantitatively evaluate the interest of our multi-feature generalization of the BPT construction framework.

7.5 Résumé : Analyse de la qualité extrinsèque d'un BPT

Dans ce chapitre, nous proposons une approche pour évaluer la qualité d'un BPT donné en ne se basant pas seulement sur des informations relatives à sa structure intrinsèque, mais en tirant avantage de certaines informations extrinsèques. Nous proposons une nouvelle approche supervisée pour l'évaluation extrinsèque de la qualité d'un BPT. Dans un premier temps, le but n'est pas de trouver une partition idéale de l'image, mais d'évaluer la capacité d'un BPT à construire des nœuds qui correspondent au mieux à un sous-ensemble d'exemples de segments de références défini par l'utilisateur.

En d'autres termes, notre carte de vérité-terrain n'est pas globale mais partielle. Dans un deuxième temps, comme les segments de vérité-terrain sont définis par des utilisateurs humains, nous supposons qu'ils peuvent être imparfaits, et nous intégrons l'incertitude induite dans notre processus d'évaluation. De plus, ces segments de vérité-terrain peuvent être labellisés par l'utilisateur et nous permettent ainsi d'intégrer des critères sémantiques dans l'approche d'évaluation, et améliorer la robustesse du score de qualité globale.

Chaîne de traitement Une fois la carte de vérité-terrain définie par l'expert, notre approche pour l'évaluation supervisée de la qualité d'un BPT se base sur trois composantes importantes : (i) une métrique locale qui évalue quantitativement le degré de correspondance entre le segment de référence G et le nœud N du BPT $\mathfrak{T}$; (ii) la recherche de nœuds du BPT qui maximisent localement leurs correspondances avec les segments de références; et (iii) un score de qualité globale qui fusionne les scores locaux pour chaque segment de vérité-terrain. La prise en compte de l'incertitude de la vérité-terrain est principalement considérée dans la métrique locale, alors que les labels sémantiques des segments de références sont principalement utilisés dans la définition de la mesure globale de la qualité d'un BPT.

Métriques de correspondances nœuds / segment en milieu incertain Comparer un nœud N et un segment de référence G consiste à quantifier le degré de similarité entre eux, à travers un score local $\Lambda(N, G) \in \mathbb{R}^+$, où Λ est une métrique ou pseuso-métrique qui évolue de manière monotonique en respectant cette similarité. D'une façon basique, une telle évaluation de la similarité peut être effectuée dans une dynamique basée région, en considérant les informations de superposition spatiale (i.e., en calculant les vrais positifs (TP), les faux positifs (FP) et les vrais négatifs (FN) entre N et G, définis par $TP = |N \cap S|$, $FP = |N \backslash G|$, et $FN = |G \backslash N|$, respectivement). Les indices de qualité les plus courants sont basés sur la combinaison de ces trois critères. Par exemple, l'indice de Jaccard $J'(N, G) = \frac{|N \cap G|}{|N \cup G|} = \frac{TP}{TP+FP+FN}$ et le coefficient de Dice $D(N, G) = \frac{2.|N \cap G|}{|N|+|G|} = \frac{2.TP}{2.TP+FP+FN}$. Le segment de référence G étant fournit par l'utilisateur humain, il est possible qu'il soit

imparfait. L'incertitude provient principalement des contours des segments qui sont généralement définis par photo-interprétation. En particulier, plus un pixel est proche du contour du segment, moins sa classification sera sure. Cette supposition motive le calcul, pour chaque segment G, d'une carte contenant pour chaque pixel $x \in \Omega$ une valeur de distance entre x et la bordure de G. Cette valeur de distance est donnée par la fonction $\sigma_G : \Omega \to \mathbb{R}$, et permet de différencier les pixels internes (i.e., $\sigma_G(x) > 0$) et les pixels externes ($\sigma_G(x) < 0$). Le degré d'incertitude sur G peut être exprimé en associant, pour chaque distance, une probabilité d'appartenance de x dans G. Une telle fonction d'appartenance est définie par $\mu : \mathbb{R} \to [0, 1]$ vérifiant $\lim_{-\infty} \mu = 1$, $\lim_{+\infty} \mu = 0$, et $\mu(0) = 0.5$. La fonction μ est définie par une fonction sigmoïde $\mu_\alpha(d) = 1/(1 + e^{\alpha d})$, où $\alpha > 0$ permet le contrôle sur le degré du flou. En se basant sur ces aspects incertain, les notions de vrais positifs, de faux positifs et de faux négatifs peuvent être reformulées tel que $TP_\alpha(N, G) = \int_N \mu_\alpha(\sigma_G(x)).dx$, $FP_\alpha(N, G) = \int_N (1 - \mu_\alpha(\sigma_G(x))).dx$ et $FN_\alpha(N, G) = \int_{\Omega \setminus N} \mu_\alpha(\sigma_G(x)).dx$. Dans la pratique, les vrais et les faux positifs peuvent être facilement calculés avec $TP_\alpha(N, G) = \sum_{x \in N} \mu_\alpha(\sigma_G(x))$ et $FP_\alpha(N, G) = |N| - \sum_{x \in N} \mu_\alpha(\sigma_G(x))$. L'évaluation des faux négatifs nécessite, en théorie, un calcul total sur tout Ω, ce qui n'est pas évident dans la pratique. En supposant que μ converge rapidement vers 0, il est cependant suffisant de calculer μ dans un sous-ensemble voisin $\Delta G \supseteq G$ de G, et $FN_\alpha(N, G)$ peut être approximé puisque $\sum_{x \in \Delta G \setminus N} \mu_\alpha(\sigma_G(x))$. Des versions incertaines des scores Λ classiques comme l'indice de Jaccard J' et le coefficient de Dice D peuvent être obtenues en utilisant le $TP_\alpha(N, G)$, $FP_\alpha(N, G)$ et $FN_\alpha(N, G)$ dans leurs formules. D'une façon générale, l'utilisation des notions incertaines du vrai / faux positifs et du faux négatif nous permet, d'une part, de prendre en compte la faible confiance sur la bordure des segments, et d'autre part, d'introduire des informations basées contours dans les métriques basées régions classiques, en pénalisant les erreurs de non correspondance des points qui sont éloignés de la bordure du segment de référence.

Recherche de correspondances de nœuds dans un BPT Une fois la métrique de similarité entre nœuds et segments de référence déterminée, notre but est de trouver pour chaque segment G le nœud N le plus correspondant. Pour k

segments de référence G_i ($i \in [\![1, k]\!]$), et un BPT $\mathfrak{T}$ composé de n nœuds N_j ($j \in [\![1, n]\!]$), nous devons résoudre k fois le problème d'optimisation $N^i_\star = \arg\max_{j \in [\![1,n]\!]} \Lambda(N_j, G_i)$. Une approche de brute-force serait de parcourir le BPT entièrement, pour chaque segment de référence. Cela emmène à calculer $\Lambda(N_j, G_i)$ pour tout $(i, j) \in [\![1, k]\!] \times [\![1, n]\!]$ avec, au minimum, un coût en temps de calcul $\mathcal{O}(k.n)$ tout en considérant le coût de Λ en terme de calcul. Il est alors cruciale de réduire ce coût, en évitant des comparaisons inutiles entre nœuds et segments de référence. Notre but est d'optimiser la correspondance entre un nœud N et un segment de référence G. D'un côté, nous souhaitons maximiser l'intersection entre G et N (i.e., $|G \cap N|$), et avoir une valeur élevée de vrais positifs TP. D'un autre côté, nous souhaitons minimiser la discordance entre G et N (i.e., $|G \setminus N|$ et $|N \setminus G|$), et ainsi avoir un faible nombre de faux positifs FP et de faux négatifs FN. Ces buts sont souvent antagonistes mais nous permettent de définir une contrainte spatiale et une heuristique quantitative pour optimiser le processus de recherche de nœuds pertinents. L'intersection entre G et N est nécessaire pour garantir une valeur non nulle de TP et évitant également des valeurs non nulles pour l'indice de qualité tel que J' ou D. Un nœud qui n'intersecte pas G est a été créé à partir de l'union de deux feuilles qui n'intersectent pas G non plus. Un nœud qui intersecte G est composé, au minimum, d'une feuille qui intersecte aussi G. En particulier, pour trouver un nœud optimale $N_\star$, il est alors suffisant de restreindre l'étude sur des branches du BPT qui contiennent des feuilles intersectant G. Cela nous donne un moyen de restreindre verticalement l'espace de recherche contenu dans le BPT, en ne considérant que les branches pertinentes. D'un autre point de vue, afin de bien correspondre au segment de référence G, une nœud N devrait avoir une taille comparable à celle de G. Dans le cas contraire, si nous avons $|N| \gg |G|$, alors la valeur de FP sera élevée. En conséquence, avoir $|N| \simeq |G|$ est une condition nécessaire (pas suffisante) pour l'obtention d'un score de correspondance satisfaisant entre nœuds et segments de référence. En pratique, il est pertinent de restreindre le calcul des valeurs de Λ à celles des nœuds ayant une taille incluse dans un intervalle de confiance $[\kappa^-, \kappa^+] \subseteq \mathbb{N}$, avec $\kappa^- < |G| < \kappa^+$. Cela est, en effet, une heuristique de réduction de l'espace de recherche qui consiste à sélectionner d'une manière horizontale un ensemble de nœuds du BPT. Les

deux stratégies proposées, pour la sélection de nœuds candidats permettent la réduction du coût de calcul du processus de recherche. En particulier, nous considérons une approche ascendante, initialisée avec l'ensemble de feuilles intersectant G qui remonte directement dans les branches jusqu'aux premiers nœuds ayant la taille κ^-. Á ce niveau, les calculs des scores Λ commencent jusqu'aux nœuds ayant la taille κ^+. Au delà de cette limite, le processus s'arrête en gardant en mémoire le score $\Lambda(N_\star, G)$ du meilleur nœud, qui sera considéré comme le score de qualité $\lambda(G)$ de $\mathfrak{T}$ pour le segment de référence G.

Score global de qualité En effectuant les procédures d'optimisation proposées précédemment, nous obtenons un ensemble de k nœuds $N_\star^i$ pour chaque segment de référence G_i. Chaque $N_\star^i$ est associé à un score (presque) optimal $\lambda(G_i) = \Lambda(G_i, N_\star^i)$ Il est maintenant nécessaire de rassembler ces informations pour finalement définir une métrique de score de qualité globale Γ qui va exprimer la qualité du BPT (i.e., sa capacité à fournir des résultats correspondant aux segments de vérité-terrain). La première idée est de définir $\Gamma = 1/k.\sum_{i=1}^{k} \lambda(G_i)$ en calculant une moyenne des valeurs des scores locaux. Cependant, comme nous supposons que chaque segment de référence est associé à un label donné, il est important de considérer ces informations sémantiques dans le but de réduire les effets de biais potentiels. Par exemple, si 25% des segments de vérité-terrain ont un label A, alors que 75% ont un label B, une simple valeur moyenne Γ pour un BPT avec une valeur de qualité moyenne 0.1 pour la classe A et 0.9 pour la classe B donnera un score de qualité $\frac{1}{4} \times 0.1 + \frac{3}{4} \times 0.9 = 0.7$, qui n'est pas pertinent si les deux classes ont une importance similaire, cela devrait donner un score global de 0.5. Il est alors crucial de modéliser dans la mesure globale l'importance relative de chaque label / classe ds segments. De plus, il est également pertinent de modéliser l'importance relative de chaque segment dans chaque classe. Pour cela, nous définissons une formulation pondérée du score de qualité globale $\Gamma = \sum_{\ell \in L} w_\ell \sum_{G_i \in \mathcal{C}_\ell} w_i.\lambda(G_i)$, avec $\sum_{\ell \in L} w_\ell = 1$, $\sum_{G_i \in C_\ell} w_i = 1$, et $w_\star \geq 0$, où L est l'ensemble des labels et $\mathcal{C}_\ell$ sont les différentes classes sémantiques des segments de référence. Les poids w_ℓ peuvent être utilisés pour évaluer l'importance relative de chaque classe sémantique $\mathcal{C}_\ell$. En

particulier, $w_\ell = 1/|L|$ si chaque classe avait la même importance. Les poids w_i peuvent être utilisés pour normaliser la métrique de score local (e.g., pour le cas des métriques extensives comme TP), et / ou discriminer l'importance de chaque segment de référence (i.e., la nécessité de correctement le segmenter). Ces poids peuvent aussi être utilisés pour quantifier la pertinence d'un segment de référence. Par exemple, si les segments sont obtenus à partir d'un crowd-sourcing, le poids w_i pourrait être proportionnel à la confiance assignée à G_i. L'utilisation des informations sémantiques pour la conception d'une métrique de qualité globale argumente en faveur d'une possibilité de conception de définitions non-linéaires de Γ comme $\Gamma_{\min} = \min_{\ell \in L} \sum_{G_i \in C_\ell} w_i.\lambda(G_i)$ et $\Gamma_{\max} = \max_{\ell \in L} \sum_{G_i \in C_\ell} w_i.\lambda(G_i)$. La première définition permet la caractérisation des BPTs qui sont capables de caractériser efficacement les classes d'objets provenant de la vérité-terrain. La deuxième permet de discriminer les BPTs qui détectent (au minimum) une classe parmi toutes les classes.

Chapter 8

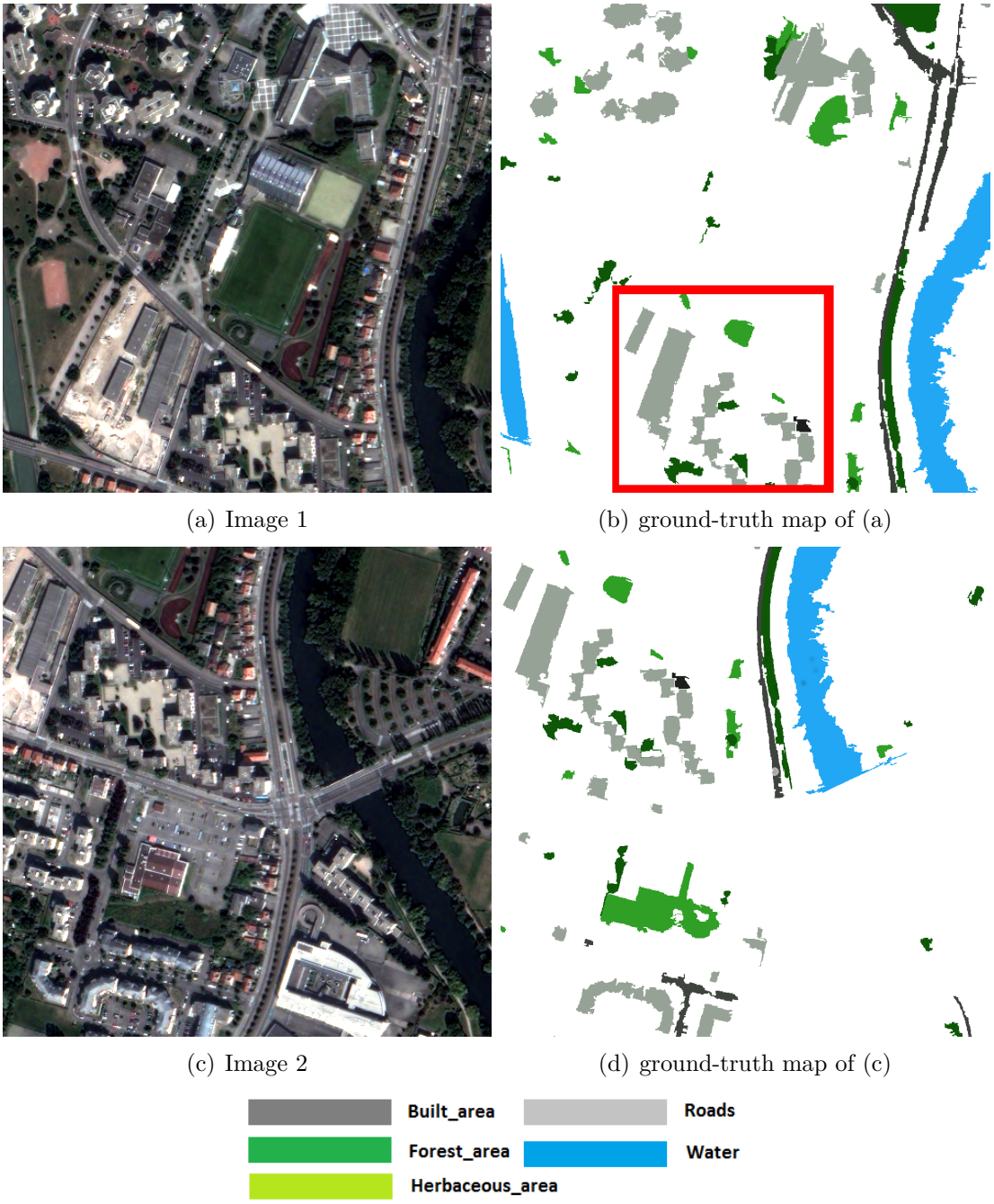
Experimental studies

Contents

In this chapter, we present different experiments to illustrate the interest of our BPT quality evaluation frameworks. First, we present the material used for the experiments (Section 8.1). Second, experiments involving combinatorial and quantitative evaluation highlight the interest of using the intrinsic quality analysis of BPTs (Section 8.2). Then, we present a set of results emphasizing the usefulness of the extrinsic quality analysis in order to assess the quality of BPTs, and also to evaluate the MBPT construction framework that we proposed (Section 8.3). Finally, we discuss about the advantages and the limits of our evaluation frameworks (Section 8.4).

8.1 Data and ground-truth map

In this section, we present the datasets used for the whole experiments about the evaluation frameworks. Two main geographical areas are considered during this study. The first is the town of Strasbourg, which is a



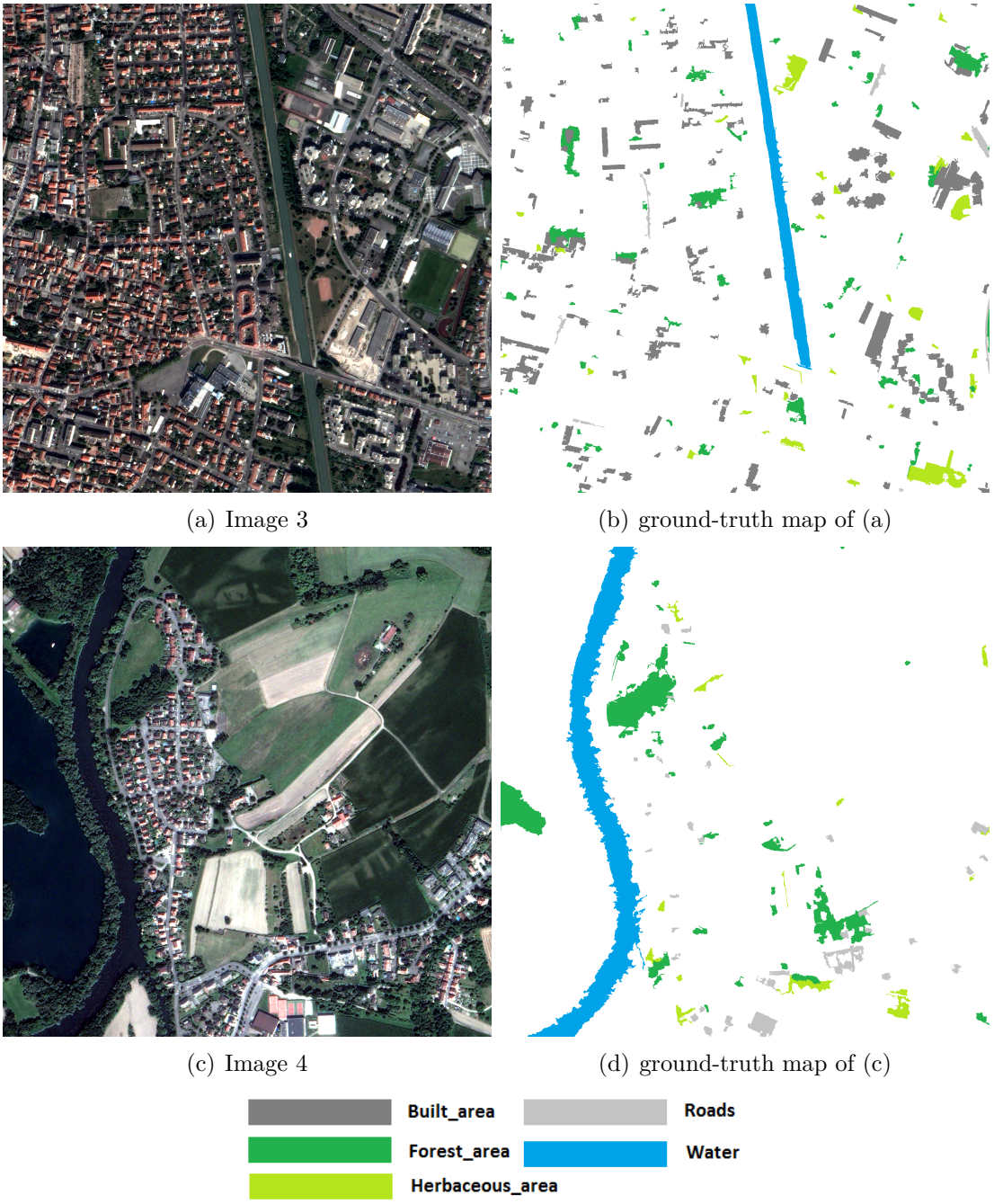
(a) Image 1 (b) ground-truth map of (a)

(c) Image 2 (d) ground-truth map of (c)

Figure 8.1: (a, c) VHSR images of 1 000 × 1 000 pixels at a spatial resolution of 60 cm. (b, d) Ground-truth maps with reference segments belonging to 5 (coloured) semantic classes.

urban area, while the second is an agricultural area located near Toulouse. In fact, the images used for the experiments derive from the same source of the datasets presented in Chapter 4.

Images of Strasbourg The first dataset (courtesy LIVE, UMR CNRS 7263) contains four VHSR images, sensed over the town of Strasbourg (France)



(a) Image 3 (b) ground-truth map of (a)

(c) Image 4 (d) ground-truth map of (c)

Figure 8.2: (a, c) VHSR images of 2 000 × 2 000 pixels at a spatial resolution of 60 cm. (b, d) Ground-truth maps with reference segments belonging to 5 (coloured) semantic classes.

by the PLÉIADES satellite, in 2012. A part of this dataset was already presented in Section 4.1 but for the sake of readability, we recall the principal information in this section.

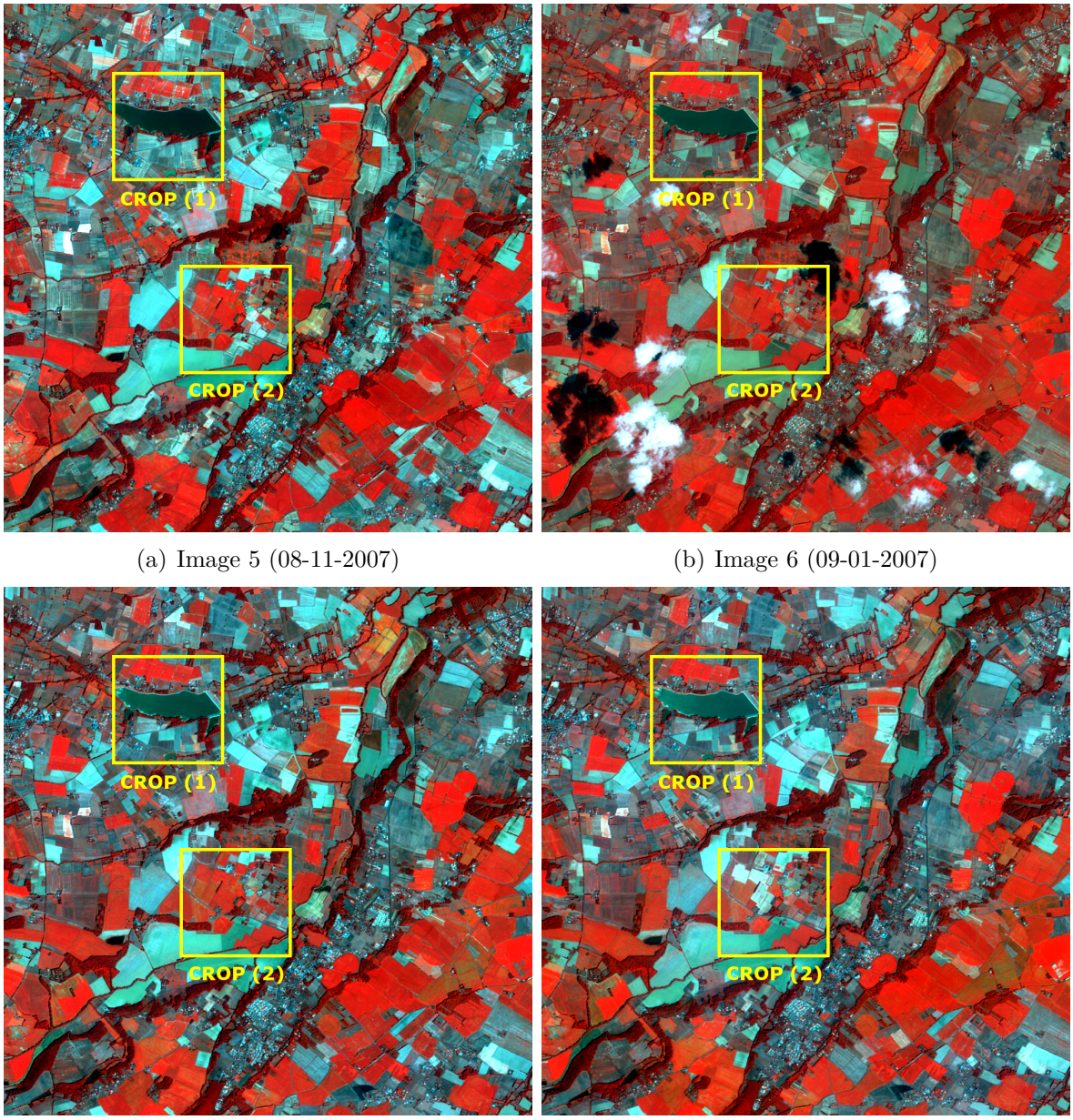
The first sample contains two images of 1 000 × 1 000 pixels (see Image 1 and Image 2 of Figure 8.1(a, b)) while the second sample is composed by

two images of $2\,000 \times 2\,000$ pixels (see Image 3 and Image 4 of Figure 8.2).

Ground-truth maps of different urban objects represented in the scene are also available. These maps were derived from a public crowd-sourcing campaign in the context of the COCLICO research project. The protocol for building this ground-truth map was the following: first, the satellite image was segmented in different semantic objects (buildings, vegetation areas, roads, bare soils, etc.) by an expert with a visual interpretation of the VHSR image combined to a supervised segmentation tool. Then, these segments were involved in a public crowd-sourcing campaign, where different questions were asked in order to evaluate the quality of each segment and its actual correspondence (i.e., region borders, semantic label) with a real urban object sensed by the image. We only retained the reference segments that led to the highest consensus between the crowders. The segments were labelled with $L = 5$ different semantic classes: *built area*, *forest area*, *herbaceous area*, *roads*, *water*. Figure 8.1(b, d) and Figure 8.2(b, d) present the partial ground-truth map of Figure 8.2(a, c) and Figure 8.2(a, c).

Images of an area near Toulouse The second dataset used here is a time series of four images ($1\,000 \times 1\,000$ pixels) sensed over an area located near Toulouse (France) by the FORMOSAT-2 satellite over the 2007 cultural year (see Section 4.2). For the sake of readability, Figure 8.3 reintroduces the four images that were already presented in Figure 4.10. Let us remind that they were all acquired in late August and September, in order to reduce temporal evolution effects, since we do not aim at carrying out time-based analysis.

A ground-truth map of different crop fields represented in the scene is also available (Figure 8.4). This map is derived from a field survey produced by the European Environment Agency (`http://ec.europa.eu/agriculture`). The semantic classes associated to this reference map reflect the temporal behavior of the considered crops over the 2007 cultural year. However, we only considered the delineation of the 198 segments of reference in the image for the quantitative evaluation of the quality of the produced BPTs.



(a) Image 5 (08-11-2007)

(b) Image 6 (09-01-2007)

(c) Image 7 (09-08-2007)

(d) Image 8 (09-15-2007)

Figure 8.3: Four satellite images ($1\,000 \times 1\,000$ pixels) sensed by the FORMOSAT-2 satellite covering the same geographical area (see Section 4.2). Note that Image 6 is partially altered by the presence of clouds.

8.2 Intrinsic quality experiments

In this section, we present some experiments based on the intrinsic quality analysis of a BPT presented in Chapter 6. Our aim is to give the intuition on how an intrinsic analysis can be used for the quality evaluation of hierarchical structures such as the BPTs.

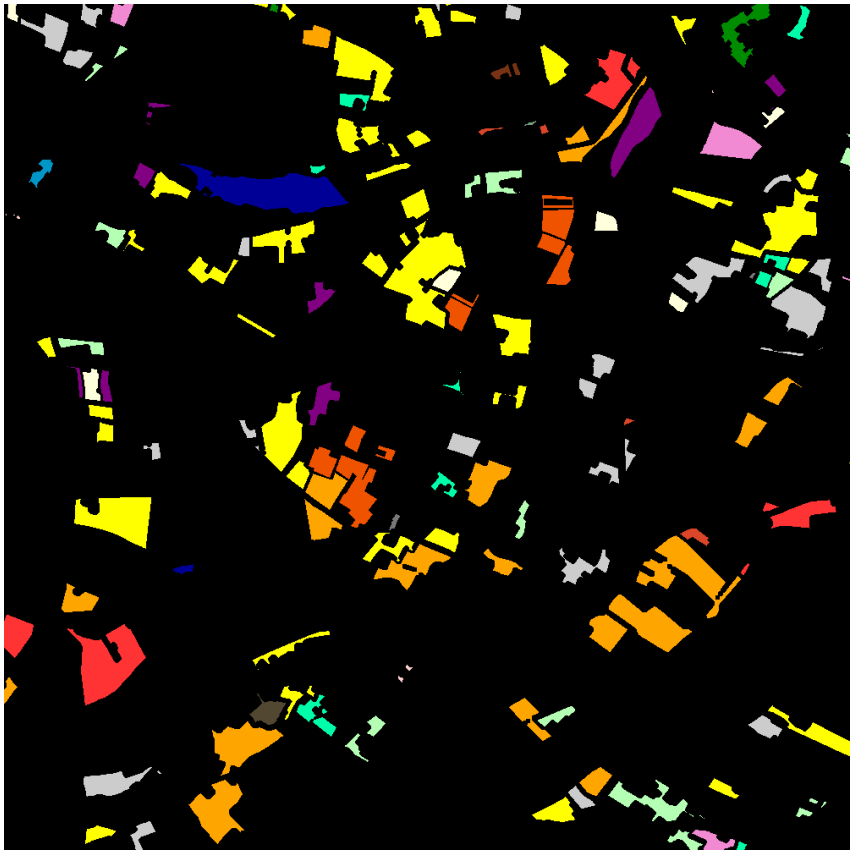
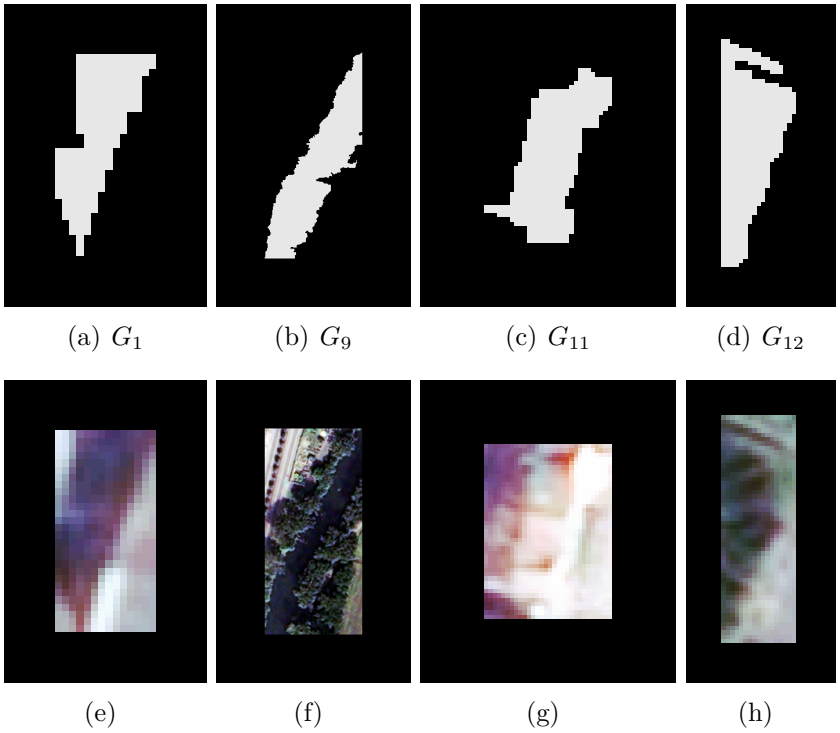


Figure 8.4: Ground-truth map representing partial crop fields of the studied area corresponding to the 4 images of Figure 8.3. This map contains 198 segments of reference delivered by the European Environment Agency.

Experimental protocol

In this experiment, we used a crop (500×500 pixels) from Image 1 (see Figure 8.1(a)). From this crop, three types of BPTs were built using various metrics. They correspond to the BPT_{std}, the BPT_{ndvi} and the BPT_{ndwi} respectively built by using the metric W_{colour}, the metric W_{ndvi} and the meric W_{ndwi} described in Section 4.1. For each type of BPT, four BPTs are built from an initial partition $\mathcal{L}$ composed of all pixels from the image support (i.e., 250 000 regions), 6 250 regions, 12 500 regions and 18 750 regions respectively. We used our intrinsic framework to estimate the ability of these BPTs to generate potential nodes well matching the set of 13 ground-truth examples. Four of them are illustrated in Figure 8.5. For each individual segment contained in the ground-truth map, we first extract a subtree (see Section 6.1). Then, we compute the granularity index and the discordance value of the set $\mathcal{L}_G$ of each subtree (see Section 6.2)



(a) G_1 (b) G_9 (c) G_{11} (d) G_{12}

(e) (f) (g) (h)

Figure 8.5: Examples of ground-truth segments used during the experiments about the intrinsic evaluation of BPTs. (a–d) Ground-truth examples G_i, in white. (e–h) The image area around each ground-truth example.

and we present a combinatorial (Section 6.3) and a quantitative analysis (Section 6.3).

In order to compute combinatorial scores normalized between 0 and 1, we define $comb1 = (b_{pp} + b_{ii})/(u_i + b_{pp} + b_{ii} + b_{pi})$ and $comb2 = b_{pp}/(b_{pp} + b_{pi})$. The definitions of u_i, b_{pp}, b_{ii} and b_{pi} are presented in Table 6.1. For a ground-truth example, the hierarchical structure is able to provide an acceptable matching region if those scores are above 0.5. The closer these values from 1 (resp. 0), the more (resp. less) acceptable BPT.

We used the definitions proposed in Section 6.3 for the quantitative analysis. For a better readability, we define $quant1 = |N_G \setminus G| - |(\cup_{L \in \mathcal{L}_G} L) \setminus G|$ and $quant2 = |N_G \setminus G| / |(\cup_{L \in \mathcal{L}_G} L) \setminus G|$. We also simplified some notations by using $n_p = (\Sigma_{N \in \mathcal{N}_p} |N|)$ and $l_p = (\Sigma_{N \in \mathcal{L}_p} |N|)$. Thus we can define $quant3 = (|G| - n_p) - (|G| - l_p)$ while $quant4$ here corresponds to the size of $\mathcal{N}_p$ (i.e., the number of nodes which are required to form the best segmentation lower than G).

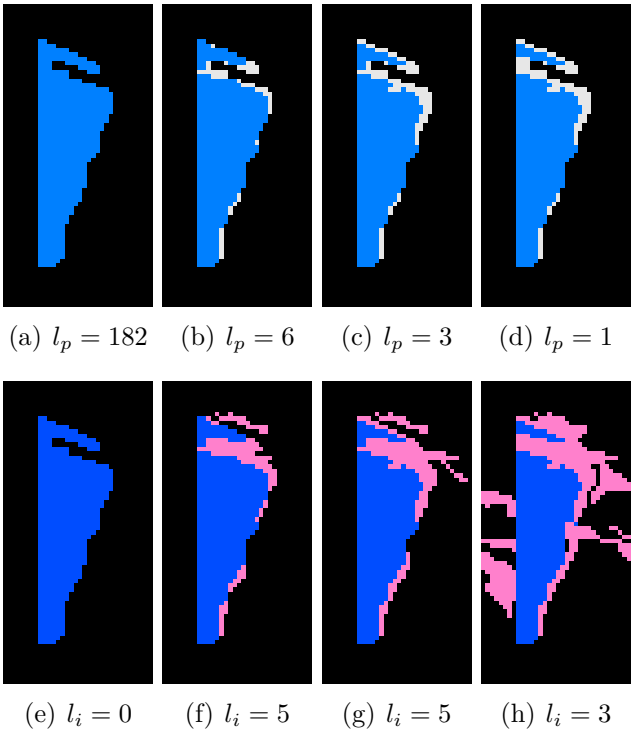
Results and discussions

Table 8.1 presents the combinatorial scores obtained from a BPT_{std} built from an initial partition $\mathcal{L}$ corresponding to all of the pixels from the image support (i.e., 250 000 initial regions). Before starting the discussion, let us first remind that a high value of granularity and a low value of discordance suggest the pertinence of the study. For this BPT_{std}, these values are optimal and a relevant quality estimation could emerge from this analysis. For this tree, we can notice that the scores are generally around 0.5 which is our point of reference. At a first glance, Table 8.1 suggests that the BPT_{std} built from its partition of pixels has an intrinsic structure that can provide acceptable segmentation result for almost each ground-truth example. However, we can observe that the scores are not that high even if the construction of the BPT_{std} started from its pixels. Indeed, the values of u_i, b_{pp}, b_{ii} and b_{pi} are still far to the values of reference, in brackets, that illustrate what a perfect structure could theoretically be (see Eqs. (6.3–6.6)).

Table 8.1: Combinatorial scores of the quality of a BPT_{std} obtained from a partition containing all pixels (i.e., granularity = 1 and discordance = 0). Each line represents the score for a particular ground-truth example G_i. The best theoretical values for unary and binary nodes are provided between brackets.

G	l_i	l_p	u_i	b_{pp}	b_{ii}	b_{pi}	$comb1$	$comb2$
G_1	0	182	17 (0)	83 (60)	33 (0)	89 (0)	0.523	0.483
G_2	0	1391	77 (0)	629 (456)	257 (0)	707 (0)	0.531	0.471
G_3	0	4953	154 (0)	2345 (1590)	967 (0)	2412 (0)	0.563	0.493
G_4	0	980	63 (0)	453 (330)	197 (0)	493 (0)	0.539	0.479
G_5	0	1054	116 (0)	479 (346)	212 (0)	524 (0)	0.519	0.478
G_6	0	4659	329 (0)	2084 (1517)	908 (0)	2384 (0)	0.524	0.466
G_7	0	3517	215 (0)	1606 (1168)	657 (0)	1782 (0)	0.531	0.474
G_8	0	266	61 (0)	118 (90)	52 (0)	133 (0)	0.467	0.470
G_9	0	22378	378 (0)	10377 (7289)	4368 (0)	11246 (0)	0.559	0.480
G_{10}	0	3062	155 (0)	1435 (1001)	607 (0)	1472 (0)	0.557	0.494
G_{11}	0	566	105 (0)	264 (182)	111 (0)	274 (0)	0.497	0.491
G_{12}	0	541	48 (0)	250 (178)	108 (0)	269 (0)	0.530	0.482
G_{13}	0	19	9 (0)	7 (5)	4 (0)	9 (0)	0.379	0.438

Table 8.2, presents some statistical information about the ground-truth segment G_1 for different cases. Each line of the first row corresponds to the case of standard BPTs built respectively from an initial partition containing all the pixels, 18 750 regions, 12 500 regions and 6 250 regions. The lines



(a) $l_p = 182$ (b) $l_p = 6$ (c) $l_p = 3$ (d) $l_p = 1$

(e) $l_i = 0$ (f) $l_i = 5$ (g) $l_i = 5$ (h) $l_i = 3$

Figure 8.6: Information about the leaves of the subtrees, corresponding to the ground-truth example G_1, extracted from the BPTs obtained for the different initial partitions. Each column corresponds to information obtained from a subtree built form an initial partition from all the pixels of the image support, from 18 750 regions, from 12 500 regions and from 6 250 regions, respectively. (a–d) In blue, the part of G_1 covered by the set of pure leaves; in gray, the uncovered part. (e–f) In pink, the set of impure leaves; in blue, the set of pure leaves.

in second row (resp. third row) contain information about the subtree extracted from a BPT_{ndvi} (resp. BPT_{ndwi}). We notice that for a same type of BPT built from various initial partitions, the resulting hierarchical representations do not provide similar structures. Indeed, we see that the number l_p of pure leaves and the number l_i of impure leaves are totally different for different initial partitions. In addition, as an example, we notice in Figure 8.6 that the leaves covering on the ground-truth segment G_1 are not similar. The values of $comb1$ for the whole experiment seem to be more sensitive to the values of the granularity and the discordance, while $comb2$ seems to have different behavior. For instance, in the third line of the BPT_{std} case, this score suggests an acceptable quality of the structure, certainly because it almost matches the good values of b_{pp}, b_{ii} and b_{pi}. However, this score is not really reliable since the granularity of the initial partition is too low. Indeed, the leaves are coarse enough to

allow only few of them to form the ground-truth example. In that case, the evaluation is more related to the initial partition than the merging decision that occurred during the construction process.

Table 8.2: Combinatorial scores of the quality of three BPTs (1: BPT_{std}, 2: BPT_{ndvi}, 3: BPT_{ndwi}) built from different initial partitions (1st line: pixels, 2nd line: 18 750, third line: 12 500, fourth line: 6 500), according to the ground-truth example G_1 illustrated in Figure 8.5(a).

	gran	dis	l_i	l_p	u_i	b_{pp}	b_{ii}	b_{pi}	comb1	comb2
1	1.000	0.000	0	182	17 (0)	83 (60)	33 (0)	89 (0)	0.523	0.483
	0.060	0.060	5	6	8 (0)	1 (5)	6 (4)	4 (1)	0.368	0.200
	0.044	0.060	5	3	7 (0)	1 (2)	5 (4)	1 (1)	0.429	0.500
	0.022	0.088	3	1	5 (0)	0 (0)	3 (2)	0 (1)	0.375	0.000
2	1.000	0.000	0	182	83 (0)	69 (63)	36 (0)	98 (0)	0.367	0.413
	0.060	0.060	5	6	81 (0)	1 (5)	8 (4)	1 (1)	0.099	0.500
	0.044	0.060	5	3	67 (0)	0 (2)	6 (4)	1 (1)	0.081	0.000
	0.022	0.088	3	1	57 (0)	0 (0)	3 (2)	0 (1)	0.050	0.000
3	1.000	0.000	0	182	82 (0)	69 (60)	30 (0)	98 (0)	0.355	0.413
	0.060	0.060	5	6	81 (0)	0 (5)	8 (4)	2 (1)	0.088	0.000
	0.044	0.060	5	3	43 (0)	1 (2)	5 (4)	1 (1)	0.120	0.500
	0.022	0.088	3	1	35 (0)	0 (0)	3 (2)	0 (1)	0.079	0.000

From another point of view, the values in Table 8.2 show a huge difference between the various BPTs. Indeed, we can notice that some structures minimize the generation of unary nodes (e.g., BPT_{std}) while the others do the contrary (BPT_{ndvi}, BPT_{ndwi}). This result suggests that the hierarchical structure of the BPT_{std}, here, are likely to provide acceptable segment matching the ground-truth example G_1.

For a first quantitative analysis of the hierarchical structures of the BPTs, Table 8.3 depicts some information allowing us to estimate the quantity of points forming potential false positives ($quant1$), hence estimate the error ($quant2$) that it could induce during the construction of the BPT. The lower these values, the better the ability of the BPT to use at best the potential adequacy of the initial partition to the ground-truth segment. In our case, the BPT_{std} built from an initial partition of 18 750 regions does not seem to show interesting prediction about the ability of the structure to provide satisfactory results for most of the ground-truth examples, because of the large difference between the size of the chosen root of the subtree and the size segments. Indeed, such phenomena brings more possibility of getting more false positive fusions during the process.

Table 8.3: Illustration of the first quantitative analysis of the quality of a BPT_{std} obtained from an initial partition of 18 750 pixels.

G	$\lvert G\rvert$	$\lvert N_G\rvert$	$\lvert N_G \setminus G\rvert$	$\lvert(\bigcup_{L\in\mathcal{L}_G} L) \setminus G\rvert$	$quant1$	$quant2$
G_1	182	259	77	15	62	5
G_2	1391	148314	146923	68	146855	2161
G_3	4953	17234	12281	282	11999	44
G_4	980	118399	117419	79	117340	1486
G_5	1054	249928	248874	80	248794	3111
G_6	4659	249979	245320	271	245049	905
G_7	3517	101223	97706	256	97450	382
G_8	266	249928	249662	3	249659	83221
G_9	22378	75948	53570	605	52965	89
G_{10}	3062	101223	98161	291	97870	337
G_{11}	566	101525	100959	78	100881	1294
G_{12}	541	30894	30353	67	30286	453
G_{13}	19	100	81	25	56	3

The choice of such huge root for the subtree is here due to the fact that the high level of details of the urban image induces the presence of disturbing objects (e.g., shadows, spots on the roof) that make small pieces of regions more similar to other far located objects than to their neighbours. Those kinds of segments tend to not fuse rapidly with their neighbours and often persist until the end of the BPTs construction.

Table 8.4: Quantitative values of the quality of three BPTs (1: BPT_{std}, 2: BPT_{ndvi}, 3: BPT_{ndwi}) built from different amount (1st line: pixels, 2nd line: 18 750, third line: 12 500, fourth line: 6 500) of initial partitions $\mathcal{L}$, according to ground-truth examples G_1 illustrated in Figure 8.5(c).

	$\lvert G\rvert$	$\lvert N_G\rvert$	$\lvert N_G \setminus G\rvert$	$\lvert(\bigcup_{L\in\mathcal{L}_G} L) \setminus G\rvert$	$quant1$	$quant2$
1	182	259	77	0	77	*
	182	259	77	15	62	5
	182	259	77	18	59	4
	182	259	77	25	52	3
2	182	30302	30120	0	30120	*
	182	31082	30900	15	30885	2060
	182	30199	30017	18	29999	1668
	182	30745	30563	25	30538	1223
3	182	20490	20308	0	20308	*
	182	108106	107924	15	107909	7195
	182	97129	96947	18	96929	5386
	182	102602	102420	25	102395	4097

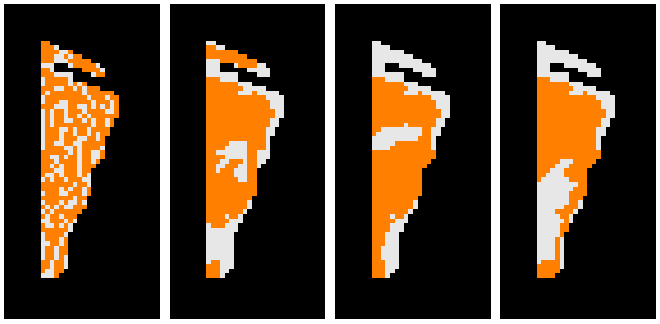
In Table 8.4, we observe more interesting values in the first row, where

BPT_{std} are involved. Indeed, we can see that from the various initial partitions, the size $|N_G|$ of the selected root is not that far from the size $|G|$ of the ground-truth G_1. The values of $quant1$ in this row suggest that the numbers of points representing potential false positives when segmenting are not high compared to the values for BPT_{ndvi} and BPT_{ndwi}. The values of $quant2$ also are not high for the BPT_{std}. This result shows that the use of the metric W_{colour} is more relevant for this study. Indeed, this is probably due to the fact that our crop contains urban objects and the corresponding ground-truth map is generally composed by built areas. Then, the metrics W_{ndvi} and W_{ndwi} do not really fit for the detection of those kinds of objects. However, even if the set of leaves fits perfectly the ground-truth segment G_1 (i.e., $|(\cup_{L\in\mathcal{L}_G} L) \setminus G| = 0$) for the trees built from the all pixels of the image support, the size of the selected root N_G still bring risks of getting more false positives.

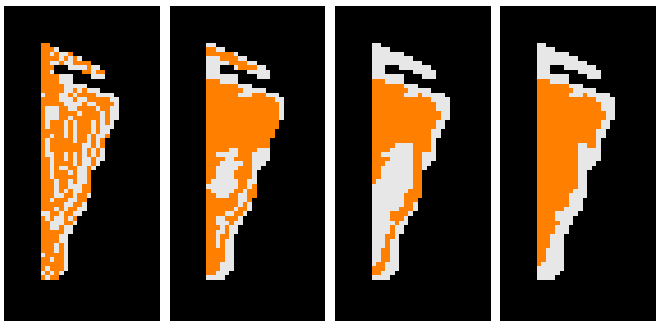
Table 8.5: Illustration of the second quantitative analysis of the quality of a BPT_{std} obtained from an initial partition of 12 500 pixels.

G	$\lvert G\rvert$	n_p	l_p	$\lvert G\rvert - n_p$	$\lvert G\rvert - l_p$	quant3	quant4
G_1	182	25	43	157	139	18	1
G_2	1391	738	1275	653	116	537	38
G_3	4953	2146	4254	2807	699	2108	12
G_4	980	221	432	759	548	211	6
G_5	1054	451	845	603	209	394	12
G_6	4659	2268	3691	2391	968	1423	47
G_7	3517	882	2479	2635	1038	1597	20
G_8	266	166	250	100	16	84	13
G_9	22378	13149	21603	9229	775	8454	92
G_{10}	3062	1593	2606	1469	456	1013	37
G_{11}	566	398	539	168	27	141	16
G_{12}	541	352	453	189	88	101	7
G_{13}	19	0	3	19	16	3	0

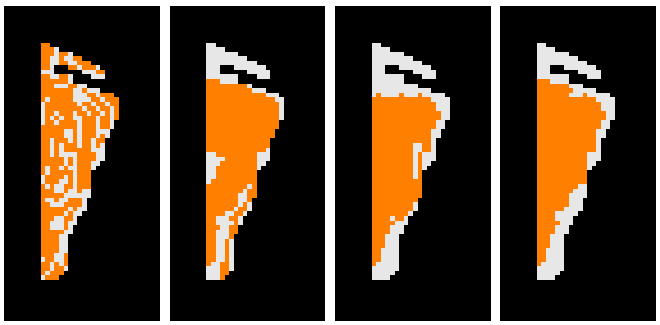
A second quantitative analysis aims to estimate the rate of false negative points that could be encountered during the process. Table 8.5 presents global quantitative information about all the subtrees extracted from the BPT_{std} built from an initial partition of 12 500 regions. The score $quant3$ shows the number of false positive points that could be obtained during the construction process of the tree, while $quant4$ shows the relevance of each subtree by showing the number of nodes that are already well merged in the hierarchical structure. The study of both scores is important since we



(a) 104 nodes (b) 10 nodes (c) 7 nodes (d) 3 nodes



(e) 78 nodes (f) 7 nodes (g) 4 nodes (h) 1 node



(i) 86 nodes (j) 5 nodes (k) 3 nodes (l) 1 node

Figure 8.7: Information about the leaves of the maximum nodes of the subtrees, corresponding to the ground-truth example G_{12}, extracted from the BPTs obtained for the different initial partitions. Each column corresponds to information obtained from a subtree built form an initial partition from all the pixels of the image support, from 18 750 regions, from 12 500 regions and from 6 250 regions. (a–d) Illustration corresponding to the subtree extracted from BPT_{std}. (e–h) Illustration corresponding to the subtree extracted from BPT_{ndvi}. (i–l) Illustration corresponding to the subtree extracted from BPT_{ndwi}.

seek for each ground-truth example a well matching segment while avoiding the risks of over-segmentation.

We can notice that almost all subtrees provide relevant information in comparison to the size of the corresponding ground-truth example. However, the last line shows that the subtree corresponding to the ground-truth example G_{13} has a less relevant hierarchical structure since no nodes are

Table 8.6: Quantitative values of the quality of three BPTs (1: BPT_{std}, 2: BPT_{ndvi}, 3: BPT_{ndwi}) built from different amount (1st line: pixels, 2nd line: 18 750, third line: 12/, 500, fourth line: 6 500) of initial partitions $\mathcal{L}$, according to ground-truth examples G_{12} illustrated in Figure 8.5(c).

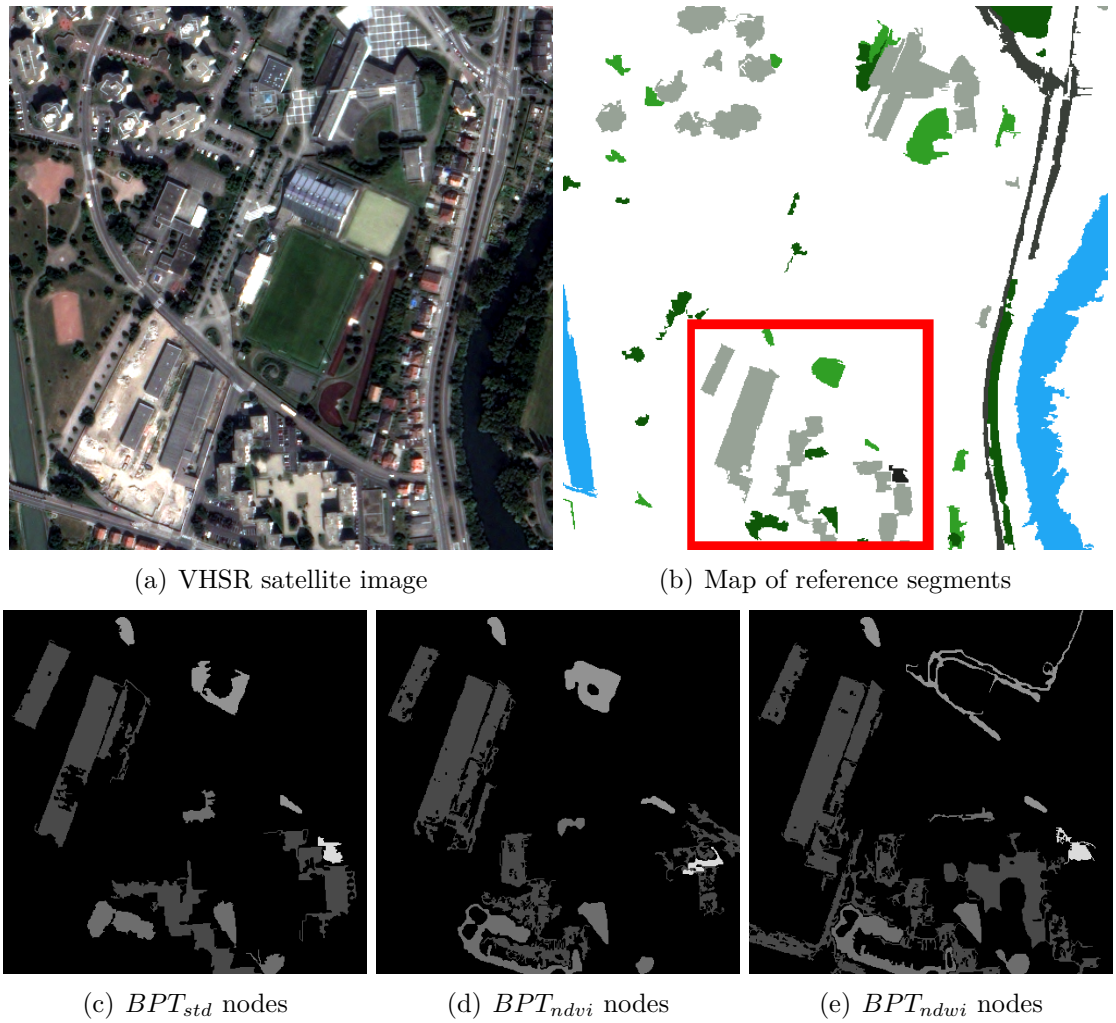
	$\|G\|$	n_p	l_p	$\|G\| - n_p$	$\|G\| - l_p$	$quant3$	$quant4$
1	541	362	541	179	0	179	104
	541	354	490	187	51	136	10
	541	352	453	189	88	101	7
	541	309	446	232	95	137	3
2	541	326	541	215	0	215	78
	541	353	490	188	51	137	7
	541	255	453	286	88	198	4
	541	328	446	213	95	118	1
3	541	347	541	194	0	194	86
	541	379	490	162	51	111	5
	541	338	453	203	88	115	3
	541	348	446	193	95	98	1

well merged. The segment to obtain from such structure is necessarily composed by the set of pure leaves that intersect the ground-truth example.

For the ground-truth example G_{12} (Figure 8.5(d)) the values of $|G| - n_p$ are visually illustrated in Figure 8.7 where the orange region corresponds to the size n_p of $quant4$ sub-regions that are already well-formed in the extracted subtree. We notice both from Table 8.6 and Figure 8.7 that all the subtrees —corresponding to the ground-truth example G_{12}— extracted from the three types of BPTs present interesting hierarchical structures by the generation of matching sub-regions with fair size. The values of quant4 in Table 8.6 suggest that the use of the metrics W_{ndvi} and W_{ndwi} compared to the metric W_{colour} seems to me more relevant for the ground-truth segment G_{12}. Indeed, Figure 8.7(j, k, l) shows that the total merging of all the elements of the set of maximum nodes $\mathcal{N}_p$ could bring a better extraction of the ground-truth example G_{12} from the BPT_{ndwi} than from the other trees.

8.3 Extrinsic quality experiments

In this section, we present some experiments based on the extrinsic quality analysis of a BPT presented in Chapter 7. This extrinsic experiment is twofold. On the one hand, we first assess the quality of the some BPTs



(a) VHSR satellite image (b) Map of reference segments

(c) BPT_{std} nodes (d) BPT_{ndvi} nodes (e) BPT_{ndwi} nodes

Figure 8.8: (a) VHSR image 1 ($1\,000 \times 1\,000$ pixels) at a spatial resolution of 60 cm. (b) Ground-truth map with reference segments belonging to 6 (coloured) semantic classes. (c) Matched nodes from the standard BPT. (d) Matched nodes from the BPT built using the NDVI metric. (e) Matched nodes from the BPT built using the NDWI metric.

and we compare them to each other according to the chosen ground-truth map. On the other hand, we present a quantitative evaluation of the multi-feature BPT construction that we presented in Chapter 2.

Validation of extrinsic quality analysis on classical BPTs

Method From the Image 1 and Image 2 (Figure 8.1(a, c)), three different BPTs —namely BPT_{std}, BPT_{ndvi} and BPT_{ndwi}— were built by choosing three simple construction metrics W_{colour}, W_{ndvi} and W_{ndwi} presented in Section 4.1.

We used our framework to evaluate the extrinsic quality of these three

BPTs relatively to the reference segments of the ground-truth maps. Our framework was parametrized as follows. To compare nodes and ground-truth segments (Section 7.2), we used the uncertain versions of the Jaccard index J' and the Dice coefficient D. The membership function μ was computed by using $\alpha = 1.2$. As quantitative heuristics to reduce the search space of matching nodes (horizontal selection, Section 7.3), we defined κ as a function depending on the size of the ground-truth segment G to be matched, by only evaluating BPT nodes whose size is included in an interval $[0.5 \times |G|, 1.5 \times |G|]$. Finally, the weights involved in the computation of the global quality scores (Section 7.4) were fixed as $w_\ell = 1/|L|$ and $\omega_i = |G_i| / \sum_{G_j \in \mathcal{C}_\ell} |G_j|$ for a ground-truth segment G_i belonging to the semantic class C_ℓ.

Table 8.7: Global quality scores of BPT_{std}, BPT_{ndvi} and BPT_{ndwi} from two VHSR images ($1\,000 \times 1\,000$ pixels). N/S: number of BPT nodes retrieved according to the number of reference segments.

Image	Index	Std	NDVI	NDWI	N/S	Time (s)
1	D	**0.632**	0.511	0.566	46/48	387
	J'	**0.480**	0.369	0.430	46/48	378
2	D	**0.670**	0.523	0.516	51/51	450
	J'	**0.531**	0.389	0.399	51/51	484

Results Table 8.7 presents the global scores obtained, assessing the extrinsic quality of the different BPTs. We notice that the best scores were obtained from the BPT_{std} built upon a classical "colour" metric. Indeed, despite the presence of some vegetation and water segments in the images, the scene is mainly composed of artificial objects such as buildings, leading to lower results for the BPT_{ndvi} and the BPT_{ndwi}. This result seems to be confirmed by Figure 8.8 (c, d, e), showing that the matched nodes for the BPT_{std} are more similar to the ground-truth segments (see red crop, Figure 8.8 (b)) than those from the other BPTs. From Table 8.7, we also remark that two ground-truth segments (out of 48) were not matched for the first image. This result is due to the quantitative heuristics applied during the research that ignored the nodes having their size not included in the restricted interval. If no restrictions were made in the optimization strategy, we could have found them with very low matching rates

(nodes too large or too small) that may impair the global score. Finally, to illustrate the interest of reducing the search space of matching nodes (Section 7.3), we provide the average computation times required to evaluate each BPT. By coupling the two proposed optimization strategies to speed-up the selection of candidate nodes, the required computation times is 387 *seconds* while 28312 *seconds* are needed by a brute force search in the BPT.

Evaluation of the MBPT construction framework vs. standard BPTs

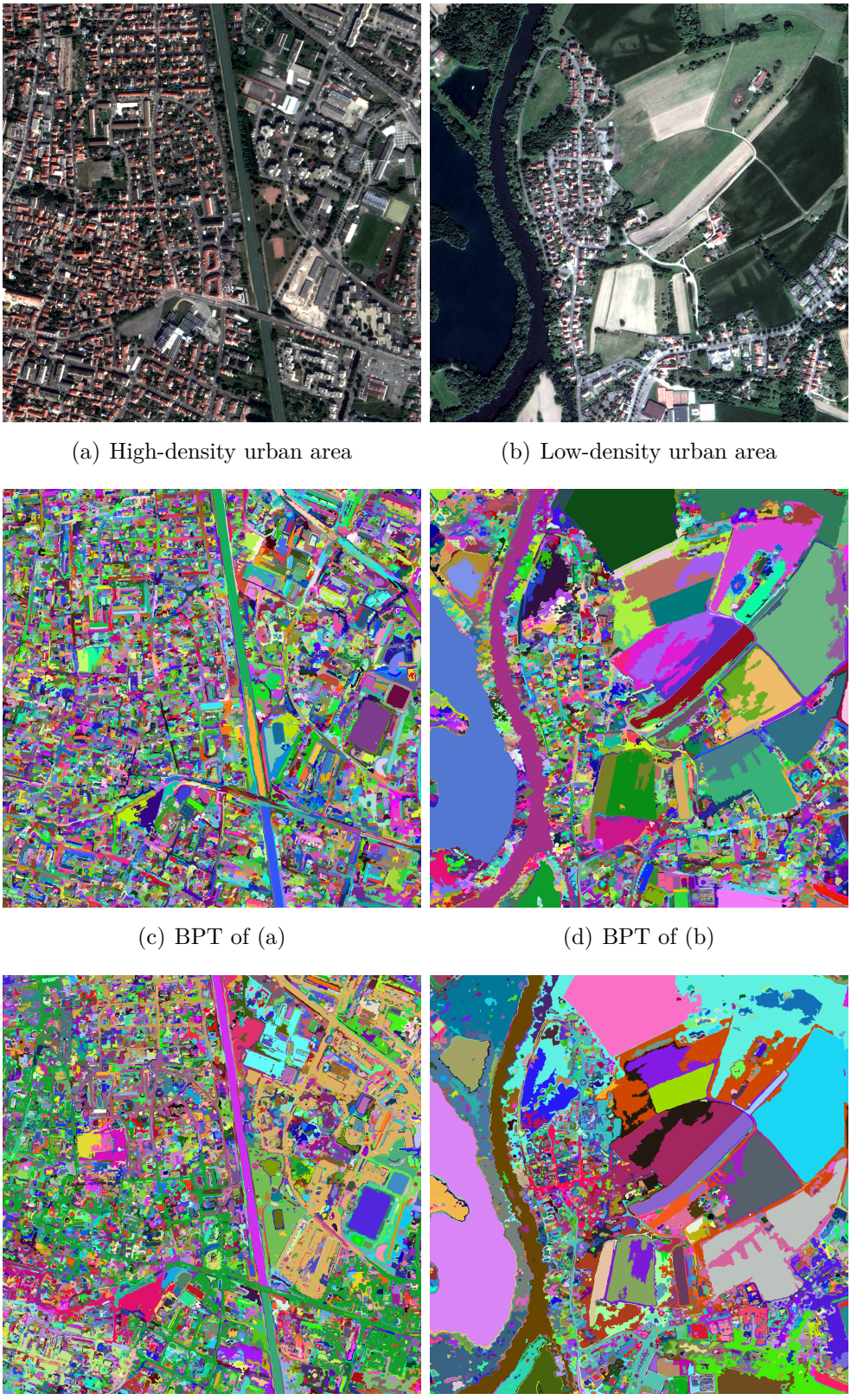
To validate the multi-feature framework for BPT construction that we proposed in Chapter 2, two application cases are considered in the domain of remote sensing image analysis. Our purpose is to highlight the versatility of the multi-feature framework. To this end, we illustrate how it can be used to build BPTs from multiple images and / or multiple metrics computed through the image content. The supervised framework employed to evaluate the quality of MBPTs and allowing us to compare them to the BPTs corresponds to the extrinsic quality analysis presented in Chapter 7. In Section 8.3, we evaluate MBPTs built from a single image, by considering simultaneously various metrics. In Section 8.3, we evaluate MBPTs built from multiple images sensed over the same area.

Evaluation of the multi-criteria BPT construction

We illustrate the interest of MBPTs for multi-feature segmentation, in the context of satellite image analysis.

Method The Image 3 and the Image 4 (Figure 8.2(a, c)) are used for this first application case. To reduce spatial complexity, all (M)BPTs are built from an initial partition $\mathcal{L}$ composed of 200 000 regions obtained from a cut on a standard BPT, thus providing superpixels composed each of 5 pixels, in average.

We consider five valuation functions $W_\star$ modeling either radiometric or geometrical information related to region dissimilarities. The first four of them (W_{colour}, W_{ndvi} , W_{ndwi}, W_{elong}) are as described in Section 4.1.



(a) High-density urban area (b) Low-density urban area

(c) BPT of (a) (d) BPT of (b)

(e) MBPT of (a) (f) MBPT of (b)

Figure 8.9: (a–b) Two VHSR satellite images ($2\,000 \times 2\,000$ pixels) at a spatial resolution of 60 cm sensed by the PLÉIADES satellite and covering different geographical areas. (c–d) Segmentation results from standard BPTs computed from (a) and (b), respectively (23 500 and 5 000 regions, respectively). (e–f) Segmentation results from MBPTs computed from (a) and (b), respectively (23 500 and 5 000 regions, respectively), using 4 features: W_{colour}, W_{elong}, W_{ndvi}, W_{ndwi}.

The fifth valuation function W_{comp} is defined as the change of the geometrical compactness values, potentially induced by the fusion of two incident regions. The geometrical compactness of a node N, noted $comp(N)$, is approximated as the ratio of its border length and the root square of its size. Practically the W_{comp} valuation function is computed for two adjacent nodes N_i, N_j as $|comp(N_i \cup N_j) - \frac{comp(N_i)+comp(N_j)}{2}|$.

As a baseline, standard BPTs are built for each of the two satellite images of Figure 8.9(a–b), by using individually the 5 above valuation functions. Some standard BPTs are also built in a mono-feature way, by considering linear combinations of several valuation functions.

For MBPTs, we consider the relative local information consensus policies *mean-of-ranks* and *most-frequent*, according to the position of the edges within the lists $\mathcal{W}_\star$. They were applied to the first 20% of their contents. To evaluate the impact of the different valuation functions on the segmentation results, the MBPTs were then built by considering various combinations of valuation functions.

Quantitative analysis For quantitative analysis, we use the extrinsic quality evaluation framework presented in Chapter 7. Our purpose is to assess the adequacy of the BPTs and MBPTs relatively to the reference segments of the ground-truth maps. Our framework is parametrized as follows.

To compare nodes and ground-truth segments, we use the Jaccard index J' and the Dice coefficient D (see Equations (7.1–7.2)). As quantitative heuristics for reducing the search space (horizontal selection), we only evaluate (M)BPT nodes whose size lies within $[0.5.|G|, 1.5.|G|]$ for a given ground-truth segment G. The weights involved in the computation of the global quality scores are set as $w_\ell = 1/|L|$ and $\omega_i = |G_i| / \sum_{G_j \in \mathcal{C}_\ell} |G_j|$, for a ground-truth segment G_i belonging to the semantic class C_ℓ.

In order to highlight the efficiency of BPTs (with one metric, or one linear combination of metrics) vs. MBPTs, we first compute the scores obtained for different semantic classes. In particular, Tables 8.8–8.11 provide scores for *built area*, *forest area*, *roads* and *water*, respectively. For the sake of readability, only the most representative –and most efficient– (M)BPTs are provided.

We can observe that for 2 of the 4 tested classes, the MBPTs improve

Table 8.8: Scores of various (M)BPTs for built areas (see Section 8.3). BPT-1: W_{colour}. BPT-2: W_{comp}. BPT-3: W_{ndvi}. BPT-4: linear combination of W_{colour}, W_{comp}. BPT-5: linear combination of W_{colour}, W_{ndvi}. MBPT-1: $\{W_{colour}, W_{comp}\}$ with *most-frequent*. MBPT-2: $\{W_{colour}, W_{ndvi}\}$ with *most-frequent*.

Index	BPT-1	BPT-2	BPT-3	BPT-4	BPT-5	MBPT-1	MBPT-2
Dice (D)	**0.561**	0.432	0.393	0.466	0.376	0.548	0.547
Jaccard (J')	**0.417**	0.283	0.276	0.313	0.266	0.403	0.409

Table 8.9: Scores of various (M)BPTs for forest areas (see Section 8.3). BPT-1: W_{colour}. BPT-2: W_{ndvi}. BPT-3: linear combination of W_{colour}, W_{ndvi}. MBPT-1: $\{W_{colour}, W_{ndvi}\}$ with *mean-of-rank*. MBPT-2: $\{W_{colour}, W_{ndvi}\}$ with *most-frequent*.

Index	BPT-1	BPT-2	BPT-3	MBPT-1	MBPT-2
Dice (D)	0.612	0.684	0.679	0.685	**0.715**
Jaccard (J')	0.464	0.553	0.541	0.548	**0.573**

both Dice and Jaccard indices. In particular, for *forest area* (Table 8.9) and *water* (Table 8.11), the results are better than for BPTs. In the case of *roads* (Table 8.10), one of the MBPTs provides better results than BPTs, while other MBPTs provide worse results. This emphasises the fact that MBPTs do not systematically provide better results than standard BPTs, especially when a given metric is already well fitted to a given class. For instance, in the fourth case, namely for *built area* (Table 8.8), a standard BPT provides better results than the MBPTs, whose scores remain however fairly close. Table 8.12 shows that the scores obtained by BPTs and MBPTs when considering all the ground-truth segments finally provide index values that are slightly better for MBPTs, compared to BPTs. This study tends to prove that well-chosen metrics, involved in the construction of a MBPT, can allow us to slightly improve the quality of the built tree structures for segmentation purpose.

Evaluation of the multi-image BPT construction

We now illustrate the interest of MBPTs for multi-image segmentation, still in the context of satellite image analysis.

Method The Image 5 to the Image 8 (Figure 8.3) are used for this second application case. To reduce spatial complexity, all (M)BPTs are built from an initial partition $\mathcal{L}$ composed of 200 000 regions obtained from a cut on

Table 8.10: Scores of various (M)BPTs for roads (see Section 8.3). BPT-1: W_{colour}. BPT-2: W_{comp}. BPT-3: linear combination of W_{colour}, W_{comp}. MBPT-1: $\{W_{colour}, W_{comp}\}$ with *mean-of-rank*. MBPT-2: $\{W_{colour}, W_{comp}\}$ with *most-frequent*. MBPT-3: $\{W_{colour}, W_{elong}, W_{ndvi}\}$ with *most-frequent*.

Index	BPT-1	BPT-2	BPT-3	MBPT-1	MBPT-2	MBPT-3
Dice (D)	0.633	0.443	0.466	0.509	**0.697**	0.515
Jaccard (J')	0.490	0.295	0.310	0.352	**0.564**	0.357

Table 8.11: Scores of various (M)BPTs for water zones (see Section 8.3). BPT-1: W_{colour}. BPT-2: W_{ndwi}. BPT-3: linear combination of W_{colour}, W_{ndwi}. MBPT-1: $\{W_{colour}, W_{ndwi}\}$ with *most-frequent*. MBPT-2: $\{W_{colour}, W_{elong}, W_{ndvi}\}$ with *most-frequent*.

Index	BPT-1	BPT-2	BPT-3	MBPT-1	MBPT-2
Dice (D)	0.803	0.845	0.906	0.906	**0.940**
Jaccard (J')	0.671	0.732	0.829	0.828	**0.887**

a standard BPT, thus providing superpixels composed each of 5 pixels, in average. This initial partition was produced from Image 1 (Figure 4.10(a)), which is not affected by the presence of clouds.

As a baseline, a standard BPT is built for each satellite image presented in Figure 4.10, with the intensity value criterion W_{colour} (denoted as W_{colour}^{img1}, W_{colour}^{img2}, W_{colour}^{img3}, W_{colour}^{img4} depending on the image where this criterion is computed). These BPTs are denoted as BPT-1, BPT-2, BPT-3 and BPT-4, respectively.

For multi-image BPTs, we consider the relative local information consensus policy *mean-of-ranks*, according to the position of the edges within the lists. It is applied to the first 15% of the lists $\mathcal{W}_\star$. To evaluate the impact of the different valuation functions on the segmentation results, two MBPTs are built. The first, denoted as MBPT-1, considers information extracted from the contents of the four satellite images. In other words, it is built from W_{colour}^{img1}, W_{colour}^{img2}, W_{colour}^{img3}, W_{colour}^{img4}. The second, denoted as MBPT-2, is built from only 3 of the 4 images, but 2 metrics (colour and NDVI). In other words, it is built from W_{colour}^{img1}, W_{colour}^{img2}, W_{colour}^{img3}, W_{ndvi}^{img1}, W_{ndvi}^{img2}, W_{ndvi}^{img3}.

Quantitative analysis For quantitative analysis, we use the extrinsic quality evaluation framework presented in Chapter 7. Our purpose is to assess the adequacy of the BPTs and MBPTs with respect to the reference seg-

Table 8.12: Scores of various (M)BPTs for the five thematic classes (see Section 8.3). BPT-1: W_{colour}. BPT-2: W_{ndvi}. BPT-4: W_{ndwi}. BPT-5: linear combination of W_{colour}, W_{ndwi}. MBPT-1: $\{W_{colour}, W_{ndvi}\}$ with *most-frequent*. MBPT-2: $\{W_{colour}, W_{ndwi}\}$ with *most-frequent*. MBPT-3: $\{W_{colour}, W_{ndvi}, W_{ndwi}\}$ with *most-frequent*.

Index	BPT-1	BPT-2	BPT-4	BPT-5	MBPT-1	MBPT-2	MBPT-3
Dice (D)	0.630	0.559	0.580	0.568	**0.658**	0.656	0.656
Jaccard (J')	0.489	0.439	0.445	0.453	**0.529**	0.525	0.526

ments of the ground-truth map. Our framework is parametrized the same way as in Section 8.3.

Table 8.13 presents the different scores obtained for the different BPTs and MBPTs. The best scores for both Dice and Jaccard indices are obtained by MBPT-2, built from 3 images and 2 metrics (W_{colour}, W_{ndvi}). The other MBPT, built from 4 images, presents lower scores than MBPT-2, which remain however slightly higher than those of standard BPTs for the 4 images.

Table 8.13: Scores assessing the capacity of each (M)BPT to provide nodes similar to the 198 segments of reference of the ground-truth map of Figure 8.4 (see Section 8.3).

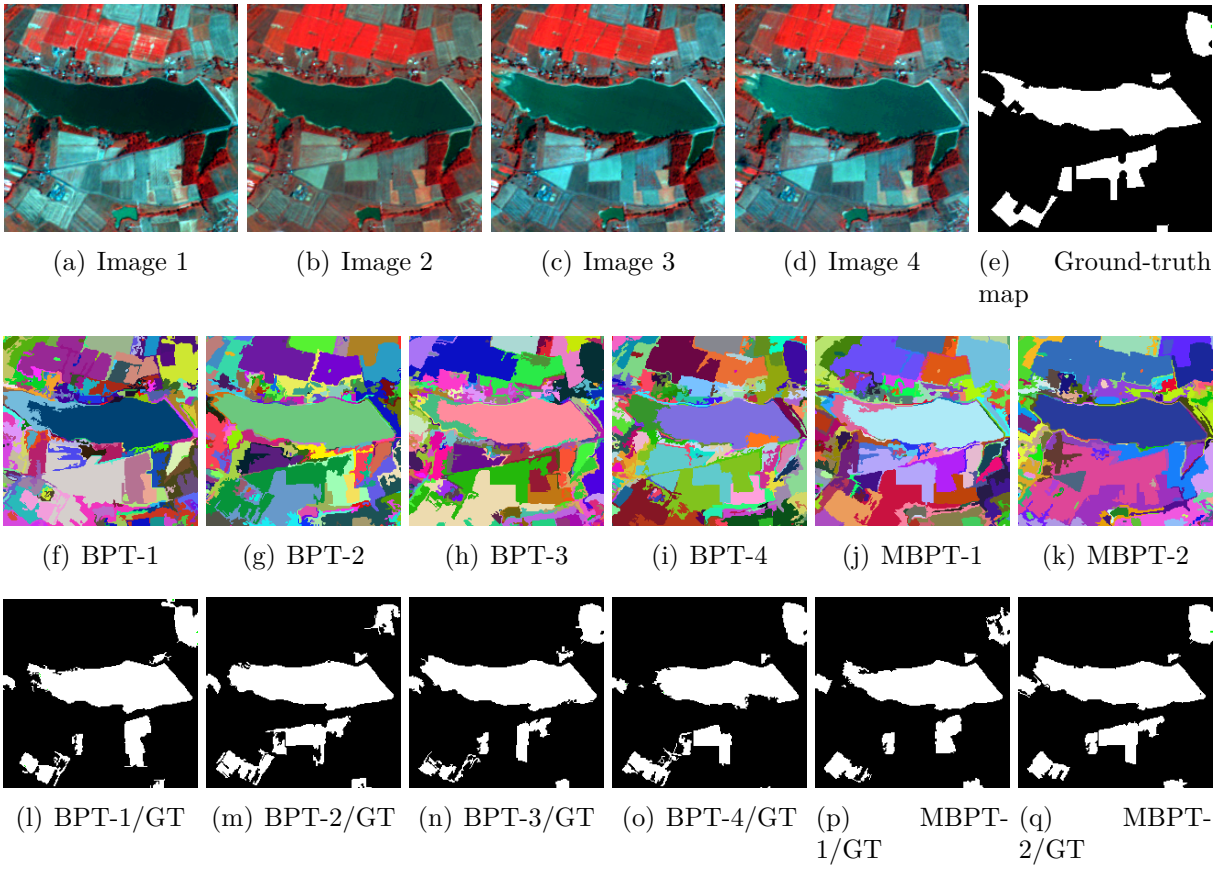
Index	BPT-1	BPT-2	BPT-3	BPT-4	MBPT-1	MBPT-2
Dice (D)	0.651	0.647	0.669	0.674	0.677	**0.700**
Jaccard (J')	0.499	0.500	0.521	0.524	0.532	**0.559**

This quantitative evaluation suggests that MBPTs are able to improve the quality of segmentation results obtained from standard BPTs. The increase of the index values is slight, but indeed sufficient to show that MBPTs constitute a fair structure for gathering and merging information collected from various images of a same scene.

8.4 Discussions

In this section, we discussed about the advantages and the limits of our evaluation frameworks based on intrinsic and extrinsic analyses.

The experiments in Section 8.2 show the relevance of the intrinsic analysis for the evaluation of the ability of a hierarchical structure to provide acceptable segmentation results for any given ground-truth examples. Indeed, the information provided by the subtrees allow one to have initial



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4 (e) Ground-truth map

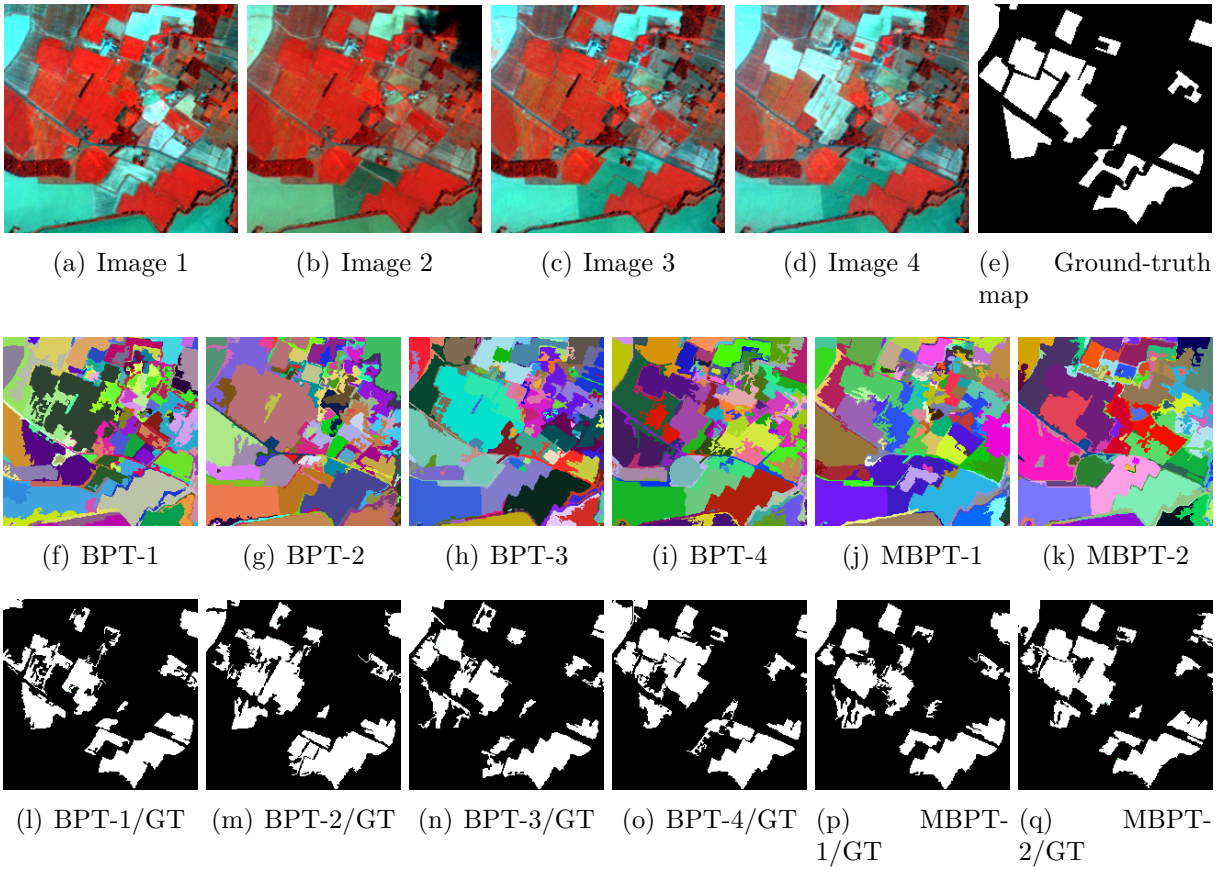
(f) BPT-1 (g) BPT-2 (h) BPT-3 (i) BPT-4 (j) MBPT-1 (k) MBPT-2

(l) BPT-1/GT (m) BPT-2/GT (n) BPT-3/GT (o) BPT-4/GT (p) MBPT-1/GT (q) MBPT-2/GT

Figure 8.10: Segmentation results from standard (i.e., mono-image) BPTs and multi-image/metric MBPTs. (a–d) Crop 1, from Figure 4.10. (e) Crop 1, from the ground-truth map of Figure 8.4. (f–i) Segmentation results from the standard BPTs (BPT-1 to BPT-4), on Images 1 to 4. (j) Segmentation result from MBPT-1 with 4-images/1-metric on Images 1–4. (k) Segmentation results from MBPT-2 with 3-images/2-metrics on Images 1–3. (l–o) Best-matching nodes of the BPTs with respect to the ground-truth map segments. (p–q) Best-matching nodes of the MBPTs with respect to the ground-truth map segments.

intuitions on the potential usefulness of the structure for further segmentation process. Such intrinsic study may —by prediction— help to better understand the potential results that we can obtain from the subtrees.

However, the relevance of the intrinsic study highly depends on the initial partition used for the construction of the structure. Indeed, the interpretation of the information is interesting only once the value of the granularity and the discordance are satisfactory. In addition, some aberrant objects (e.g., shadows derived from some semantic information) may lead to the choice of irrelevant roots for the subtrees. Such phenomenon may be frequent when using remote sensing images where the contents are well detailed and semantic noise deriving from some objects such as the



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4 (e) Ground-truth map

(f) BPT-1 (g) BPT-2 (h) BPT-3 (i) BPT-4 (j) MBPT-1 (k) MBPT-2

(l) BPT-1/GT (m) BPT-2/GT (n) BPT-3/GT (o) BPT-4/GT (p) MBPT-1/GT (q) MBPT-2/GT

Figure 8.11: Segmentation results from standard (i.e., mono-image) BPTs and multi-image/metric MBPTs. (a–d) Crop 2, from Figure 4.10. (e) Crop 2, from the ground-truth map of Figure 8.4. (f–i) Segmentation results from the standard BPTs (BPT-1 to BPT-4), on Images 1 to 4. (j) Segmentation result from MBPT-1 with 4-images/1-metric on Images 1–4. (k) Segmentation results from MBPT-2 with 3-images/2-metrics on Images 1–3. (l–o) Best-matching nodes of the BPTs with respect to the ground-truth map segments. (p–q) Best-matching nodes of the MBPTs with respect to the ground-truth map segments.

shadows and the clouds may be present.

The experiments in Section 8.3 highlight the usefulness of the extrinsic quality analysis for evaluating hierarchical structures such as BPTs but also comparing the quality of our MBPT with the BPT, hence helping in quantitatively validating our generalized multi-feature BPT construction framework. By using extrinsic metrics helping the matching between the inner nodes of the hierarchical structure and the chosen set of segments contained in a given ground-truth map, the extrinsic analysis allows one to assess the ability of the structure to provide relevant segmentation results for the user.

For the extrinsic approach, we proposed a way to find in the hierarchi-

cal structure a single node matching a ground-truth example. However, in practice, such situation is not common. Segments are generally composed by some regions represented by many nodes of the structure. This phenomenon affect the scores and show the limit of this approach.

8.5 Résumé : Études expérimentales

Dans ce chapitre, nous avons présenté différentes expérimentations qui illustrent l'intérêt des approches d'évaluations que nous avons proposées. Premièrement, nous avons présenté les données nécessaires pour ces expérimentations. Ensuite, des expérimentations impliquant des évaluations combinatoires et quantitatives ont mis en évidence l'intérêt de l'analyse intrinsèque de la qualité des BPTs. Puis, nous avons présenté un ensemble de résultats montrant l'utilité de l'analyse extrinsèque pour l'évaluation de la qualité des BPTs, mais aussi pour évaluer l'approche de construction des MBPTs que nous avons proposée.

Données et cartes de vérité-terrain Deux principales régions géographiques sont considérées durant les expérimentations. La première correspond à la ville de Strasbourg, qui représente une zone urbaine, et la deuxième correspond à une zone de culture près de Toulouse. Le premier jeu de données contient des images VHSR de la ville de Strasbourg, captées par le satellite PLÉIADES en 2012. Le premier échantillon contient deux images de $1\,000 \times 1\,000$ pixels et la deuxième contient deux images de $2\,000 \times 2\,000$ pixels. Des cartes de vérité-terrain représentant divers objets urbains de la scène sont également disponibles grâce à une campagne de crowd-sourcing dans le cadre du projet COCLICO. Les segments sont labellisés avec $L = 5$ différentes classes sémantiques : *built area*, *forest area*, *herbaceous area*, *roads*, *water*. Le second jeu de données contient quatre images ($1\,000 \times 1\,000$ pixels) d'une série temporelle captées par le satellite FORMOSAT-2 près de Toulouse durant l'année agricole 2007. Une carte de vérité-terrain des différents champs représentés dans la scène est également disponible où les classes sémantiques associées reflètent le comportement temporel de ces champs durant l'année 2007.

Expérimentations de la qualité intrinsèque Notre but est de donner des intuitions sur l'utilisation de l'analyse intrinsèque pour l'évaluation de la qualité des structures hiérarchiques comme les BPTs. A partir d'une imagette (500×500 pixels) obtenue à partir d'un image de Strasbourg, trois BPTs ont été construits (BPT_{std}, BPT_{ndvi} et BPT_{ndwi}) en considérant les métriques W_{colour}, W_{ndvi} et W_{ndwi}. Pour chaque type de BPT, quatre arbres ont été construits à partir de différentes partitions initiales $\mathcal{L}$ (i.e., contenant, les pixels de l'image support, 6 250 régions, 12 500 régions et 18 750 régions). Nous avons utilisé notre approche intrinsèque pour évaluer la capacité de ces BPTs à générer des nœuds qui correspondent potentiellement à un ensemble de 13 segments de référence. Pour chaque segment de référence, on extrait premièrement un sous-arbre. Les valeurs de l'indice de granularité et de la discordance ont été calculé à partir des différentes partitions initiales afin de déterminer la pertinence de l'analyse intrinsèque. Dans ces expérimentations, nous montrons l'utilité de l'analyse combinatoire et de l'analyse quantitative pour prédire d'une manière intrinsèque la capacité potentielle d'une structure hiérarchique comme les BPTs à fournir un résultat de segmentation satisfaisant relatifs aux segments de références de la vérité-terrain.

Expérimentations de la qualité extrinsèque Notre but est de montrer l'utilité de l'utilisation de l'analyse extrinsèque pour l'évaluation de la qualité des structures hiérarchiques comme les BPTs. Une première expérimentation vise à évaluer la capacité de plusieurs BPTs à fournir des résultats satisfaisants par rapport aux segments de références de la vérité-terrain. A partir des images de $1\,000 \times 1\,000$ pixels du premier jeu de données, trois BPTs ont été construits (BPT_{std}, BPT_{ndvi} et BPT_{ndwi}) en considérant les métriques W_{colour}, W_{ndvi} et W_{ndwi}. Nous avons utilisé sur ces BPTS notre approche extrinsèque en considérant la version incertaine de l'indice de Jaccard J' et du coefficient de Dice D. La fonction μ a été calculée en utilisant $\alpha = 1.2$. Pour la sélection horizontale, nous avons utilisé l'intervalle $[0.5 \times |G|, 1.5 \times |G|]$. Les poids utilisés pour le calcul des scores de qualité globale des BPTs ont été fixés tel que $w_\ell = 1/|L|$ et $\omega_i = |G_i| / \sum_{G_j \in \mathcal{C}_\ell} |G_j|$ pour un segment de référence G_i appartenant à la classe sémantique C_ℓ. Les résultats de ses expériences montrent la

pertinence de notre approche extrinsèque pour évaluer les qualités des différents BPTs mais aussi pour les comparer. En effet, l'analyse a montré que l'utilisation de certaines métriques est plus favorable pour certains segments de référence. Une deuxième expérimentation évalue quantitativement l'approche de construction de BPTs multicritères que nous avons proposée. Pour cela, deux cas d'applications ont été considérés dans le domaine de la télédétection. Notre but est de démontrer la versatilité de notre approche multicritère en comparant la qualité des MBPTs à celle des BPTs. Le premier cas vise à montrer la pertinence de l'utilisation de plusieurs métriques. A partir des images de $2\,000 \times 2\,000$ pixels de Strasbourg, des (M)BPTs ont été créés. Afin de réduire la complexité, les (M)BPTs ont été construit à partir d'une partition initiale comportant 200 000 régions. Pour cette étude, nous considérons cinq métriques (W_{colour}, W_{ndvi} , W_{ndwi}, W_{elong}). Comme base, les BPTs ont été construits individuellement suivant chaque métrique dans un paradigme mono-critère. Les MBPTs ont été construits en utilisant diverses combinaisons des métriques et en utilisant la stratégie de consensus *most-frequent* appliquée sur 20% des têtes des listes $\mathcal{W}_\star$. Le second cas vise à montrer la pertinence de l'utilisation de plusieurs images. Les images utilisées proviennent, cette fois-ci, du second jeu de données. Afin de réduire la complexité spatiale, les (M)BPTs sont construits à partir d'une partition initiale $\mathcal{L}$ comportant 200 000 régions. Comme base, un BPT classique est construit pour chacune des quatre images du jeu de données. Pour les BPT multi-images, le nombre d'images a été varié et la stratégie de consensus *mean-of-ranks* appliquée sur 15% des têtes des listes $\mathcal{W}_\star$. Les résultats montrent la pertinence de l'évaluation extrinsèque pour évaluer la qualité des BPTs, mais aussi d'évaluer la pertinence de notre approche multicritère grâce à des comparaisons entre BPTs et MBPTs.

Discussions D'une part, Les expérimentations montrent la pertinence de l'analyse intrinsèque pour l'évaluation de la capacité des structures hiérarchiques à fournir des résultats de segmentation acceptables pour des segments de référence donnés. En effet, les informations provenant des sous-arbres extraits nous permettent d'avoir des premières intuitions sur l'utilité potentielle de la structure dans un processus de segmentation. Une telle

analyse intrinsèque pourrait —par prédiction— aider à mieux comprendre les résultats potentiels que nous pouvons obtenir des sous-arbres. Cependant, la pertinence de cette étude intrinsèque dépend fortement de la partition initiale utilisée lors de la construction des structures hiérarchiques; En effet, l'interprétation des informations n'est intéressante que si les valeurs de l'indice de granularité et celles de la discordance sont optimales. De plus, la présence de certains objets aberrants (e.g., ombre) peuvent mener à un choix non pertinent de la racine du sous-arbre. Ce phénomène peut être fréquent quand on utilise des images satellites où les contenus sont très détaillés et les bruits sémantiques dérivant des objets —comme les ombres et les nuages— peuvent être présents. D'autre part, les expérimentations montrent aussi l'utilité de l'analyse extrinsèque pour évaluer des structures hiérarchiques comme les MBPTs, et permet aussi de les comparer aux BPTs. Cela nous permet d'évaluer quantitativement notre approche multicritère pour la construction de MBPTs. Il a aussi été montré que l'utilisation d'une métrique extrinsèque aide à la comparaison entre les nœuds de la structure et les segments de référence de la vérité-terrain. Pour l'évaluation extrinsèque, nous avons proposé un moyen de trouver, dans la structure hiérarchique, un nœud correspondant au segment de référence. Cependant, ce n'est pas courant dans la pratique. En effet, les segments sont généralement composés de plusieurs sous-régions correspondant à plusieurs nœuds du BPT. Ce phénomène affecte les scores et montre la limite de notre approche dans son état actuel.

Conclusions et perspectives

Conclusions et perspectives

Conclusions

Dans cette thèse, notre première contribution a été une généralisation du cadre de construction des BPT, construit classiquement de manière mono-critère, permettant alors de considérer divers paradigmes multicritères. Un tel cadre multicritère permet de réduire la tâche de l'utilisateur, en offrant plus de flexibilité lors de la création de BPT. En effet, la négociation entre les différentes métriques, à chaque étape de la construction BPT, sont intrinsèquement traitées par l'algorithme, en fonction des politiques de consensus choisies. Cela réduit le dur travail de définition des connaissances préalables obligatoires à fournir par l'utilisateur au seul choix des métriques impliquées et les stratégies de consensus. C'est aussi une alternative à la construction complexe d'une métrique à partir de plusieurs autres, éventuellement non comparables.

Des expériences utilisant des images de télédétection VHSR mettent en évidence la polyvalence de notre environnement et son intérêt en démontrant comment il peut être utilisé pour construire des BPTs utilisant plusieurs critères d'une manière consensuelle à partir de plusieurs images et / ou de plusieurs métriques calculées à travers le contenu de l'image. Les résultats obtenus suggèrent que notre cadre est capable d'extraire ou de capturer des objets complexes et hétérogènes d'une scène unique. En effet, nous remarquons que divers types d'objets d'intérêt (par exemple des routes homogènes, de vastes étendues de végétation et des zones forestières) peuvent être extraites du contenu global par des MBPT, éventuellement à différentes échelles. De plus, les résultats font apparaître la capacité de notre généralisation à gérer l'information provenant de diverses sources (c'est-à-dire, les métriques, les images) et en tirer parti.

Notre deuxième contribution est un cadre d'évaluation de la qualité des

BPT de manière supervisée. Le besoin de validation quantitative de notre BPT multicritère nous a motivé à proposer deux approches d'évaluation. D'une part, des analyses d'évaluation intrinsèque s'appuyant sur la structure interne du BPT ont montré qu'il est possible d'estimer la qualité potentielle d'un BPT en fonction des nœuds fusionnés et de sa hiérarchie interne. Une telle analyse donne une estimation de l'aspect correct de la construction du BPT et reflète sa capacité à fournir (ou non) des résultats de segmentation pertinents pour l'expert. D'autre part, une analyse d'évaluation extrinsèque permet à l'utilisateur de définir une métrique de la qualité aidant la recherche d'un nœud du BPT qui corresponde bien à n'importe quel segment de vérité-terrain sélectionné ou fourni par l'utilisateur. Un score de qualité global d'un BPT est calculé en prenant en tenant compte de l'information incertaine / sémantique d'une carte de vérité-terrain donnée. Les expériences mettent en évidence l'utilité de nos cadres dans le contexte de l'analyse d'image de télédétection. La stratégie d'optimisation proposée pour réduire l'espace de recherche des nœuds correspondants du BPT est utile sur le plan informatique mais peut biaiser légèrement les résultats et devrait être utilisée avec précaution.

Enfin, nous avons proposé une validation quantitative de notre environnement de construction de BPT multicritère en comparant les MBPT avec divers BPT construits à partir de métriques uniques. Les résultats ont montré que les scores de qualité des MBPT, fournis par nos expériences, sont souvent plus élevés que ceux obtenus à partir des BPT standard. Cependant, cela ne signifie pas que les MBPT sont systématiquement plus pertinents / utiles que les BPT. Effectivement, ces résultats dépendent fortement de la combinaison de métriques utilisées pour tout objet d'intérêt spécifique. Nos expériences montrent que les MBPT peuvent être pertinents pour une application définie à la fois en fonction du choix des paramètres et du choix de la stratégie consensuelle utilisée.

Perspectives

La pierre angulaire du cadre de construction de BPT multicritère est la stratégie utilisée lors de la mise en place du consensus entre diverses sources d'information. Celui-ci définit toute décision de fusion au cours de l'algo-

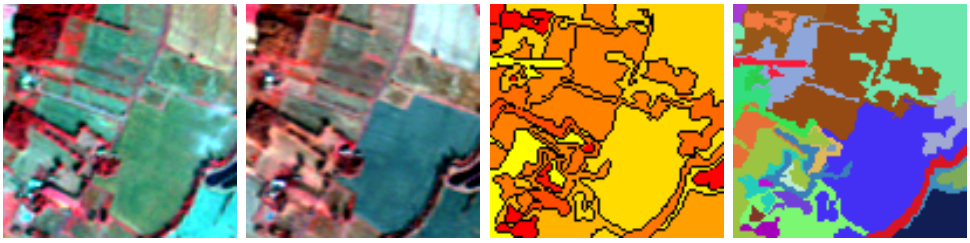
rithme de construction du MBPT. Une première perspective directe de cette thèse serait d'intégrer un consensus de haut niveau, ce qui pourrait aussi nous permettre d'améliorer la pertinence des hiérarchies et la segmentation induite.

Les évolutions algorithmiques liées à la construction de BPT multicritères nécessitent la manipulation de structures de données plus complexes par rapport aux « BPT standard ». Pour s'attaquer à la mémoire induite et aux problèmes de complexité temporelle soulevés par ce cadre, des approfondissements et des améliorations sont à apporter aux implémentations relatives aussi bien aux heuristiques séquentielles qu'à celles qui sont distribuée et ce en se basant sur la distribution d'algorithmes de graphes. Une telle amélioration pourrait aider à mettre à l'échelle notre algorithme pour gérer des images de très grandes dimensions.

Au-delà des exemples d'application décrits dans cette thèse, d'autres applications pertinentes pourraient également être envisagées pour le traitement de plusieurs familles d'images. Par exemple, il est possible d'appliquer des BPT multicritères pour segmenter des images hyperspectrales, en établissant un consensus entre l'information complémentaire (et potentiellement corrélée) portée par les différentes bandes spectrales ; cela nécessiterait d'abord de prétraiter les bandes les plus pertinentes impliquées dans le processus, pour éviter les problèmes liés à la complexité espace / temps. L'imagerie multi-temporelle pourrait également être envisagée, en établissant un consensus temporel de niveau supérieur entre les différentes acquisitions d'images de la même scène.

Un autre défi méthodologique, éventuellement induit par des séries temporelles, est soulevé par la possible divergence entre les valeurs différentes recueillies par les métriques / images, qui peuvent conduire à des décisions consensuelles occasionnelles non pertinentes. Nous prévoyons d'étudier comment les informations non-consensuels (voir la Figure 8.12) pourraient être utilisées pour suivre un consensus local entre les mesures / images, conduisant à des hyper-arbres où les branches modélisent les décisions locales de fusion.

Dans l'analyse extrinsèque de la qualité, nous avons présenté un moyen de trouver dans une structure hiérarchique un nœud correspondant bien à un exemple donné de vérité-terrain. Le cas idéal est de trouver un nœud



(a) Image 1 (b) Image 2 (c) Conflictual map (d) MBPT result

FIGURE 8.12 – Example highlighting a conflictual map showing non-consensual merging decisions during the construction of a MBPT from two images. (a) First image. (b) Second image. (c) Heatmap showing conflictual merging decision for each segment (yellow : low / non-conflictual, orange : medium-conflictual, red : high-conflictual). (d) Partition obtained from a cut applied on the MBPT.

correspondant parfaitement à l'attente de l'utilisateur. Cependant, ce cas n'est pas courant en pratique. En effet, un segment de vérité-terrain est généralement composé de nombreux autres pièces (cas de sur-segmentation) ou totalement inclus dans un grand segment (cas de sous-pondération). Une perspective consisterait à étendre l'approche proposée de la structure hiérarchique en un ensemble de nœuds correspondant bien à un seul exemple de vérité du sol. Une double approche d'optimisation est alors nécessaire, car un bon appariement devrait maximiser le chevauchement entre l'exemple de vérité-terrain et l'ensemble des nœuds tout en minimisant la cardinalité de cet ensemble.

Il y a plusieurs façons de tirer parti du cadre d'évaluation proposé. Effectivement, il peut être utilisé à la fois de manière rétrospective ou prospective. Dans le premier —rétrospective— cas, il peut permettre à l'utilisateur expert de l'aider à évaluer ses choix en termes de métriques de construction BPT. En particulier, il / elle peut choisir a posteriori quelle métrique, parmi plusieurs proposées, est la meilleure pour une application de segmentation donnée. Dans le deuxième —prospective— cas, il peut permettre de déterminer automatiquement (en utilisant une étape d'apprentissage) un ensemble de métriques de construction pertinentes dans un ensemble peut-être énorme. Ce serait particulièrement pertinent pour les applications qui construisent des MBPT à partir de nombreuses images et / ou de nombreux paramètres de construction, basés sur des stratégies de consensus. En effet, la qualité d'un MBPT dépend des métriques à utiliser et de la stratégie de consensus en fonction des attentes de l'expert dans

son travail d'analyse. Nous prévoyons utiliser notre cadre d'évaluation pour aider et guider les choix de paramètres liés à la construction des MBPTs, conduisant à l'amélioration des résultats de segmentation d'image.

Conclusions and perspectives

Conclusions and perspectives

Conclusions

In this thesis, our first contribution was a generalization of the BPT construction framework, classically built in a mono-feature way, thus allowing one to consider various multi-feature paradigms. Such a multi-feature framework enables to reduce the task of the user, by offering more flexibility for the BPT creation. Indeed, the negociation between the different features, at each step of the BPT construction, is intrinsically dealt with by the algorithm, with respect to the chosen consensus policies. This reduces the hard prior knowledge mandatory from the user to the only choice of the involved features and the global strategies for their consensus. It is also an alternative to the complex metric construction issue from several, possibly non-comparable, features.

Experiments using VHSR remote sensing images highlight the versatility of our framework and its interest by demonstrating how it can be used to build consensual multi-feature BPTs from multiple images and / or multiple metrics computed through the image content. The obtained results suggest that our framework is able to extract or capture complex and heterogeneous objects of a unique scene. Indeed, we notice that various kinds of objects of interest (e.g., homogeneous roads, vast vegetation and forest areas) can be extracted from the content of the MBPTs possibly at different scales. In addition, the results reflect the ability of our framework to manage information from various sources (i.e., features, images) and take advantage from them.

Our second contribution was a framework for evaluating the quality of the BPTs in a supervised way. The need for quantitative validation of our multi-feature BPT framework motivated us to propose two evaluation approaches. On the one hand, an intrinsic evaluation analyses relying on

the internal structure of the BPT showed that it is possible to estimate the potential quality of a BPT according to merged nodes and the hierarchy within it. Such analysis gives an estimation on the correctness construction of the BPT and reflects its ability to provide (or not) relevant segmentation results for the expert.

On the other hand, an extrinsic evaluation analysis allows the user to define a quality metric helping the search of a node of the BPT that matches well any given ground-truth segment example chosen by the user. A global quality score of a BPT is computed taking into consideration the uncertainty / semantic information of a given ground-truth map. The experiments highlight the usefulness of our frameworks in the context of remote sensing image analysis. The optimization strategy proposed to reduce the search space of matching BPT nodes is computationally helpful but may slightly bias the results and should be used carefully.

Finally, we proposed a quantitative validation of our multi-feature BPT construction framework by comparing the MBPTs with various BPTs built from single different metrics. The results showed that the quality scores of the MBPTs, provided by our experiments, are often higher than those obtained from standard BPTs. However, it does not mean that the MBPTs are systematically more relevant / useful than BPTs. Indeed, those results depend highly on the combination of metrics used for any specific object of interest. Our experiments show that the MBPTs could be relevant for a defined application both according to the choice of metrics and also the choice of consensus strategy to be used.

Perspectives

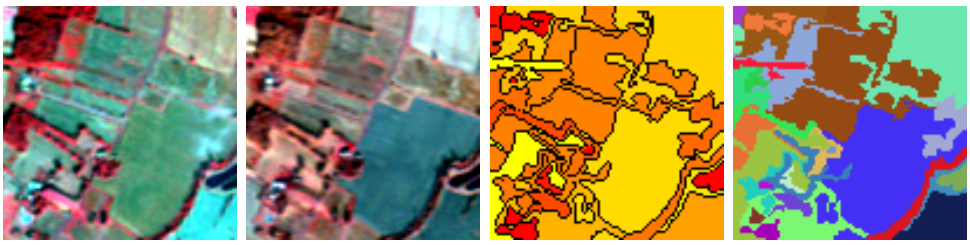
The cornerstone of the multi-feature BPT construction framework is the strategy used to set up a consensus between various sources of information. It defines all merging decisions during the algorithm of construction of the MBPT. A first direct perspective of this thesis would be to integrate higher-level consensus, which may also allow us to improve the relevance of the hierarchies and the induced segmentation.

The algorithmic evolutions related to the multi-feature BPT construction require the handling of more complex data-structures, compared to

"standard" BPTs. In order to tackle the induced memory and time complexity issues raised by this framework, a deeper investigation and implementation on the sequential and distributed heuristics based on graph-based distribution algorithmic could be interesting. Such improvement could help scaling up our algorithm to handle images of huge dimensions.

Beyond the application examples described in this thesis, other relevant applications could also be considered for the processing of different families of images. As an example, it may be possible to apply multi-feature BPTs to segment hyperspectral images, by establishing a consensus between the complementary (and potentially correlated) information carried by the different spectral bands ; this would require to first preprocess the most relevant bands involved in the process, for avoiding space / time complexity issues. Multi-temporal imaging could also be considered, by establishing a higher-level "temporal" consensus between the different image acquisitions of the same scene.

Another methodological challenge, possibly induced by time series, is raised by the possible divergences between the different values gathered by the metrics / images, which may lead to occasional irrelevant consensual decisions. We plan to study how non-consensual information (see Figure 8.13) could be used to follow local consensus between metrics / images, leading to hypertrees where the branches model local fusion decisions.



(a) Image 1 (b) Image 2 (c) Conflictual map (d) MBPT result

FIGURE 8.13 – Example highlighting a conflictual map showing non-consensual merging decisions during the construction of a MBPT from two images. (a) First image. (b) Second image. (c) Heatmap showing conflictual merging decision for each segment (yellow : low / non-conflictual, orange : medium-conflictual, red : high-conflictual). (d) Partition obtained from a cut applied on the MBPT.

In the extrinsic quality analysis, we presented a way to find in a hierarchical structure a single node well matching a given ground-truth example. The ideal case is to find a node perfectly matching the expectancy of the

user. However, this case is not common in practice. Indeed, a ground-truth segment is generally either composed by many other pieces (case of over-segmentation) or totally included in a large segment (case of under-segmentation). A perspective would be to extend the proposed approach by finding in the hierarchical structure a set of nodes well matching a single ground-truth example. A double optimization approach is then required, since a good matching should maximize the overlapping between the ground-truth example and the set of nodes while minimizing the cardinality of this set.

There are many ways to take advantage of the proposed evaluation framework. Indeed, it can be used both in a retrospective or a prospective way. In the first —retrospective— case, it can allow the expert user to help him / her to assess his / her choices in terms of BPT construction metrics. In particular, he / she can choose a posteriori which metric, among several proposed ones, is the best for a given segmentation application. In the second —prospective— case, it can allow for automatically determining (using a learning step) a set of relevant construction metrics within a possibly huge set. This would be particularly relevant for applications building MBPTs from many images and / or many construction metrics, based on consensus strategies. Indeed, the quality of a MBPT depends on the metrics to use and the consensus strategy according to the expectations of the expert in his / her analysis. We plan to use our evaluation framework to help and guide the parameter choices related to the construction of MBPTs, leading to improvement of image segmentation results.

Bibliographie

[1] A. Rosenfeld, "Connectivity in digital pictures," *Journal of the ACM*, vol. 17, no. 1, pp. 146–160, 1970.

[2] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

[3] L. Vincent and P. Soille, "Watersheds in digital spaces : An efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.

[4] S. L. Horowitz and T. Pavlidis, "Picture segmentation by a directed split-and-merge procedure," in *Second International Joint Conference on Pattern Recognition, Proceedings*, vol. 424, p. 433, 1974.

[5] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*, vol. 1, pp. 3–20, 1991.

[6] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[7] P. Salembier and S. Foucher, "Optimum graph cuts for pruning binary partition trees of polarimetric sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5493–5502, 2016.

[8] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[9] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed : A unifying graph-based optimization framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1384–1398, 2011.

[10] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1153–1160, 1995.

[11] P. Salembier and M. Wilkinson, "Connected operators : A review of region-based morphological image processing techniques," *Ieee signal processing magazine*, vol. 26, pp. 136–157, 2009.

[12] E. Carlinet and T. Géraud, "A comparative review of component tree computation algorithms," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3885–3895, 2014.

[13] M.-M. Yau and S. N. Srihari, "A hierarchical data structure for multidimensional digital images," *Communications of the ACM*, vol. 26, no. 7, pp. 504–515, 1983.

[14] A. Montanvert, P. Meer, and A. Rosenfeld, "Hierarchical image analysis using irregular tessellations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 307–316, 1991.

[15] P. Salembier, A. Oliveras, and L. Garrido, "Anti-extensive connected operators for image and sequence processing," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570, 1998.

[16] P. Monasse and F. Guichard, "Scale-space from a level lines tree," *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 224–236, 2000.

[17] P. Monasse and F. Guichard, "Fast computation of a contrast-invariant image representation," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 860–872, 2000.

[18] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1163–1173, 1996.

[19] B. Perret, S. Lefèvre, C. Collet, and E. Slezak, "Hyperconnections and hierarchical representations for grayscale and multiband image processing," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 14–27, 2012.

[20] G. Ouzounis and P. Soille, *The Alpha-Tree Algorithm. Theory, Algorithms, and Applications.* 2012.

[21] J. Havel, F. Merciol, and S. Lefèvre, "Efficient tree construction for multiscale image representation and processing," *Journal of Real-Time Image Processing*, 2016.

[22] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 561–576, 2000.

[23] R. Jones, "Connected filtering and segmentation using component trees," *Computer Vision and Image Understanding*, vol. 75, no. 3, pp. 215–228, 1999.

[24] P. Bosilj, B. B. Damodaran, E. Aptoula, M. D. Mura, and S. Lefèvre, "Attribute profiles from partitioning trees," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, pp. 381–392, Springer, 2017.

[25] E. Aptoula, M. D. Mura, and S. Lefèvre, "Vector attribute profiles for hyperspectral image classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3208–3220, 2016.

[26] L. Guigues, J.-P. Cocquerez, and H. Le Men, "Scale-sets image analysis," *International Journal of Computer Vision*, vol. 68, no. 3, pp. 289–317, 2006.

[27] N. Passat, B. Naegel, F. Rousseau, M. Koob, and J.-L. Dietemann, "Interactive segmentation based on component-trees," *Pattern Recognition*, vol. 44, no. 10–11, pp. 2539–2554, 2011.

[28] B. R. Kiran, *Energetic-lattice based optimization. (L'optimization par trellis-énergetique).* PhD thesis, University of Paris-Est, France, 2014.

[29] Y. Cui, L. Chapel, and S. Lefèvre, "Scalable bag of subpaths kernel for learning on hierarchical image representations and multi-source remote sensing data classification," *Remote Sensing*, vol. 9, no. 3, 2017.

[30] N. Passat and B. Naegel, "Component-trees and multivalued images : Structural properties," *Journal of Mathematical Imaging and Vision*, vol. 49, no. 1, pp. 37–50, 2014.

[31] C. Kurtz, B. Naegel, and N. Passat, "Connected filtering based on multivalued component-trees," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5152–5164, 2014.

[32] E. Carlinet and T. Géraud, "MToS : A tree of shapes for multivariate images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5330–5342, 2015.

[33] N. Passat and B. Naegel, "Component-hypertrees for image segmentation," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, vol. 6671 of *Lecture Notes in Computer Science*, pp. 284–295, Springer, 2011.

[34] B. Perret, J. Cousty, O. Tankyevych, H. Talbot, and N. Passat, "Directed connected operators : Asymmetric hierarchies for image filtering and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1162–1176, 2015.

[35] Y. Xu, T. Géraud, and L. Najman, "Connected filtering on tree-based shape-spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 6, pp. 1126–1140, 2016.

[36] É. Grossiord, B. Naegel, H. Talbot, N. Passat, and L. Najman, "Shape-based analysis on component-graphs for multivalue image processing," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, vol. 9082 of *Lecture Notes in Computer Science*, pp. 446–457, Springer, 2015.

[37] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[38] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1132–1145, 2008.

[39] S. Valero, P. Salembier, and J. Chanussot, "Comparison of merging orders and pruning strategies for binary partition tree in hyperspectral data," in *ICIP, International Conference on Image Processing, Proceedings*, pp. 2565–2568, 2010.

[40] X. Giro and F. Marqués, "From partition trees to semantic trees," in *MRCS, Multimedia Content Representation, Classification and Security, Proceedings*, vol. 4105 of *Lecture Notes in Computer Science*, pp. 306–313, Springer, 2006.

[41] V. Vilaplana, F. Marques, and P. Salembier, "Binary partition trees for object detection," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2201–2216, 2008.

[42] S. Valero, P. Salembier, and J. Chanussot, "Object recognition in hyperspectral images using binary partition tree representation," *Pattern Recognition Letters*, vol. 56, pp. 45–51, 2015.

[43] J. A. Benediktsson, L. Bruzzone, J. Chanussot, M. Dalla Mura, P. Salembier, and S. Valero, "Hierarchical analysis of remote sensing data : Morphological attribute profiles and binary partition trees," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, Lecture Notes in Computer Science, pp. 306–319, Springer, 2011.

[44] C. Kurtz, N. Passat, P. Gançarski, and A. Puissant, "Extraction of complex patterns from multiresolution remote sensing images : A hierarchical top-down methodology," *Pattern Recognition*, vol. 45, no. 2, pp. 685–706, 2012.

[45] F. Calderero, F. Eugenio, J. Marcello, and F. Marques, "Multispectral cooperative partition sequence fusion for joint classification and hierarchical segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1012–1016, 2012.

[46] C. Kurtz, A. Stumpf, J.-P. Malet, P. Gançarski, A. Puissant, and N. Passat, "Hierarchical extraction of landslides from multiresolution

remotely sensed optical images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 122–136, 2014.

[47] S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image representation and processing with binary partition trees," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1430–1443, 2013.

[48] M. A. Veganzones, G. Tochon, M. Dalla Mura, A. J. Plaza, and J. Chanussot, "Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3574–3589, 2014.

[49] A. Alonso-González, C. López-Martínez, and P. Salembier, "Filtering and segmentation of polarimetric SAR data based on binary partition trees," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 593–605, 2012.

[50] P. Salembier, "Study of binary partition tree pruning techniques for polarimetric SAR images," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, vol. 9082 of *Lecture Notes in Computer Science*, pp. 51–62, Springer, 2015.

[51] A. Alonso-González, S. Valero, J. Chanussot, C. López-Martínez, and P. Salembier, "Processing multidimensional SAR and hyperspectral images with binary partition tree," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 723–747, 2013.

[52] A. Alonso-González, C. López-Martínez, and P. Salembier, "PolSAR time series processing with binary partition trees," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3553–3567, 2014.

[53] C. Kurtz, N. Passat, A. Puissant, and P. Gançarski, "Hierarchical segmentation of multiresolution remote sensing images," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, vol. 6671 of *Lecture Notes in Computer Science*, pp. 343–354, Springer, 2011.

[54] L. Garrido, P. Salembier, and D. Garcia, "Extensive operators in partition lattices for image sequence analysis," *Signal Processing*, vol. 66, no. 2, pp. 157–180, 1998.

[55] J. Cousty, L. Najman, and B. Perret, "Constructive links between some morphological hierarchies on edge-weighted graphs," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, vol. 7883 of *Lecture Notes in Computer Science*, pp. 86–97, Springer, 2013.

[56] V. Machairas, M. Faessel, D. Cárdenas-Peña, T. Chabardes, T. Walter, and E. Decencière, "Waterpixels," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3707–3716, 2015.

[57] A. Al-Dujaili, F. Merciol, and S. Lefèvre, "GraphBPT : An efficient hierarchical data structure for image representation and probabilistic inference," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, vol. 9082 of *Lecture Notes in Computer Science*, pp. 301–312, Springer, 2015.

[58] M. Wilkinson, P. Soille, M. Pesaresi, and G. K. Ouzounis, "Concurrent computation of differential morphological profiles on giga-pixel images," in *ISMM, International Symposium on Mathematical Morphology, Proceedings*, pp. 331–342, Springer, 2011.

[59] U. Moschini, A. Meijster, and M. Wilkinson, "A hybrid shared-memory parallel max-tree algorithm for extreme dynamic-range images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 513–526, 2018.

[60] M. H. Wilkinson, M. Pesaresi, and G. K. Ouzounis, "An efficient parallel algorithm for multi-scale analysis of connected components in gigapixel images," *ISPRS International Journal of Geo-Information*, vol. 5, no. 3, 2016.

[61] J. Pont-Tuset and F. Marques, "Supervised assessment of segmentation hierarchies," in *ECCV, European Conference on Computer Vision, Proceedings*, vol. 7575 of *Lecture Notes in Computer Science*, pp. 814–827, 2012.

[62] P. Jaccard, "The distribution of the flora in the Alpine zone," *New Phytologist*, vol. 11, pp. 37–50, 1912.

[63] L. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297–302, 1945.

[64] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons," *Biologiske Skrifter*, vol. 5, pp. 1–34, 1948.

[65] M. Polak, H. Zhang, and M. Pi, "An evaluation metric for image segmentation of multiple objects," *Image and Vision Computing*, vol. 27, pp. 1223–1227, 2009.

[66] H. Vojodi and A. Moghadam, "A supervised evaluation method based on region shape descriptor for image segmentation algorithm," in *AISP, International Symposium on Artificial Intelligence and Signal Processing, Proceedings*, pp. 18–22, 2012.

[67] J. Pont-Tuset and F. Marques, "Measures and meta-measures for the supervised evaluation of image segmentation," in *CVPR, Computer Vision and Pattern Recognition, Proceedings*, pp. 2131–2138, 2013.

[68] H. Li, J. Cai, T. Nguyen, and J. Zheng, "A benchmark for semantic image segmentation," in *ICME, International Conference on Multimedia and Expo, Proceedings*, pp. 1–6, 2013.

[69] J. Pont-Tuset, P. Arbelaez, J. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping for image segmentation and object proposal generation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 128–140, 2017.

[70] J. Pont-Tuset and F. Marques, "Supervised evaluation of image segmentation and object proposal techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 1465–1478, 2016.

[71] R. Fabbri, L. Costa, J. Torelli, and O. Bruno, "2D Euclidean distance transform algorithms : A comparative survey," *ACM Computing Surveys*, vol. 40, pp. 1–44, 2008.

[72] J. Serra, "Hierarchies and optima," in *Discrete Geometry for Computer Imagery (DGCI)*, vol. 6607 of *Lecture Notes in Computer Science*, pp. 35–46, 2011.

Appendix

A Agat : An open-source library for multi-feature BPT construction

We provide, in this Appendix, a short discussion about an implementation as a library of the proposed MBPT construction strategy. This open-source JAVA library, namely AGAT, is available at `https://bitbucket.org/agat-team/agat-v0.3`. It implements the proposed algorithm (Section 2.3) and the required data-structures (B) for the creation of a MBPT. It should be noticed that other open-source libraries were already proposed to create BPTs [52, 57]. However they rely on the creation of classical mono-feature BPTs.

AGAT is a generic library. The user can tune the construction of his/her BPT, by easily implementing his/her own metric and consensus strategies as new JAVA classes. We use the same pattern factory mechanism for both cases. Then, the user only need to create new classes that extend either a predefined metric or a consensus class parent. To create a new metric, two methods have to be designed. The first defines the computation of the underlying feature value for a particular node. The second defines the metric (dissimilarity) computation between two adjacent nodes according to this feature. To create a new consensus, the user only has to implement a method that applies the desired consensus strategy and returns the next couple of nodes to be merged.

AGAT also integrates a TIFF library for loading only the subdivisions of the images that are necessary to the current segmentations. It is then possible to reduce the memory resources required by the application. This is useful when dealing with large images.

Once a BPT has been constructed, a saving system enables to store the tree structure in a file. Such file, encoded in the JSON format, contains

information about the tree, e.g., the pixels contained in each leaf, the merging order of the nodes, the paths of the images in the file system, the metrics and the consensus strategy. The BPT can be rebuilt without recomputation, by simply reading this JSON file.

For image segmentation purpose, a standard cut can be performed from the BPT, by specifying the number of desired regions. As output, standard segmentation images (border image or label image based on random colours) are generated. Other segmentation paradigms proposed in the literature, mainly based on energy optimization, or criterion-based node selection can be easily interfaced with AGAT.

B Implementation details of the data-structures

This Appendix describes the main data-structures involved in AGAT. For the sake of concision and readability, only the most important structures (MBPT, RAG, nodes, adjacencies, and lists $\mathcal{W}_i$) are mentioned here. Some of these data-structures are strongly dependent on the JAVA philosophy.

B.1 Data-structure of the MBPT

The MBPT is modelled by a class whose most important attributes are: a list of the image(s); a list of lists $\mathcal{W}_i$; the RAG; the consensus strategy; and an indexation matrix for mapping pixels from the image to the nodes. Another optional attribute of this class is a pre-segmented image (i.e., an initial partition of the image support) for non-trivial initialization.

This MBPT class does not actually represent a physical structure of tree. However it gathers all the required information for its creation and management. Its most important methods are: the growing method (essential for the creation process, since it contains the three main points of the algorithm described in Section 2.3); the regrowing method that allows for rebuilding the MBPT from a saved JSON structured file; and the cutting method for segmentation.

B.2 Data-structure of the RAG

As an attribute of the MBPT class, the RAG contains both the structure of the tree $\mathfrak{T}$, and that of the graph $\mathfrak{G}$. In particular, it contains a table

that gathers the nodes of $\mathfrak{G}$ and $\mathfrak{T}$. Indeed, each node of the graph is also one of the tree. This table is static. The required memory space allocation is done once, when preparing the RAG.

It also contains a static table of all adjacencies. This table regroups all the active adjacencies of $\mathfrak{G}$. Unlike the lists $\mathcal{W}_i$, this set is not ordered. Its main purpose is to store the adjacencies in the memory. During the creation process of the tree, the content of this table can only decrease. Indeed, some adjacencies are deleted and others are updated during node mergings. When an element is removed from this table, it is replaced by the last not empty element. Then, all empty cells remain at the end of the table. This leads to a compact structure. At the end of the node merging iterations, this table is totally empty.

B.3 Data-structure of a node

The nodes are important elements of both the MBPT tree $\mathfrak{T}$ and the graph $\mathfrak{G}$. Each node is modelled by a class whose principal attributes are: the type of the node (leaf, simple node or root); a list of pixels contained in the region modelled by the node; the left and right children of the node (plus its father); a list of adjacencies linking the node to its neighbours; and a list of feature values associated to all metrics and prepared by each of them.

B.4 Data-structure of an adjacency

The adjacencies are edges representing the links between neighbouring regions. Each adjacency is associated to n values computed from each metric. Adjacencies are modelled by a class, with the following attributes:

- a couple of neighbouring nodes;
- a table of metric values representing the distances between the two nodes (relatively to the n metrics);
- an index giving its position in the table of adjacencies of the RAG;
- a table of ranks;
- two tables of previous and next adjacencies.

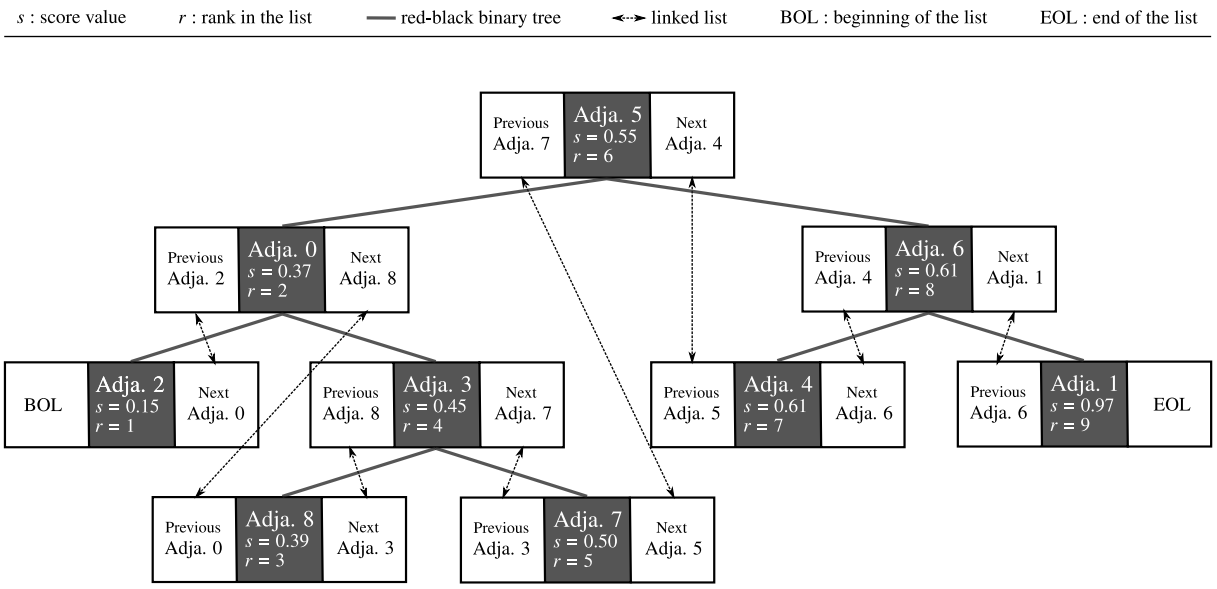


Figure 14: Illustration of the hybrid data-structure of a list $\mathcal{W}_i$ containing all adjacencies of the RAG (9 in this example) represented as black boxes. The list is implemented using a red-black binary search tree, maintained sorted according to the metric values (also called scores s) saved in each adjacency. The bold relations model the father-son relationships of the binary search tree. On top of this structure, a linked list of adjacencies is implemented (see B.5) to store the ranks r of each adjacency in the list $\mathcal{W}_i$: white boxes model pointers (previous and next elements in the list) while dashed relations model double linkages.

The last three attributes are used when handling the n lists $\mathcal{W}_i$ by particular consensus strategies. Their interest will be explained in the next section. Each adjacency implements some methods of the JAVA comparable interface. Surcharging those methods allows for the comparison between adjacencies. This facilitates their sorting in the lists $\mathcal{W}_i$, by using the table of metric values.

B.5 Data-structure of a list $\mathcal{W}_i$

The data-structures used for the n lists $\mathcal{W}_i$ are crucial in this algorithm. Indeed, most instructions related to the MBPT creation rely on them. Each list $\mathcal{W}_i$ contains all adjacencies of the RAG (see B.2) and must be maintained sorted. The sorting process of each list is based on the corresponding metric values (also called scores) saved in each adjacency (see B.4). The data-structure of the lists $\mathcal{W}_i$ is designed in order to lighten the MBPT creation process. In AGAT, a list $\mathcal{W}_i$ is modelled by a class containing a hybrid data-structure based on a red-black binary search tree.

This provides guaranteed logarithmic time cost for basic operations. An illustration of this hybrid data-structure is given in Figure 14. This structure is maintained sorted thanks to a specific comparator that determines which adjacency has to be placed before or after another. If some adjacencies have the same score, the least recent are placed before the most recent (e.g., Adja. 6 and Adja. 4 in Figure 14).

The proposed structure of a list $\mathcal{W}_i$ also allows each adjacency to know its rank r in each list. Then, the process can get rapidly the position of the adjacency in a list $\mathcal{W}_i$. This is useful to speed up the computation carried out by the consensus strategies, in particular in the case of relative and global consensus strategies (see Sections 2.2 and 2.2). Indeed, the ranks of each adjacency can be determined rapidly without having to iterate on all lists content. When a list $\mathcal{W}_i$ is modified (adding, removing, sorting elements), three specific attributes of each adjacency (see B.4) are updated:

- the table of ranks;
- the table of previous adjacencies;
- the table of next adjacencies.

These attributes allow for the formation of a linked list of adjacencies (combined to the red-black binary search tree), whose links are built according to the ranks of the adjacency in the list $\mathcal{W}_i$. In order to reduce the task and time computation related to the modification of each list, the ranks are only updated once per node fusion.

C Publications and communications

Publication in international journal (under review)

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Binary partition tree construction from multiple features for image segmentation,* Submitted, september 2017.

Publications in international conferences

- T.J.F. Randrianasoa, C. Kurtz, P. Gançarski, E. Desjardin and N. Passat, *Supervised evaluation of the quality of binary partition trees*

based on uncertain semantic ground-truth for image segmentation purpose, pp. 3874–3878, IEEE International Conference on Image Processing, Beijin, China, 2017.

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Multi-image segmentation: A collaborative approach based on binary partition trees,* International Symposium on Mathematical Morphology (ISMM), Lecture Notes in Computer Science, vol. 9082, pp. 253–264, Springer, Reykjavik, Iceland, 2015.

Publication in national conferences and workshops

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Réflexions sur l'évaluation supervisée de la qualité d'un arbre binaire de partitions d'image,* Conférence Extraction et Gestion des Connaissances – Atelier Fouille de Données Complexes (FDC@EGC), Grenoble, France, 2017.

- T.J.F. Randrianasoa, *Segmentation multi-image : une approche collaborative pour la construction d'arbres binaires de partitions,* Journées francophones des jeunes chercheurs en vision par ordinateur (ORASIS), Amiens, 2015.

Communications

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Construction d'arbres binaires de partitions à partir de caractéristiques multiples pour la segmentation d'images,* Spatial Analysis and GEOmatics – Atelier collaboration, classification, connaissances et données de l'environnement (C3DE@SAGEO), Nice, France, 2016.

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Construction d'arbres binaires de partitions multi-critères pour la segmentation d'images,* Poster, Journée du Gdr Madics, Paris, France, 2016.

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Construction d'arbres binaires de partitions multi-critères pour la segmentation d'images satellites,* Big Data Mining and Visualization, Journées des groupes de travail EGC, Metz, France, 2016

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Approches hiérarchiques et gestion de l'incertitude pour l'analyse d'images de télédétection,* Poster, Fouille Collaborative Incrémentale de Masses de Données Multisources Multitemporelle : Application en Sciences de l'Environnement, Strasbourg, France, 2016.

- T.J.F. Randrianasoa, C. Kurtz, E. Desjardin and N. Passat, *Segmentation multi-image : approche collaborative basée sur les arbres binaires de partitions,* 38e journée ISS France, Paris, France, 2015.

Représentation d'images hiérarchique multicritère

La segmentation est une tâche cruciale en analyse d'images. L'évolution des capteurs d'acquisition induit de nouvelles images de résolution élevée, contenant des objets hétérogènes. Il est aussi devenu courant d'obtenir des images d'une même scène à partir de plusieurs sources. Ceci rend difficile l'utilisation des méthodes de segmentation classiques. Les approches de segmentation hiérarchiques fournissent des solutions potentielles à ce problème. Ainsi, l'Arbre Binaire de Partitions (BPT) est une structure de données représentant le contenu d'une image à différentes échelles. Sa construction est généralement mono-critère (i.e. une image, une métrique) et fusionne progressivement des régions connexes similaires. Cependant, la métrique doit être définie a priori par l'utilisateur, et la gestion de plusieurs images se fait en regroupant de multiples informations issues de plusieurs bandes spectrales dans une seule métrique. Notre première contribution est une approche pour la construction multicritère d'un BPT. Elle établit un consensus entre plusieurs métriques, permettant d'obtenir un espace de segmentation hiérarchique unifiée. Par ailleurs, peu de travaux se sont intéressés à l'évaluation de ces structures hiérarchiques. Notre seconde contribution est une approche évaluant la qualité des BPTs en se basant sur l'analyse intrinsèque et extrinsèque, suivant des exemples issus de vérités-terrains. Nous discutons de l'utilité de cette approche pour l'évaluation d'un BPT donné mais aussi de la détermination de la combinaison de paramètres adéquats pour une application précise. Des expérimentations sur des images satellitaires mettent en évidence la pertinence de ces approches en segmentation d'images.

Arbre binaire de partitions, hiérarchie morphologique, segmentation d'image, multicritère, traitement d'image par graphe, évaluation de la qualité supervisée

Hierarchical multi-feature image representation

Segmentation is a crucial task in image analysis. Novel acquisition devices bring new images with higher resolutions, containing more heterogeneous objects. It becomes also easier to get many images of an area from different sources. This phenomenon is encountered in many domains (e.g. remote sensing, medical imaging) making difficult the use of classical image segmentation methods. Hierarchical segmentation approaches provide solutions to such issues. Particularly, the Binary Partition Tree (BPT) is a hierarchical data-structure modeling an image content at different scales. It is built in a mono-feature way (i.e. one image, one metric) by merging progressively similar connected regions. However, the metric has to be carefully thought by the user and the handling of several images is generally dealt with by gathering multiple information provided by various spectral bands into a single metric. Our first contribution is a generalized framework for the BPT construction in a multi-feature way. It relies on a strategy setting up a consensus between many metrics, allowing us to obtain a unified hierarchical segmentation space. Surprisingly, few works were devoted to the evaluation of hierarchical structures. Our second contribution is a framework for evaluating the quality of BPTs relying both on intrinsic and extrinsic quality analysis based on ground-truth examples. We also discuss about the use of this evaluation framework both for evaluating the quality of a given BPT and for determining which BPT should be built for a given application. Experiments using satellite images emphasize the relevance of the proposed frameworks in the context of image segmentation.

Binary partition tree, hierarchical morphology, image segmentation, multi-feature, graph-based image processing, supervised quality evaluation

Discipline : INFORMATIQUE

Spécialité : Analyse et traitement d'images

Université de Reims Champagne-Ardenne

CReSTIC EA 3804

Chemin des Rouliers, CS 30012, 51867 Reims