Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology

Camille Kurtz, Nicolas Passat, Pierre Gançarski, Anne Puissant

Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology

Camille Kurtz[a,*], Nicolas Passat[a], Pierre Gançarski[a], Anne Puissant[b]

[a] *Université de Strasbourg, LSIIT, UMR CNRS 7005, France*
[b] *Université de Strasbourg, LIVE, ERL CNRS 7230, France*

Abstract

The extraction of urban patterns from Very High Spatial Resolution (VHSR) optical images presents several challenges related to the size, the accuracy and the complexity of the considered data. Based on the availability of several optical images of a same scene at various resolutions (Medium, High, and Very High Spatial Resolution), a hierarchical approach is proposed to progressively extract segments of interest from the lowest to the highest resolution data, and then finally determine urban patterns from VHSR images. This approach, inspired by the principle of photo-interpretation, has for purpose to use as much as possible the user's skills while minimising his/her interaction. In order to do so, at each resolution, an interactive segmentation of one sample region is required for each semantic class of the image. Then, the user's behaviour is automatically reproduced in the remainder of the image. This process is mainly based on tree-cuts in binary partition trees. Since it strongly relies on user-defined segmentation examples, it can involve only low level –spatial and radiometric– criteria, then enabling fast computation of comprehensive results. Experiments performed on urban images datasets provide satisfactory results which may be further used for classification purpose.

Keywords: hierarchical segmentation, clustering, multisource images, multiresolution approaches, binary partition trees, remote sensing, urban analysis

1. Introduction

In the field of Earth observation, a new generation of sensors of submetric resolution [1] has led, at the end of the 90's, to the production of Very High Spatial Resolution (VHSR, less than 1m) optical images, and to an improved ability to analyse urban scenes. In such images, basic urban patterns (*e.g.*, individual houses, gardens, roads) are formed by different materials (*e.g.*, red or grey roofs, different asphalts or different kinds of vegetation), while complex ones (*e.g.*, urban districts, urban blocks) generally contain different kinds of basic patterns (Figure 1). Thus, by opposition to lower resolution images, such patterns are not necessarily composed of homogeneous pixels but are often hierarchically organised.

These specific properties induced by VHSR images lead to several new challenges. On one hand, the size and the complexity of the images[1] make the visual analysis a time consuming and error prone task for human experts. On the other hand, new image analysis tools have to be developed since methods developed for lower resolutions, *e.g.*, region-based ones [2, 3], are generally designed to extract segments based on radiometric homogeneous hypotheses.

In this context, and due to the actual importance to analyse VHSR images [4] in addition to lower spatial resolution ones, it is then relevant to develop tools adapted to the extraction of complex patterns from such data, and in particular (low-level) segmentation ones. Moreover, the availability of data with a large range of spatial resolutions can enable

*Corresponding author – LSIIT, Bd S. Brant, BP 10413, 67412 Illkirch Cedex, France – Tel.: +33 3 68 85 45 78 – Fax: +33 3 68 85 44 55
Email addresses: ckurtz@unistra.fr (Camille Kurtz), passat@unistra.fr (Nicolas Passat), gancarski@unistra.fr (Pierre Gançarski), anne.puissant@live-cnrs.unistra.fr (Anne Puissant)

[1] For instance, satellite VHSR imaging can produce multiresolution data with dimensions close to 20 000 × 20 000 pixels.

Figure 1: Example of object of interest (an urban block) bounded in red on a panchromatic satellite VHSR image with a spatial resolution of 60 cm (QuickBird, © DigitalGlobe Inc.).

the extraction of potentially hierarchical patterns, especially when such data are provided by various acquisition devices, providing complementary information at distinct radiometric bands [5] (Figure 2).

Such new segmentation tools should allow the user to obtain satisfactory results, at possibly different levels of pattern extraction (*i.e.*, scales), with minimal time (by automating the tasks which do not require human expertise), minimal efforts (by reducing the parameters induced by *a priori* knowledge), and ergonomic interaction.

In order to do so, it is possible to involve the data available at several resolutions (from Medium Spatial Resolution (MSR, 30–5m) to VHSR ones) [6] in a hierarchical strategy which enables, at a given resolution, the exploration of the whole structure of an urban scene [7]. By analysing first the image content at a coarse resolution and then gradually increasing this resolution, it is in particular possible to detect complex patterns (which structure the scene) while avoiding the semantic noise induced by the details [8].

Based on these considerations, a hierarchical approach is proposed to progressively extract from multiresolution urban images, segments of interest from the lowest to the highest resolution data (by opposition to ascendant approaches often proposed in the literature [9]), and then finally determine urban patterns from VHSR images. This approach, inspired from the principle of photo-interpretation, has for purpose to use as much as possible the user's skills while minimising his/her interaction. In order to do so, at each resolution, an interactive segmentation of one sample region is required for each semantic class of the image. Then, the user's behaviour is automatically reproduced in the remainder of the image. This process is mainly based on tree-cuts in binary partition trees. Since it strongly relies on user-defined segmentation examples, it can involve only low level –spatial and radiometric– criteria, then enabling fast computation of comprehensive results.

This article, which is an extended and improved version of the preliminary work described in [10], is organised as follows. Section 2 provides a state of the art of hierarchical and multiresolution segmentation approaches dealing with (but not restricted to) remote sensing data. Section 3 introduces some definitions and notations needed for the remainder of this article. Section 4 describes the proposed segmentation methodology. Section 5 gathers experiments enabling to assess the relevance of this methodology. Conclusions and perspectives will be found in Section 6.

2. Related works

2.1. Complex objects segmentation

Many efforts have been conducted to automatically extract features from satellite images, in order to involve them into learning systems. This extraction, often performed thanks to low-level processing, generally relies on radiometric homogeneousness hypothesis. This can lead to valid results for basic objects extraction (*e.g.*, individual houses, gardens, roads) from High Spatial Resolution (HSR, 3–1m) images [2], but not for images (*e.g.*, VHSR ones) and/or objects of higher complexity [3] (*e.g.*, urban districts, urban blocks).

A way to extract complex objects is by grouping several basic ones, using, for instance, a graph-based approach. A representative example is proposed in [9], where a graph-based structural pattern recognition system is used to infer broad categories of urban land use from HSR images. (This system has been considered to analyse discrete land cover

(a) 1 pixel = 9.6 m × 9.6 m (b) 1 pixel = 2.4 m × 2.4 m (c) 1 pixel = 60 cm × 60 cm

Figure 2: Satellite images representing the same geographical area with different spatial resolutions and in different radiometric bands. (a) Medium Spatial Resolution (MSR) image with four spectral bands (SPOT-5, © CNES). (b) High Spatial Resolution (HSR) image with four spectral bands (QUICKBIRD, © DigitalGlobe Inc.). (c) Very High Spatial Resolution (VHSR) image with one spectral band (QUICKBIRD, © DigitalGlobe Inc.).

parcels by taking into account the structural properties and the relations between simple objects.) Another example of such approaches can be found in [11] in which a set of particular subgraphs of a valued graph is introduced. However, two major problems are inherent to such approaches. Firstly, the computation of all the grouping possibilities within the (large) space of candidate segments is not actually tractable. Secondly, the ability to detect complex objects is directly linked to the quality of the initial partition of the image. Such techniques, devoted to the "first" semantic level of complex objects (*e.g.*, complex buildings) then seem unable to directly extract more complex structures at higher semantic levels.

Moreover, when dealing with HSR images, different composite objects could be merged to form new kinds of structures of interest, enabling different levels of analysis. This frequent issue is often known as the "scale problem". For instance, the main environments, such as urban areas, rural zones, or forests, can be identified at coarsest levels, while more detailed structures, such as buildings and roads, will emerge at the finest ones [2]. Thus, different objects can be extracted at various scales and be related according to some suitable criteria in a hierarchical structure. Consequently, "grouping approaches" have to be improved by considering hierarchical strategies to enable the (potential) extraction of more complex structures (with a higher semantic level). To this end, some techniques providing a multiscale partitioning have been proposed.

2.2. Multiscale partitioning

Multiscale/hierarchical segmentation methods compute a series of partitions of an image with an increasing (or decreasing) level of details. Such methods have been widely studied for the last decades (see [12] for an example of pioneering work).

In the field of remote sensing (and especially for HSR images), several techniques have been proposed. In [13], compositions of opening and closing operations with structuring elements (SEs) of increasing sizes generate morphological profiles for any pixel, enabling their characterisation. Although morphological profiles are sensitive to different pixel neighbourhoods, the segmentation decision is performed by individually evaluating pixels without considering the neighbourhood information. However, the assumption that all pixels in a structure have only one significant derivative maximum occurring at the same SE size, often does not hold for VHSR images. To overcome this limitation, new approaches have been proposed. In [7], morphological profiles are enriched with neighbourhood and spectral information, while in [14], a framework is proposed to detect complex objects in HSR images by combining spectral information with structural ones. These approaches emphasise the potential of hierarchical segmentation. However, these *pixel-based* methods hardly take into account the intrinsic and semantic information of the images, by opposition to *object-based* ones.

An object-based segmentation hierarchy is a set of image segmentations at different levels of details in which the segmentation at a level can be produced by merging regions at finer ones. Such hierarchies can be built by following two opposite paths. In the top-down approaches, the process starts from a coarse segmentation and successively refines the regions, as in [15], where segmentation is treated as a graph partitioning problem. However, such an approach, which assumes that the images contain only few objects of interest is not adapted to capture the richness

and complexity of (V)HSR images. Another approach can be found in [16], where a top-down construction scheme for irregular pyramids is presented. Starting from an initial topological map, regions are successively refined by splitting operations. However, finding a relevant (and robust) splitting function remains an open issue. In the, more frequent, bottom-up approaches ("region merging" or "split and merge" methods), the finest segmentations are produced first, and their regions are then merged, based on similarity criteria. In remote sensing, various algorithms use this principle. For instance, in [17], a hierarchical segmentation algorithm that combines spectral clustering with iterative region growing is proposed. The multiscale segmentation algorithm presented in [18] also consists of bottom-up region merging, where each pixel is initially considered as a separate object, and pairs of objects are iteratively merged to form new larger ones. In [19], segmentation is performed through a region merging process carried out by hierarchical stepwise optimisation. The main issue with such approaches is that the segmentation results depend on a user-defined threshold related to local homogeneity criteria. An alternative solution is proposed in [20]. In this approach, the goal is to detect complex urban structures using a hierarchical multiple Markov chain model. It considers the image as a complex collection of textures, emerging at different scales of observation, and non-textured patches. The merging process exploits textural image properties, together with spatial and spectral ones, in order to recognise the semantic unity of complex regions. However, such criteria, useful to extract textural objects, are not relevant for objects formed by several heterogeneous components.

In mathematical morphology, connected operators [21, 22] may be used in a hierarchical segmentation fashion by using, for instance, tree data structures. Notions such as component-tree [23] and level-lines tree [24], potentially enable to perform hierarchical segmentation, by fusion of flat zones. However, such structures strongly rely on the image intensity and in particular on extremal values. Thus, the obtained segmented components may be non relevant in the case of satellite images. By opposition, the binary partition tree (BPT) [25] reflects a (chosen) similarity measure between neighbouring regions, and models the hierarchy between these regions *via* the tree structure. The BPT represents a set of regions at different scales and its nodes provide good estimates (with respect to the chosen measure) of the objects in the scene. It has been used to extract complex objects from various kinds of images [26, 27]. A last approach, based on the constrained connectivity paradigm, has been recently introduced in [28] and applied to process (V)HSR images in [29]. The connectivity relation generates a partition of the image definition domain. Fine to coarse partition hierarchies are then produced by varying a threshold value associated with each connectivity constraint.

Approaches based on connected operators have shown encouraging results in the context of complex objects extraction. However, in the case of remote sensing, they are limited by the spatial and spectral properties of the images. Indeed, complex objects appear in (V)HSR images too much heterogeneous to be extracted in an ascendant way. This justifies the use of multiresolution data to enhance their ability to extract such complex objects.

2.3. Exploiting multiresolution data

Structures of interest, in satellite images, are generally formed by various heterogeneous objects. Consequently, such structures may have very different sizes and shapes. To deal with this variability, two solutions can be considered: (1) to use size/scale invariant features, or (2) to process the image at different spatial resolutions. This last solution enables, in particular, to adapt the level of the objects extraction task to a specific "synthetic" spatial resolution. For instance, if the resolution becomes coarser than that of the original image, the larger (and thus, complex) structures, which provide the global image context, can be extracted without being bothered with the details [30]. This property has led to the development of segmentation methods using multiresolution data.

A way to deal with multiresolution data is to generate images with lower resolution than the original (monoresolution) one in order to enable the extraction of different levels of details. Numerous approaches have been proposed using the wavelet transform [31], which provides a hierarchical framework for interpreting the image. In particular, in [32, 33], some extensions of the watershed segmentation have been proposed to deal with multiresolution images provided by wavelet approaches. However the major drawback of this family of methods is to introduce new contours during the segmentation process, which are not relevant to the extraction of complex objects.

In remote sensing, the wide variety of sensors directly provides multiresolution sets of images (*e.g.*, MSR, HSR, VHSR). Thus, it is not required to produce degraded images. A way to deal with such multiresolution satellite images consists of combining all the descriptions of the objects associated to the different resolutions into a unique image at the highest resolution [34, 35] and then segmenting the output result. For instance, several approaches [36, 37] use a pansharpening fusion technique, which fuses low spatial resolution multispectral images with high

spatial resolution panchromatic images to obtain high spatial resolution multispectral images. In such cases, the segmentation performances are, of course, affected by possible errors induced by this fusion step.

Recently, new approaches dealing with multiresolution images without fusion have been proposed. These methods aim at discovering the structural decomposition of the studied scenes by using images with different spatial resolutions. For instance, in [6], a hierarchical multiresolution segmentation method is proposed to extract complex object from such images. Based on a bottom-up approach, the proposed algorithm works first on the high-resolution data, performing an over-segmentation devoted to preserve fine details. This initial over-segmentation produces a large number of elementary regions which are then progressively merged, based on both spectral and spatial properties, in a hierarchical fashion. This method provides promising results but does not fully exploit the richness offered by the images at low resolutions. Indeed, it may seem relevant to possibly adopt an opposite strategy. By analysing first the image content at a coarse resolution and then gradually increasing this resolution, it is possible to detect complex patterns while avoiding the semantic noise induced by the details, as proposed, *e.g.*, in [38]. This strategy, which is also comparable to the human vision system [39], has already been considered in [8] to create thematic maps from HSR and MSR images.

2.4. Purpose

To conclude on this state of the art, two main problems can occur when dealing with (Very) High Spatial Resolution data. The first one is a complexity issue which is directly linked to the size, the complexity and the accuracy of this data. The second one is the scale problem, which appears when different objects of interest can emerge at various scales through the same image.

Based on these considerations, we propose a hierarchical top-down analysis methodology involving segmentation and clustering. It combines the advantages of multiresolution strategies and the efficiency of the connected operators approaches, in the context of the mapping of urban areas. It relies on interactive BPT segmentation (based on the skills of the user), defined interactively on a part of the images, and automatically reproduced on the whole remaining data.

This multiresolution top-down property enables to deal with the complexity and memory issues of the considered data while the proposed example-based *modus operandi* (combined with the multiresolution clustering process) enables to deal with the "scale" issue. The main idea is to adapt the segmentation process (and/or the segmentation parameters) to local areas of homogeneous classes of radiometric intensity instead of segmenting a whole image using only one segmentation parameter (*e.g.*, only one scale parameter).

The method operates first on the low resolution data, extracting the global structure of the scene, and subsequently enriches this description thanks to the high resolution data.

3. Definitions and notations

3.1. Sets and functions

Let X be a finite set. The set of all the subsets of X, namely $\{Y \mid Y \subseteq X\}$ is noted 2^X. The cardinal of X is noted $|X|$. If a set $\{X_i\}_{i=1}^t \in 2^X$ $(t \geq 1)$ of subsets of X is a partition of X, we note that $X = \bigsqcup_{i=1}^t X_i$ (or $X = X_1 \sqcup X_2 \ldots \sqcup X_t$).

A function F from a set X to a set Y is noted $F : X \to Y$. For any $Z \subseteq X$, the image of Z by F, namely $\{F(z) \mid z \in Z\}$ is noted $F(Z)$. For any $T \subseteq Y$, the preimage of T by F, namely $\{t \mid F(t) \in T\}$ is noted $F^{-1}(T)$.

An interval on $\mathbb{R}$, bounded by $a, b \in \mathbb{R}$, will be noted $[a, b]$. An interval on $\mathbb{Z}$, bounded by $a, b \in \mathbb{Z}$, will be noted $[\![a, b]\!]$.

3.2. Images

Let $E = [\![0, d_x - 1]\!] \times [\![0, d_y - 1]\!] \subset \mathbb{Z}^2$. The set E corresponds to the discretisation of the continuous space (*i.e.*, the part of $\mathbb{R}^2$) which will be visualised in the images. (Note that, without loss of generality, E may also be any connected subset of $\mathbb{Z}^2$, for a given connectivity, *e.g.*, the 4- or 8-connectivity.) An element $\mathbf{x} = (x, y) \in E$ is called a pixel. It physically corresponds to a cubic square region in the continuous counterpart of E.

Let $V_b = [\![0, v_b - 1]\!] \subset \mathbb{Z}$. The set V_b corresponds to the discrete sampling of the intensities observed for a given spectral band. Let $V = \prod_{b=1}^s V_b \subset \mathbb{Z}^s$ $(s \geq 1)$. The set V corresponds to the discrete sampling of the intensities observed for s given spectral bands.

A monovalue image is a function $\mathcal{I}_b : E \to V_b$ which to each point $\mathbf{x} = (x, y) \in E$ of the scene, associates a spectral intensity $\mathcal{I}_b(\mathbf{x}) = v$ in exactly one spectral band.

A multivalue image is a function $\mathcal{I} : E \to V$ (with $s > 1$) which to each point $\mathbf{x} = (x, y) \in E$ associates a s-uple of spectral intensities $\mathcal{I}(\mathbf{x}) = \mathbf{v} = \prod_{b=1}^{s} \mathcal{I}_b(\mathbf{x})$ in the considered spectral bands.

3.3. Segmentation, clustering

A segmentation of an image $\mathcal{I} : E \to V$ is a partition $\mathfrak{S} = \{R_i\}_{i=1}^{r}$ ($r \geq 2$) of E. Equivalently, such a segmentation $\mathfrak{S}$ can be considered as a function $\mathcal{I}_{\mathfrak{S}} : E \to [\![1, r]\!]$ (*i.e.*, a "false colour" image) unambiguously defined by $R_i = \mathcal{I}_{\mathfrak{S}}^{-1}(\{i\})$ for all $i \in [\![1, r]\!]$.

Let $\mathfrak{S} = \{R_i\}_{i=1}^{r}$ be a partition of E, associated to an image $\mathcal{I} : E \to V$, and $\mathcal{I}_{\mathfrak{S}} : E \to [\![1, r]\!]$ be the image induced by $\mathfrak{S}$. A clustering[2] of $\mathcal{I}$ into u clusters is provided by the definition of a map $C : [\![1, r]\!] \to [\![1, u]\!]$ which, to each one of the r regions R_i, associates one of the u clusters $C(i)$. A cluster K_i induced by such a clustering is then defined by $K_i = \bigcup_{j \in C^{-1}(\{i\})} R_j$, *i.e.*, by gathering all the regions R_j which correspond to a same cluster. Similarly to the case of segmentation, each clustering C of an image $\mathcal{I}$ partitioned by $\mathfrak{S}$ can be considered as a function $\mathcal{I}_C : E \to [\![1, u]\!]$ (*i.e.*, a "false colour" image) unambiguously defined by $\mathcal{I}_C = C \circ \mathcal{I}_{\mathfrak{S}}$.

3.4. Histograms

The histogram of an image $\mathcal{I} : E \to V$, is the function $\mathcal{H}_{\mathcal{I}} : V \to \mathbb{N}$ which associates to each value $v \in V$ the number $\mathcal{H}_{\mathcal{I}}(v) = |\mathcal{I}^{-1}(\{v\})|$ of pixels of $\mathcal{I}$ of value v.

The histogram of $\mathcal{I}$ associated to a subset $X \subseteq E$ is the function $\mathcal{H}_{\mathcal{I},X} : V \to \mathbb{N}$ which associates to each value $v \in V$ the number $\mathcal{H}_{\mathcal{I}}(v) = |\mathcal{I}^{-1}(\{v\}) \cap X|$, *i.e.*, the histogram of the restriction of $\mathcal{I}$ to X.

3.5. Binary partition tree

Let $\mathcal{I} : E \to V$ be an image. A binary partition tree (BPT) [25] of $\mathcal{I}$ is a tree data-structure that provides a hierarchy of regions of E with respect to $\mathcal{I}$. More formally, a BPT of $\mathcal{I}$ is a couple $(\mathcal{N}, \varphi)$ such that $\mathcal{N} \subseteq 2^E$ is a set of subsets of E verifying $E \in \mathcal{N}$, and $\varphi : \mathcal{N} \setminus \{E\} \to \mathcal{N}$ is a function verifying the following property:

Property for any $N \in \varphi(\mathcal{N} \setminus \{E\})$ we have $\varphi^{-1}(\{N\}) = \{N_1, N_2\}$ such that $N_1 \neq N_2 \in \mathcal{N}$ and $N = N_1 \sqcup N_2$.

The elements of $\mathcal{N}$ are called the nodes of the BPT. The function φ models the "parent" relation between the nodes: broadly speaking, if $N = \varphi(N')$, then N (resp. N') is the "father" (resp. a "child") of N' (resp. N). The node E is the root of the BPT. The nodes of $\mathcal{N} \setminus \varphi(\mathcal{N} \setminus \{E\})$, *i.e.*, those having no children, are the leaves of the BPT.

Practically, φ enables to recursively divide E into several partitions, successively obtained by splitting exactly one element of the current partition into two subsets. Note in particular that the set $\mathcal{N} \setminus \varphi(\mathcal{N} \setminus \{E\})$ (resp. $\{E\}$) constitutes a partition, and actually the finest (resp. the coarsest) one, of E with respect to φ.

Each subset $C \subseteq \mathcal{N}$ of nodes such that C is a partition of E is called a cut. Practically, the nodes of C define a subtree of the initial BPT, of root E and of leaves C (this tree being also a BPT).

4. Methodology

The proposed multiresolution methodology is dedicated to hierarchically segment $n \geq 2$ images of a same scene at various resolutions, from the lowest to the highest one, enabling different scales of interpretation. In the standard case, three images are considered, namely a MSR (30–5m), a HSR (3–1m) and a VHSR (less than 1m) image.

This segmentation methodology, which constitutes the main contribution of this article, performs n successive steps (one step per resolution), each step being iteratively composed of:

(*i*) an example-based segmentation approach;
(*ii*) a multiresolution clustering approach.

[2] This definition which enables to conveniently formalise object-based clustering, also enables to deal with pixel-based clustering by simply assuming that $\mathfrak{S} = \{\{\mathbf{x}\} \mid \mathbf{x} \in E\}$, *i.e.*, by partitioning the image into pixels instead of larger regions.

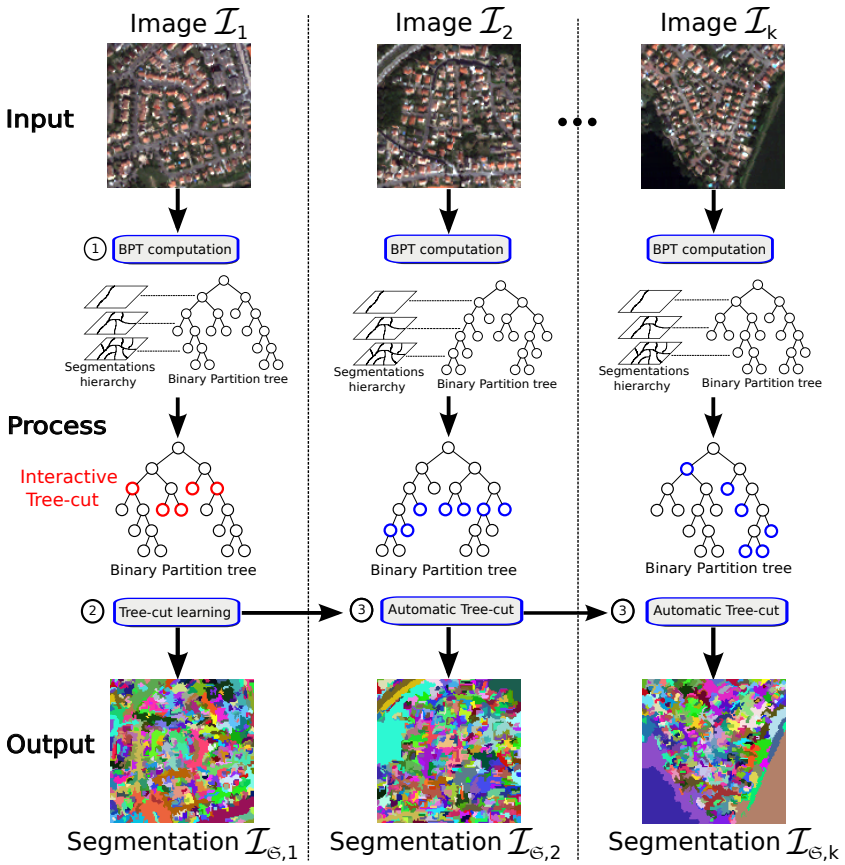


Figure 3: Example-based segmentation approach (see Section 4.1). In red: user interactions. In blue: automatic processing.

At each resolution/step, the output of the process (namely a segmentation map) is embedded into the next resolution image to be treated as input of the next step.

The example-based segmentation approach (*i*) is an original interactive strategy, which constitutes another contribution of this article. It is first presented in details in Section 4.1.

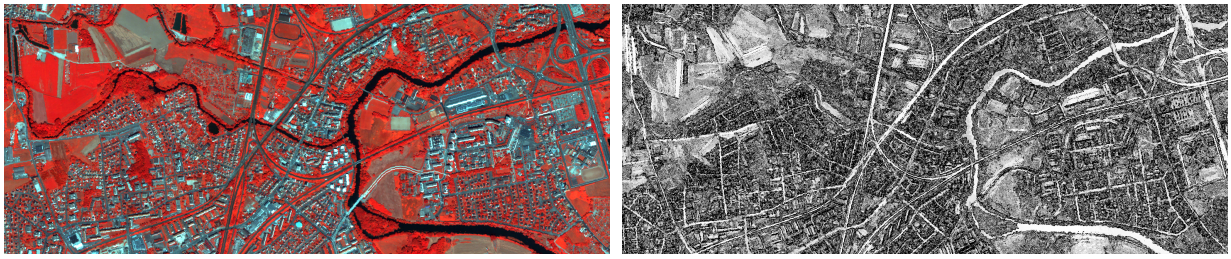
The whole segmentation methodology, is then presented in details in Section 4.2. Since the multiresolution clustering approach (*ii*) has already been fully described by the authors in [40], it is only briefly recalled in Section 4.2.2.

4.1. Example-based segmentation

One of the main steps of the proposed segmentation methodology consists of a segmentation approach visually summarised in Figure 3.

This approach takes as input $k \geq 2$ images (or $k \geq 2$ parts of the same image) representing k different (but specific) areas. All the k considered images (or sub-images) must have the same resolution and semantic, and must be provided by the same sensor (*e.g.*, 10 images of 10 distinct urban districts composed of roads and individual houses).

For one of the k images, the user first interactively performs a segmentation, by providing a cut in its BPT (Figure 3-①). This cut is assumed to correctly characterise the user-defined segmentation, and is then considered as an example to reproduce in the BPTs of the $k-1$ other images (Figure 3-②, ③). It is then possible to re-apply this approach to segment each one of the different sets of areas which have the same semantic content and which could compose a sensed scene (*i.e.* the different thematic/semantic classes). The three key-points of the approach are then (*i*) the way to build the BPTs (Section 4.1.1), (*ii*) the learning of the cut example on one image (Section 4.1.2), and (*iii*) its automatic reproduction in the $k-1$ other ones (Section 4.1.3).



(a) HSR image (b) Elongation map

Figure 4: Elongation map computation. (a) HSR image with a spatial resolution of 2.4 m (QuickBird, © DigitalGlobe Inc.). (b) Corresponding elongation map (elongated structures in light grey, non-elongated ones in dark grey).

4.1.1. Computing the binary partition trees

As stated in Section 3.5, the BPT of an image $\mathcal{I} : E \to V$ is built in a bottom-up fashion, *i.e.*, from its leaves to its root. Practically, based on an initial partition of E (generally composed by the singleton sets $\{\mathbf{x}\}$, for all $\mathbf{x} \in E$, or by the flat zones of $\mathcal{I}$), the nodes of $\mathcal{N}$ (and thus φ) are successively defined by fusion of couples of (already defined) nodes of $\mathcal{N}$ for which φ has not been defined yet. (In the context of image segmentation, such couples of nodes are generally chosen as spatially adjacent ones, thus leading to connected nodes in $\mathcal{N}$.)

A huge number of distinct BPTs may be obtained for a unique initial partition of E. In order to decide which one among them will be the most relevant, it is then necessary to define a "merging order", *i.e.*, to decide of the priority of the fusions between nodes. A BPT generation then relies on two main notions: a *region model* (which specifies how regions are characterised), and a *merging criterion* (which defines the similarity of neighbouring regions and thus the merging order).

The basic models and criteria used in most of image segmentation approaches are generally radiometric homogeneity [41]. However, when dealing with (V)HSR images, geometric details also have to be taken in consideration. Consequently, we propose to rely on both the increase of the ranges of the intensity values (for each spectral band) and on area and elongation of the regions in order to merge in priority objects which do not structure the scene.

In the sequel, the chosen region model and merging criterion are defined. It may be noticed that it has been chosen to involve only low-level properties in these notions, since we consider that the high-level (semantic) knowledge is provided to the approach by the user, *via* his/her segmentation example.

Region model. A node/region $R_i \in \mathcal{N}$ ($R_i \subseteq E$) is modelled here by a couple of values

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

where $v_b^\star$ provides the extremal values for the b-th spectral band in $\mathcal{I}$ (*i.e.*, in $\mathcal{I}_b$), while e and a provide the elongation and the area, respectively. Broadly speaking, M_r and M_g provide (low-level) spectral and geometric information. During the merging process, the region model of two merged regions R_i and R_j is then provided by

$$\begin{aligned} M_r(R_i \cup R_j) &= \langle (\min\{v_b^-(R_i), v_b^-(R_j)\}, \max\{v_b^+(R_i), v_b^+(R_j)\}) \rangle_{b=1}^s \\ M_g(R_i \cup R_j) &= (e(R_i \cup R_j), a(R_i) + a(R_j)) \end{aligned} \tag{2}$$

By opposition to M_r and a, whose computation is actually straightforward, the elongation e requires to (pre)compute an elongation map associated to $\mathcal{I}$ (which will emphasise linear structures, *e.g.*, roads, rivers and railways) thus dividing E into (large) regions. The detection of linear, or more generally elongated, structures has led to a huge literature (the description of which is out of the scope of this article). Our purpose, here, is not to get the best elongation results, but to be able to compute correct elongations with a low computational cost. Following this heuristic (but pragmatic) policy, the following strategy is considered for generating the elongation map e:

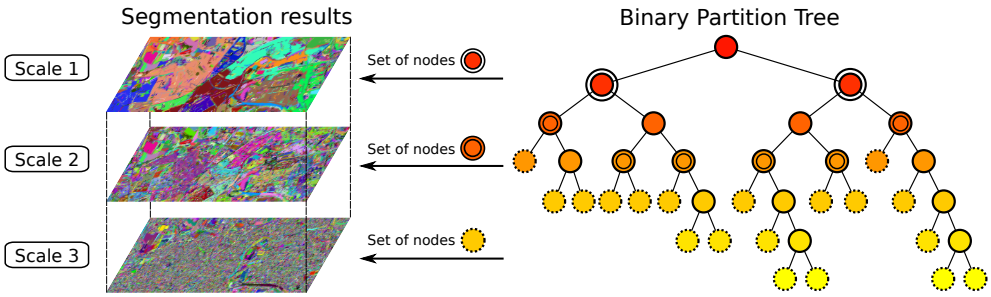


Figure 5: An example of BPT associated to the HSR image $\mathcal{I} : E \to V$ presented in Figure 4(a) (the number of nodes is significantly reduced, for the sake of readability). The nodes of $\mathcal{N}$ are depicted by colour disks (the root E is the highest node). The function φ is modelled by the couples of edges (linking two nodes N_1, N_2 with their common parent node $\varphi(N_1) = \varphi(N_2)$. The colours of the nodes (from yellow to red) symbolise the decrease of the similarity measure O_r between two neighbouring regions (and thus, also the decrease of the function α controlling the trade-off between O_r and O_g). For the sake of visualisation, three partitions of E associated to three cuts of the BPT are depicted.

1. for each pixel $\mathbf{x} \in E$ (considered as a seed), a series of region-growing segmentations (based on radiometric intensity) is performed with an increasing tolerance $\lambda \in [\![0, v_b^+(E) - v_b^-(E)]\!]$;
2. for each segmentation, a score is computed using the ratio width/length of the best bounding box of the region (computed in several discrete orientations);
3. the best (*i.e.*, the highest) elongation value is then assigned to $\mathbf{x}$.

This approach presents an algorithmic cost bounded, for each pixel, by the area of the neighbourhood where Step (1.) is carried out (which, in practice, needs not to be high). The computation of the elongation map is then globally linear with respect to the size of E. Figure 4 provides an example of an elongation map computed on a HSR image with a spatial resolution of 2.4 m and obtained thanks to this heuristic strategy.

Merging criterion. At each step, the algorithm determines the pair of most similar connected regions minimising the increase of the ranges of the intensity values (for each spectral band) and having low elongation/area properties. This leads to the following merging criteria

$$\begin{aligned} O_r(R_i, R_j) &= \frac{1}{s} \sum_{b=1}^{s} \frac{\max\{v_b^+(R_i), v_b^+(R_j)\} - \min\{v_b^-(R_i), v_b^-(R_j)\}}{v_b^+(E) - v_b^-(E)} \\ O_g(R_i, R_j) &= \frac{1}{2}(e(R_i \cup R_j) + a(R_i \cup R_j)) \end{aligned} \tag{3}$$

where $e(R_i \cup R_j)$ and $a(R_i \cup R_j)$ are normalised using the maximal possible values of these features. The similarity measure between two neighbouring regions R_i and R_j can then be computed as

$$O(R_i, R_j) = \alpha.O_r(R_i, R_j) + (1 - \alpha).O_g(R_i, R_j) \tag{4}$$

with $\alpha \in [0, 1]$. Note that $O_r(R_i, R_j)$ and $O_g(R_i, R_j)$ are normalised by construction. In practice, the closer the nodes are to the root, the less relevant O_r is. Consequently, the weight α can be defined as a function depending directly on the value of O_r (and decreasing when O_r increases). In particular, it has been experimentally observed that a standard Gaussian formulation

$$\alpha(O_r) = \exp(-O_r^2) \tag{5}$$

provides a satisfactory behaviour of the merging function O. (Note in particular that the user may be free to tune this function, by easily introducing parameters in $\alpha(O_r)$ enabling to control, *e.g.*, the asymptotic behaviour of α and/or the value of O_r for which $\alpha = 1 - \alpha$.)

Based on these chosen region model and merging criterion, the BPT can then finally be built, as exemplified in Figure 5.

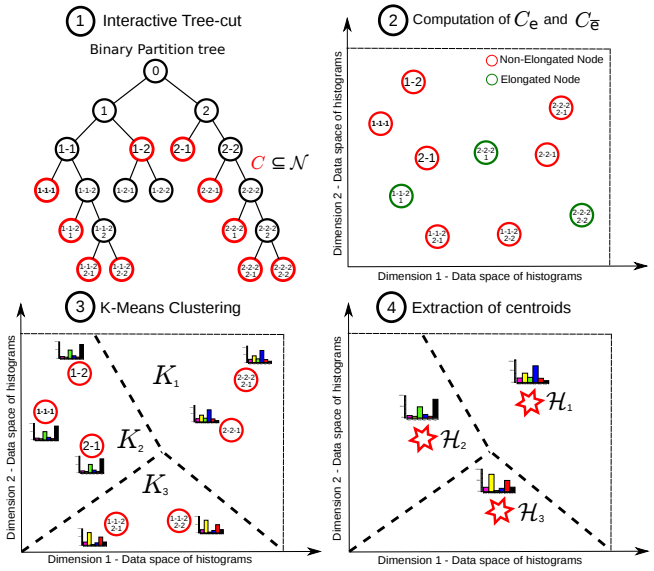


Figure 6: Illustration of the proposed example-based segmentation approach: learning of the cut example (see Section 4.1.2)

4.1.2. Learning of the cut example

By opposition to other strategies devoted to automatically extract cuts from partition hierarchies [7, 42], with the risk of generating non-relevant results, we propose to learn the user's behaviour from a segmentation example.

Indeed, the proposed approach allows the user to interactively select a relevant segmentation in one[3] of the k images, and equivalently, a relevant cut in the BPT of this image. In order to be able to "reproduce at best" this example in the other $k-1$ images/BPTs, it is first necessary to learn this example, *i.e.*, to extract the main properties and features which characterise it and then enable its reproduction.

In previous works [10], the cuts in these BPTs were straightforwardly obtained by performing a thresholding on the similarity measure related to the O function (Equation (4)) and attached to each node of the tree (by saving for a node $R_i \in \mathcal{N}$, the value of O required to create this node), at the value induced by the user's example. In the sequel of this section, we provide an alternative strategy designed to more accurately mimic the user's behaviour. The reader may refer to Figure 6 for a visual outline of this approach.

Let $C \subseteq \mathcal{N}$ be the cut defined in the BPT interactively processed by the user (Figure 6-①). In order to extract the most relevant nodes which characterise this cut, it is necessary to first separate linear elements from those which correspond to the areas "bounded" by these linear elements. Thus, C is partitioned into two subsets C_e and $C_{\bar{e}}$, corresponding to the nodes/regions being elongated and non-elongated, respectively (Figure 6-②). Such a partition can be straightforwardly obtained by a 2-class clustering process, *e.g.*, a K-means based on the attribute e of the nodes. The objects of interest for the proposed approach are then those of $C_{\bar{e}}$, which correspond to the areas "bounded" by the linear elements of C_e.

In order to extract the most relevant elements which characterise $C_{\bar{e}}$, a partitional clustering process is then carried out on the regions of $C_{\bar{e}}$ (Figure 6-③). This clustering is based on the histogram of each region, *i.e.*, for each region $R \in C_{\bar{e}}$ of $\mathcal{I} : E \rightarrow V$ (with $R \subseteq E$), the criterion characterising R is its (normalised) histogram $\mathcal{H}_{\mathcal{I},X} : V \rightarrow \mathbb{N}$ (with $\sum_{v \in V} \mathcal{H}_{\mathcal{I},X}(v) = 1$). Partitional clustering algorithms require a distance (and an averaging method) to compare each object to classify. In the following, we will denote by d the distance used to compare two histograms. The number of clusters u, which characterises the number of "relevant" families of nodes composing $C_{\bar{e}}$, can be parametrised by the user. This process leads to the definition of a set of u clusters $\{K_i\}_{i=1}^{u}$, associated to a set of u centroids $\{\mathcal{H}_i\}_{i=1}^{u}$, each centroid $\mathcal{H}_i$ being actually an "averaged" (normalised) histogram of the cluster K_i (Figure 6-④).

[3] Without loss of generality, the approach also enables to select more than one example in order to enhance the learning process. This will be discussed in Section 5.3.

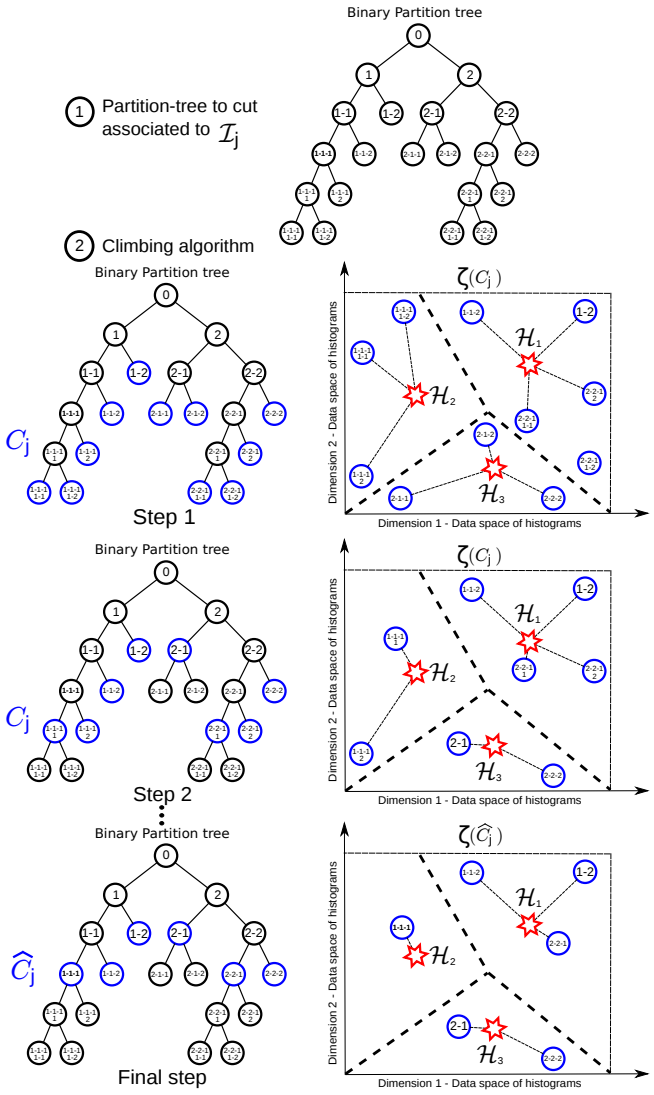


Figure 7: Illustration of the proposed example-based segmentation approach: automatic reproduction of the cut example (see Section 4.1.3).

4.1.3. Automatic reproduction of the cut example

The segmentation example provided by the user is then modelled by the u centroids obtained from the cut of the BPT of one of the k images (Figure 6-④). These centroids then have to be involved in the automatic segmentation of the $k-1$ other images (Figure 7).

This can be conveniently done by finding, for each one of the $k-1$ images $\mathcal{I}_j$ ($j \in [\![2,k]\!]$), a cut $\widehat{C}_j$ in the BPT of $\mathcal{I}_j$, minimising a scatter measure between the set of centroids $\{\mathcal{H}_i\}_{i=1}^u$ and the set of nodes C_j (or, more precisely, the set of normalised histograms $\{\mathcal{H}_{\mathcal{I}_j,X}\}_{X\in C_j}$). This scatter measure $\zeta(C_j)$ associated to a cut C_j, with respect to the set of clusters $\{\mathcal{H}_i\}_{i=1}^u$, can be defined as

$$\zeta(C_j) = \sum_{i=1}^{u} \frac{|\bigcup_{X\in C_j^i} X|}{|\bigcup_{X\in C_j} X|} \cdot d(\overline{\mathcal{H}_{i,j}}, \mathcal{H}_i) \tag{6}$$

where d is the distance used to compare two histograms and $C_j^i \subseteq C_j$ is the set of the nodes whose histogram is closer (with respect to d) of the centroid $\mathcal{H}_i$ than of any other $u-1$ centroids (note that $C_j = \bigsqcup_{i=1}^u C_j^i$), and $\overline{\mathcal{H}_{i,j}}$ is the

(weighted) mean histogram of the nodes $X \in C_j^i$, *i.e.*

$$\overline{\mathcal{H}_{i,j}} = \sum_{X \in C_j^i} \frac{|X|}{|\bigcup_{Y \in C_j^i} Y|} \cdot \mathcal{H}_{\mathcal{I}_j, X} \tag{7}$$

A climbing algorithm can then be applied to find the best cut $\widehat{C}_j \subseteq \mathcal{N}_j$ among the set of nodes $\mathcal{N}_j$ of the BPT of $\mathcal{I}_j$. This algorithm can be formalised[4] as

$$\widehat{C}_j = \mathcal{F}(E) \tag{8}$$

where $\mathcal{F} : \mathcal{N}_j \to 2^{\mathcal{N}_j}$ is (recursively) defined as

$$\mathcal{F}(N) = \{N\} \tag{9}$$

if $N \notin \varphi(\mathcal{N}_j \setminus \{E\})$, *i.e.*, if N is a leaf of the BPT, and as

$$\mathcal{F}(N) = \begin{cases} \{N\} & \text{if } \zeta(\{N\}) \leq \sum_{N' \in \varphi^{-1}(\{N\})} \zeta(\mathcal{F}(N')) \\ \bigcup_{N' \in \varphi^{-1}(\{N\})} \mathcal{F}(N') & \text{otherwise} \end{cases} \tag{10}$$

if $N \in \varphi(\mathcal{N}_j \setminus \{E\})$, *i.e.*, if N is not a leaf of the BPT.

By performing this algorithm on each one of the $k-1$ images, we then automatically obtain $k-1$ segmentations with a level of details (*e.g.*, a scale) similar to that of the segmentation example provided by the user.

Figure 7 exemplifies the automatic reproduction of the cut example on one BPT (associated to one image $\mathcal{I}_j$) among the $k-1$ BPTs to cut. The BPT associated to $\mathcal{I}_j$ is depicted in Figure 7-①. The application of the climbing algorithm in order to cut this BPT is then illustrated in Figure 7-②.

4.2. Multiresolution methodology

In this section, we now describe the whole multiresolution methodology, which constitutes the core of this article. This methodology is devoted to hierarchically segment several images of a same scene, at various resolutions, from the lowest to the highest one.

Practically, it takes as input:

- a set $\{\mathcal{I}_i : E_i \to V_i\}_{i=1}^n$ of $n \geq 2$ images ($n = 3$ in the general cases, see Section 5) of a same scene, at increasing resolutions, and with possibly different sensors (and thus different spectral bands);

and provides as output:

- a set of $\{\mathcal{I}_{\overline{\varepsilon},i}\}_{i=1}^n$ of segmented/clustered images (one per considered image/resolution), hierarchically linked, enabling different scales of interpretation.

The methodology is divided into n successive (and similar) steps, each step being devoted to the analysis of one image $\mathcal{I}_i$ among the n ones (from the image $\mathcal{I}_1$ of lowest resolution, to the image $\mathcal{I}_n$ of highest resolution). At the t-th step, the image $\mathcal{I}_t$, is considered (it is then assumed that the images $\mathcal{I}_i$ ($i \in [\![1, t-1]\!]$) have already been processed).

Each step relies on (*i*) the segmentation of the current image (Section 4.2.1), and (*ii*) its multiresolution clustering (Section 4.2.2). The reader may refer to Figure 8 to visually follow the description of the methodology.

[4] It can be noticed that this algorithm is actually better suited to be applied to a restricted part of the BPT of $\mathcal{I}_j$, which corresponds to the tree induced by the subset $\mathcal{N}_{j,\overline{\varepsilon}} \subseteq N_j$ of the non-linear regions of $\mathcal{I}_j$. This is justified by the fact that the involved u centroids have been obtained from the clustering of the subset of non-linear nodes $C_{\overline{\varepsilon}}$, as described in Section 4.1.2. From a practical point of view, this reduction of the BPT does not intrinsically modify the algorithmic process proposed here. The main two difference are (*i*) the fact that the considered tree is no longer a binary one, since a node may have 0, 1 or 2 children, instead of either 0 or 2, and (*ii*) the fact that $\widehat{C}_j$ does no longer constitute a partition of E. However, the "missing" nodes necessary to recover a partition may be easily (and deterministically) retrieved by embedding $\widehat{C}_j$ in the initial BPT. For the sake of readability (and without loss of correctness), we then preferred to present the formalised algorithm on the whole BPT.

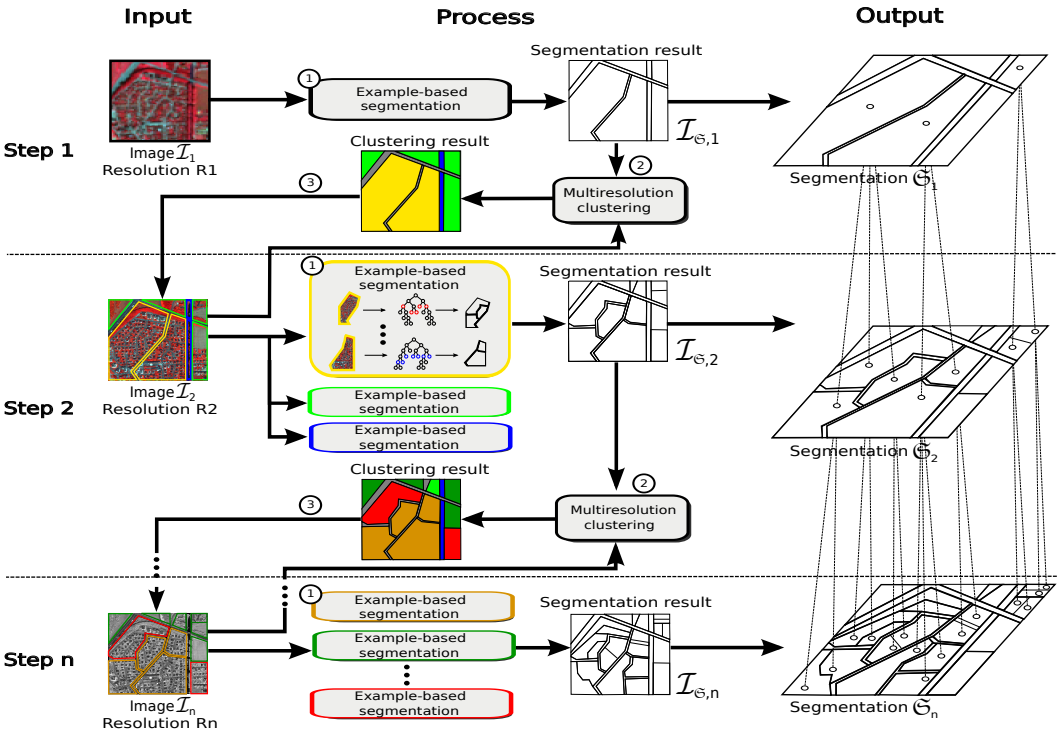


Figure 8: Work-flow overview of the multiresolution methodology (see Section 4.2). The reader may refer to Figure 3 for the description of the interactive segmentation boxes and to Figure 9 for the description of the multiresolution clustering boxes. Note that the output results (*i.e.*, the n segmentation maps) can be coupled with the clustering results.

4.2.1. Image segmentation

Thanks to the previous processing of $\mathcal{I}_{t-1}$, a clustering of $\mathcal{I}_t : E_t \to V_t$ into ω_{t-1} clusters is already available[5]. For instance, in Figure 8, a clustering of $\mathcal{I}_2$ into three (blue, green and yellow) clusters is available. These clusters enable to divide $\mathcal{I}_t$ into ω_{t-1} semantic families corresponding of the level of details of the (lower) resolution of $\mathcal{I}_{t-1}$ (Figure 8-③). The number of clusters ω_{t-1} has to be set by the user. It corresponds to the number of different urban semantic elements that the user want to extract (and refine) from the image $\mathcal{I}_{t-1}$.

Each one of these families may then be decomposed into new ones corresponding to the level of details of $\mathcal{I}_t$ (Figure 8-①). In order to do so, it is necessary to perform a segmentation of the part of the image corresponding to each one of the ω_{t-1} semantic families, *i.e.*, to segment the subimage $\mathcal{I}_{t,i} : K_i \to V_t$ of $\mathcal{I}_t$ defined on the cluster $K_i \subseteq E_t$ for any $i \in [\![1, \omega_{t-1}]\!]$ (note that the user may however choose to restrict his/her study to only certain of these families, thus leading to a partial analysis of the images).

For each considered semantic family $i \in [\![1, \omega_{t-1}]\!]$, the segmentation of $\mathcal{I}_{t,i} : K_i \to V_t$ is carried out thanks to the approach proposed in Section 4.1. Indeed, K_i can be partitioned into several (disconnected) regions, inducing several subimages of $\mathcal{I}_{t,i}$ of same resolution and semantic, and provided by the same sensor. These subimages can then conveniently be used as input for the previously described example-based image segmentation approach. The user then performs the segmentation of one of these images (see Sections 4.1.1 and 4.1.2), and this segmentation example is then automatically reproduced in all the other subimages (see Section 4.1.3).

The segmentation image $\mathcal{I}_{\mathfrak{S},t}$ obtained by gathering the ω_{t-1} segmented subimages corresponding to the ω_{t-1} semantic families constitutes the output of the step (and a partial output of the whole methodology).

[5] In the case of $\mathcal{I}_1$, we consider, without loss of generality, that there is only one cluster, the semantics of which is the one of the whole image.

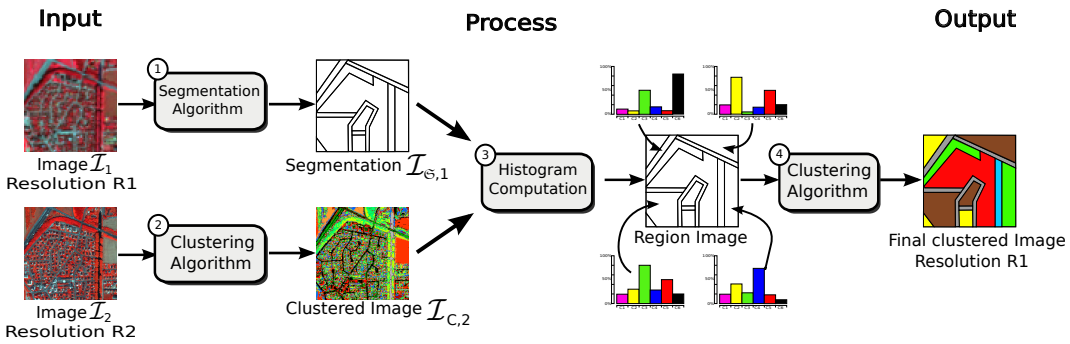


Figure 9: Multiresolution clustering approach (see Section 4.2.2).

4.2.2. (Multiresolution) image clustering

As stated above, at any step t, the segmentation of $\mathcal{I}_t$ relies on a clustering performed at step $t-1$, on the image $\mathcal{I}_{t-1}$, which provides ω_{t-1} semantic families. In order to enable to correctly perform step $t+1$, it is then necessary to perform a clustering of $\mathcal{I}_{\mathfrak{S},t}$ at the current step t (except, possibly for the last step n, where no clustering is mandatory).

This clustering relies on a multiresolution approach introduced in [43] and fully described in [40]. For a better understanding, we briefly recall this approach hereafter. The reader may also refer to Figure 9 for a visual outline.

This approach takes as input the image $\mathcal{I}_{t-1} : E_{t-1} \rightarrow V_{t-1}$, namely the image to be clustered, the segmentation $\mathcal{I}_{\mathfrak{S},t-1}$, which provides a partition $\mathfrak{S}$ of E_{t-1}, and the "next" image $\mathcal{I}_t : E_t \rightarrow V_t$. The main idea is to fuse the information provided by (1) the analysis of the "low" resolution regions of $\mathfrak{S}$ (provided by a partitioning algorithm, Figure 9-①) and (2) the "high" resolution semantic clustering of $\mathcal{I}_t$ (provided by a standard clustering method directly applied on the radiometric values of the pixels, Figure 9-②), to obtain a final clustering result corresponding to a mixed semantic level. For each region $R \in \mathfrak{S}$, a "composition" histogram is indeed computed, taking into account the distribution of the pixels of R in terms of cluster values in the highest resolution clustered image (Figure 9-③). The final clustering result is computed by classifying (in an unsupervised way) the regions of the lowest resolution segmented image using these composition histograms (Figure 9-④).

Finally, these classified segments are embedded in the next resolution, thus forming, for each resulting cluster, a new family of subimages which can be processed by following the same strategy.

5. Experimental studies

This section describes the experiments carried out with the proposed methodology to extract complex patterns from multiresolution satellite images.

The applicative context of these experimental studies is first introduced in Section 5.1. The material which were used to test the method is then presented in Section 5.2. To evaluate this methodology, our experimental protocol was based on three main criteria:

1. Efficiency: the amount of time/effort required to parametrise the methodology and to perform the segmentations (Section 5.3).
2. Accuracy: the degree of concordance between the extracted objects and the ground-truth maps provided by experts in urban planning (Section 5.4).
3. Stability: the determinism degree of the method, by studying the impact of the choice of the tree-cut example (Section 5.5).

Finally, we detail in Section 5.6 the computational complexity of this methodology and we provide runtime for the segmentations of the datasets.

5.1. Applicative context: urban mapping of complex objects

Urban planning and development organisations, disaster or environment agencies need to manage and to follow the increase of urban settlements, in particular on high risk areas. To this end, it is necessary to map these urban areas in order to store useful information (*e.g.*, urban development, natural disaster damage).

Table 1: Typologies and levels used to map urban areas at different scales.

Level	Urban areas	Urban blocks	Urban objects
Scale	1:100 000–1:25 000	1:10 000	1:5 000
Objects of interest	⋆ High-density urban fabric ⋆ Low-density urban fabric ⋆ Industrial areas ⋆ Forest zones ⋆ Agricultural zones ⋆ Water surfaces ⋆ Bare soil	⋆ Continuous urban blocks ⋆ Discontinuous urban blocks - Individual urban blocks - Collective urban blocks ⋆ Industrial urban blocks ⋆ Urban vegetation ⋆ Forest ⋆ Agricultural zones ⋆ Water surfaces ⋆ Roads	⋆ Building/roofs - Red tile roofs - Light grey residential roofs - Light commercial roofs ⋆ Vegetation - Green vegetation - Non-photosynthetic vegetation ⋆ Transportation areas - Streets - Parking lots ⋆ Water surfaces - Rivers - Natural water bodies ⋆ Bare soil ⋆ Shadows

A wide range of object typologies has been defined in order to analyse the territory at different scales. We present hereafter a hierarchical nomenclature with three semantic levels enabling to map the urban surfaces:

1. The first level of nomenclature is called the "urban areas" level (Table 1, left column). It allows to map the territory from MSR images (provided, for instance, by the SPOT constellation), with a low level of details, from 1:100 000 to 1:25 000 (enabling, for instance, to specify the density of an urban fabric). This first semantic level is used to study the large urban districts which compose the territories.
2. The second level of nomenclature is called the "urban blocks" level (Table 1, centre column). It permits to map the territory at the scale of 1:10 000 (enabling, for instance, to analyse urban blocks, defined by the minimal cycles formed by communication ways). This second semantic level has been proposed to study the urban elements which compose the urban districts. It is generally used to map the territory from HSR or/and MSR images.
3. The third level of nomenclature is called the "urban objects" level (Table 1, right column). It allows to map the urban areas at a scale of 1:5 000 enabling to deal with urban objects (*e.g.*, individual houses, gardens, roads) with their material (*e.g.*, houses with orange tile roof). Since the end of the 90's, VHSR images provide a level of details allowing to map the urban areas at this level. In the coming years, the VHSR data will become widely available thanks to the European PLEIADES program [44].

In this context, we propose to use the presented methodology to extract hierarchies of complex urban patterns (urban districts, urban blocks, urban objects) from multiresolution datasets. Such hierarchies could then be used to help urban planners to map the territories at the different levels defined previously.

5.2. Material

We describe here the datasets used for these experiments, and the software which was designed to test the methodology.

5.2.1. Images

Experiments have been performed on the urban area of Strasbourg (France) on three sites called **ELSAU**, **ILLKIRCH** and **HAUTEPIERRE** (Figure 10). These sites present typical suburban environments composed of classical objects (water surfaces, forest areas, industrial areas, individual/collective housing blocks and agricultural zones). The **ELSAU** site corresponds to a pericentral zone (6 576 m × 2 793 m) while the **ILLKIRCH** site is localised in the first ring (6 576 m × 2 803 m) in the South part of Strasbourg. The **HAUTEPIERRE** site is localised also in the first ring (3 293 m × 2 246 m) and in the Northwest part of Strasbourg (Figure 10).

To each site, a dataset is associated. Each dataset is composed of:

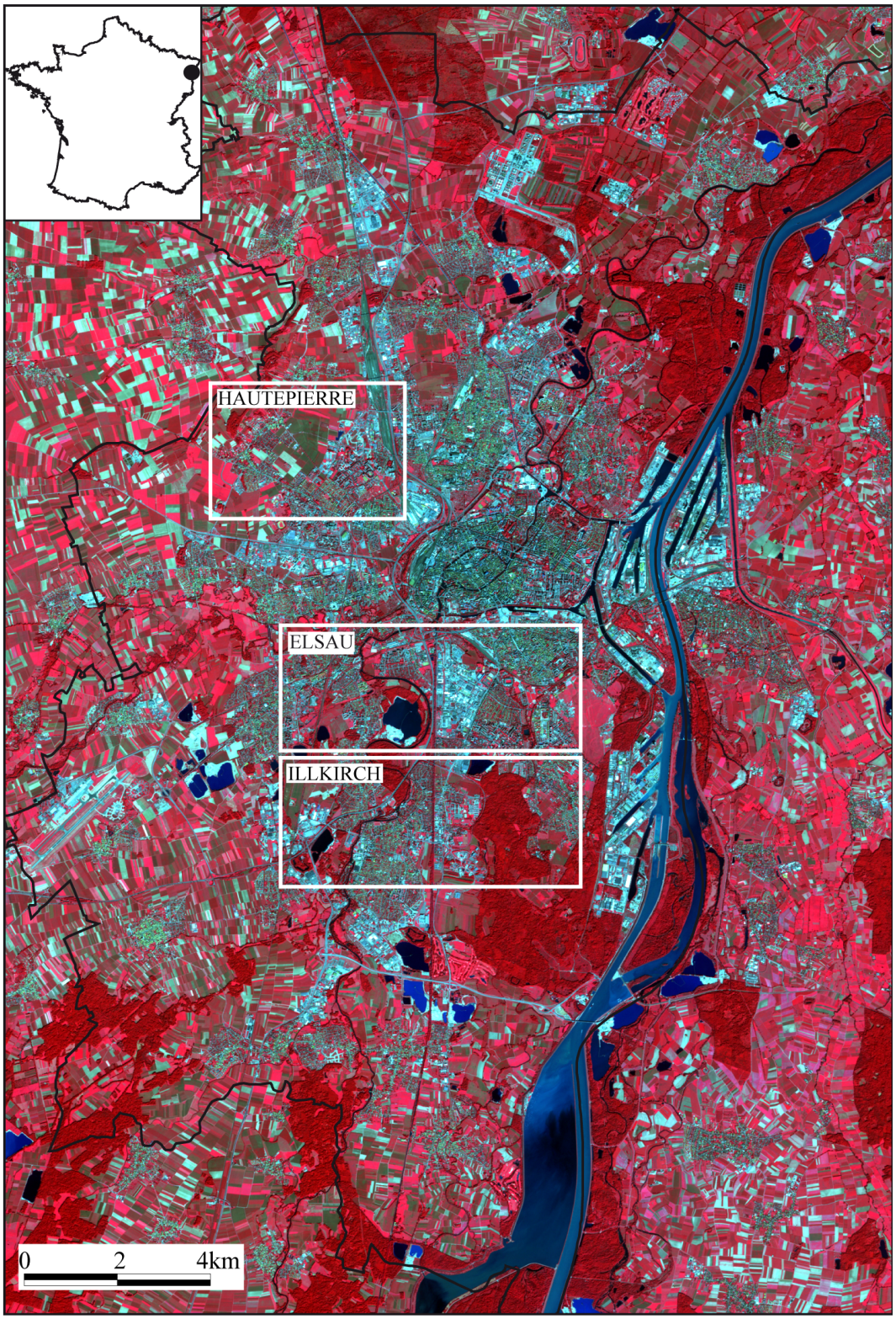


Figure 10: The urban area of Strasbourg (France) with the three study sites (**Elsau**, **Illkirch**, **Hautepierre**) localised on the Spot-5 MSR (9.6 m) multispectral image (Spot-5, © CNES), 2002.

Table 2: Satellite data used for the experiments.

Dataset	Image	Spectral resolution	Spatial resolution	Size - pixels	Surface	Size - memory
Elsau	MSR	Multispectral - 4 bands	9.6 m	685 × 291	6 576 m × 2 793 m	1.5 MB
	HSR	Multispectral - 4 bands	2.4 m	2 740 × 1 164		26 MB
	VHSR	Panchromatic - 1 band	60 cm	10 960 × 4 656		98 MB
Illkirch	MSR	Multispectral - 4 bands	9.6 m	685 × 292	6 576 m × 2 803 m	1.5 MB
	HSR	Multispectral - 4 bands	2.4 m	2 740 × 1 168		27 MB
	VHSR	Panchromatic - 1 band	60 cm	10 960 × 4 672		101 MB
Hautepierre	MSR	Multispectral - 4 bands	9.6 m	343 × 234	3 293 m × 2 246 m	600 KB
	HSR	Multispectral - 4 bands	2.4 m	1 372 × 936		3 MB
	VHSR	Panchromatic - 1 band	60 cm	5 488 × 3 744		24.5 MB

- a single Spot-5 MSR (9.6 m) multispectral image with four spectral bands (Spot-5, © CNES). The original image has a spatial resolution of 10 m. It has been resampled to a 9.6 m spatial resolution image in order to obtain a multiple of the other data;
- a couple of QuickBird images (QuickBird, © DigitalGlobe Inc.) composed by a HSR (2.4 m) multispectral image, with four spectral bands, and a VHSR (60 cm) panchromatic one.

All the images have been acquired in 2002 and for the same seasons. This guarantees the feasibility of the multiresolution approach. All the data used in these experiments are summarised in Table 2. Moreover, the images of the **Elsau** dataset are presented in Figure 12(d, e, f).

5.2.2. *Software*

In order to allow the user to actually test the proposed multiresolution methodology, a software has been designed. In this tool, each BPT can be interactively browsed, in a "threshold-like" fashion, thus enabling to easily determine the most satisfactory segmentation examples (globally, and/or by refining one or several branches).

Due to the pre-processing of the data structures, the short computation times (less than 30 seconds CPU, see Section 5.6) authorise, in particular, to carry out several segmentations to finally select the best one. We have also developed (and integrated in this software) a TIFF library which allows to load only the subdivisions of the images that are necessary to the current segmentations. This library enables to reduce the memory resources required by the application.

This tool has been implemented using the Java Mustic library and the Orfeo Toolbox (OTB) framework. Both are open source libraries and are freely downloadable[6]. It is planned to fully integrate the proposed methodology into these libraries and to distribute this software under a free licence.

5.3. *Efficiency*

To assess the efficiency of the proposed methodology, several tests have been performed to extract hierarchies of complex urban patterns from the three datasets. The process was run with three images as input ($n = 3$) in order to extract three levels of details: urban districts from the MSR images $\mathcal{I}_1$ (Step 1), urban blocks from the HSR images $\mathcal{I}_2$ (Step 2) and buildings from the VHSR images $\mathcal{I}_3$ (Step 3).

To run the proposed methodology, it is necessary to set two groups of parameters: (1) those related to the example-based segmentation approach (*e.g.*, the parametrisation of the BPT computation, the number of centroids and tree-cut examples in the learning step) and (2) those related to the multiresolution image clustering approach (*e.g.*, the number of clusters ω_t for each step $t \in [\![1, 3]\!]$). To make these experiments reproducible, we discuss and provide hereafter the details about our experimental settings.

[6]The Mustic Java library, developed by some of the authors, can be downloaded at the following url http://lsiit-cnrs.unistra.fr/fdbt-fr/index.php/Logiciels. The OTB framework is an open source set of tools for remote sensing data exploitation. It has been developed by the French Space Agency (CNES) to prepare and promote the use and the exploitation of the future images derived from the Pleiades systems [44]. It can be downloaded at the following url http://otb.cnes.fr.

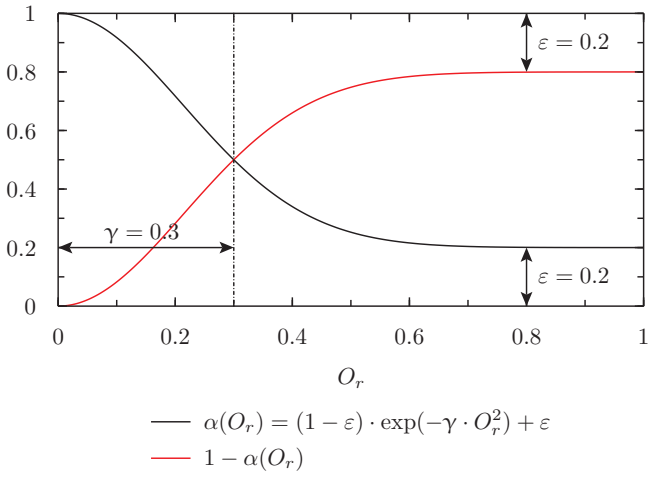


Figure 11: Representation of the α and $(1-\alpha)$ functions. $\varepsilon = 0.2$ and $\delta = 0.3$ (see Equations (11) and (12)).

5.3.1. Example-based segmentation

At each step $t \in [\![1, 3]\!]$, the example-based segmentation algorithm has been applied in each semantic class K_i (composed of k_i regions) provided by processing the level of details of the (lower) resolution $t-1$. We describe in the following the parametrisation of the three steps of this approach:

Computing the binary partition trees. The computation of the BPTs depends on two elements which are easily adjustable: one related to the elongation map computation and one related to the tuning of the α function.

As stated previously, the aim of this article is not to get the best elongation results, but to be able to compute correct elongations with a low computational cost. Thus, we have run the interactive segmentation algorithm using different elongation maps obtained by varying the growing tolerance parameter $\lambda \in [\![0, v_b^+(E) - v_b^-(E)]\!]$. For each experiment, the bounding box was computed in eight principal directions (*i.e.*, each $\pi/8$). A good balance between the computation times and the quality of the results was found when $\lambda = \frac{1}{10}\cdot(v_b^+(E) - v_b^-(E))$.

Regarding the weight $\alpha \in [0, 1]$ (which affects the similarity measure between two neighbouring regions R_i and R_j), the user may be free to tune the function proposed in Equation (5), by introducing parameters in $\alpha(O_r)$. In the present case, we have straightforwardly adapted the formulation of α as

$$\alpha(O_r) = (1-\varepsilon)\cdot\exp(-\gamma\cdot O_r^2) + \varepsilon \tag{11}$$

where γ is defined as

$$\gamma = \frac{1}{\delta^2}\cdot\ln\left(\frac{2-2\varepsilon}{1-2\varepsilon}\right) \tag{12}$$

with $\delta \in [0, 1]$ and $\varepsilon \in [0, \frac{1}{2}[$. The variable δ permits to parametrise the inversion of the trade-off between the α function (*i.e.*, the confidence into the radiometric homogeneity O_r) and the $1-\alpha$ function (*i.e.*, the confidence into the geometric homogeneity O_g). The variable ε enables to control the asymptotic behaviour of the α function. It allows to keep a minimal threshold of confidence into the radiometric homogeneity O_r. Figure 11 exemplifies the use of δ and ε. These parameters can be easily determined: a first run of the process with $\varepsilon = 0$ and $\delta = 0$ can provide a first overview of the global behaviour of the weight α. Then, the value of δ (and then ε) can be adapted to fit with user's expectations. For instance, if the processed image is not totally structured by linear elements, the parameters ε and δ can be increased. In these experiments, ε and δ were respectively set to $\varepsilon = 0.2$ and $\delta = 0.3$ (Figure 11).

Learning of the cut example. Once the k BPTs have been built, it is necessary for the user to determine one (or more) relevant tree-cut examples among the k available BPTs. In these experiments, we have studied the impact of the ratio $\eta_{ex} \in [0, 1]$ of surface (in terms of pixels) covered by the tree-cut examples provided by the user, on the quality of

the obtained segmentation/clustering results. In the case where the k regions have a similar size (in terms of pixels), this ratio can be defined as $\eta_{ex} = \frac{\#\text{tree-cut examples}}{k}$. To process, we have run the example-based segmentation approach on images at different resolutions, by varying this ratio (η_{ex} = 3%, 5%, 7%, 10%, 15%, 30%, 50%, and 70%). Each experiment has then been assessed by using the evaluation measures presented hereafter. The result of this impact study is presented in Section 5.4.

In order to be able to reproduce at best the tree-cut examples in the other remaining images/BPTs, it is necessary to learn these examples. To process, the K-MEANS clustering algorithm has been used to extract u centroids which characterise the cut examples provided by the user. We have studied the impact of the number of centroids u on the quality of the final segmentation/clustering results. To process, we have applied the example-based segmentation approach as it is explained in the following protocol. First, a BPT has been built for an image $\mathcal{I}$. From this BPT, the user has interactively selected a relevant cut C through the tree. This cut C has then been learnt with the presented learning process using different numbers of centroids ($u \in [1, 20]$). Finally, this learnt example has been used to (re)-cut the previous BPT associated to the same image $\mathcal{I}$. Each experiment (and obtained cut) has been quantitatively compared to the cut C previously provided by the user. To find relevant values of u (one per resolution), these experiments have been carried out using all the available resolutions. The result of this impact study is presented in Section 5.4.

Automatic reproduction of the cut example. The standard distance used to compare two histograms is the Euclidean one. However, this distance suffers from the problem of the *shuffling invariance* property [45] which is not desirable when comparing two histograms of ordinal type measurements.

To deal with this issue, a solution consists of using the constrained Dynamic Time Warping similarity measure [46] which enables small distortions on the radiometric axis. This similarity measure requires longer computation times than the Euclidean distance, but provides better results. We have experimentally found that the computation of the DTW similarity measure requires 10 times more operations than the computation of the Euclidean distance. The computational complexity required to compare two histograms $\mathcal{H}_{I_1}, \mathcal{H}_{I_2} : V \rightarrow \mathbb{N}$ when using the Euclidean distance is in $\Theta(\text{card}(V))$ while the complexity is in $\Theta(\text{card}(V) \times c)$ (where c measures the tolerance of distortions on the radiometric axis) when using the constraint DTW similarity measure. In the worst case, when running the climbing algorithm, the number of histograms to compare is equal to $(\mathcal{N} \times u)$. Thus, the computation of the DTW similarity measure remains tractable. Associated to this measure, an averaging method was introduced in [47].

5.3.2. (Multiresolution) image clustering

At each step $t \in [\![1, 3]\!]$, the resulting regions (obtained by gathering the k produced segmentations) have been classified in an unsupervised way, using the multiresolution method (Section 4.2.2), and then embedded in the next resolution $t + 1$ (except for the last step (Step 3) where a standard clustering has been carried out on the resulting regions).

Experiments have shown that the method did not directly find all the appropriate clusters with respect to the different urban levels. To tackle this problem, the standard solution consists of extracting a higher number of clusters ω_t than the number of thematic classes contained in the expected results. For instance, six thematic classes (at the districts level) can be extracted from the MSR image (see Table 3). Consequently, it has been chosen in agreement with the expert, to extract $\omega_1 = 10$ clusters from the MSR image. In the same way, we have set $\omega_2 = 15$ to process the HSR image, and $\omega_3 = 13$ to process the VHSR one.

5.4. Accuracy

As unsupervised evaluation techniques are inappropriate for interactive segmentation, we consider supervised evaluation. This requires the creation of a set of ground-truth maps for the evaluation.

5.4.1. Ground-truth map

To effectively measure accuracy, the ground-truth maps must be as precise as possible. Indeed, errors in these ground-truth maps may directly affect the accuracy benchmarks. Thus, results produced by the method have been assessed by quantitative comparisons with certified ground-truth maps extracted from land-cover/land-use databases. For each dataset, a set of three ground-truth maps is available (*i.e.*, one per considered level).

Table 3: Thematic classes associated to the ground-truth maps. The colours of the classes refer to those of the ground-truth maps of the **Elsau** dataset (Figure 12).

Level	Class	Colour
Districts	(1) Housing urban districts	Orange
	(2) Specific urban districts	Pink
	(3) Water surfaces	Blue
	(4) Roads	Grey
	(5) Agricultural surfaces	Yellow
	(6) Vegetation areas	Dark green
Blocks	(1) Collective housing urban blocks	Light orange
	(2) Individual housing urban blocks	Dark orange
	(3) Dense urban blocks	Red
	(4) Industrial urban blocks	Pink
	(5) Water surfaces	Blue
	(6) Roads	Grey
	(7) Agricultural zones	Yellow
	(8) Urban vegetation	Light green
	(9) Forest areas	Dark green
Objects	(1) Buildings	Red

Figure 12 presents the ground-truth maps associated to the **Elsau** dataset. The first ground-truth map contains six thematic classes at the semantic district level (Figure 12(a)) This map has been produced by the expert with a visual interpretation of the MSR image. The second ground-truth map contains nine thematic classes at the semantic block level (Figure 12(b)) and is extracted from a regional database © BDOCS – CIGAL2002. The last certified map is used to evaluate the extraction of the urban objects. This map contains only one thematic class corresponding to the buildings (Figure 12(c)). This class has been extracted from a building database produced by the national cartographic agency © IGN, 2002. All the thematic classes associated to the different ground-truth maps are summarised in Table 3.

5.4.2. Results evaluation

In order to compare the obtained segmentation/clustering results to land cover reference maps, we have computed local and global evaluation indexes (Table 4).

Local evaluation indexes enable to independently assess the extraction of each thematic class. To process, for each thematic class, the best corresponding clusters (in terms of partitions) were extracted. Then, we have computed:

1. the percentage of false positive denoted by $f^{(p)}$;
2. the percentage of false negative denoted by $f^{(n)}$;
3. the percentage of true positive denoted by $t^{(p)}$.

These measures are used to estimate the precision $\mathcal{P}$ and the recall $\mathcal{R}$ of the results obtained by using the proposed method:

$$\mathcal{P} = \frac{t^{(p)}}{t^{(p)} + f^{(p)}} \qquad \text{and} \qquad \mathcal{R} = \frac{t^{(p)}}{t^{(p)} + f^{(n)}} \tag{13}$$

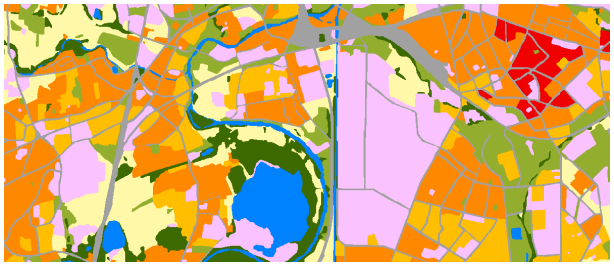
We have also computed the standard F-measure $\mathcal{F}$ which is the harmonic mean of precision and recall:

$$\mathcal{F} = 2 \cdot \frac{\mathcal{P} \cdot \mathcal{R}}{\mathcal{P} + \mathcal{R}} \tag{14}$$

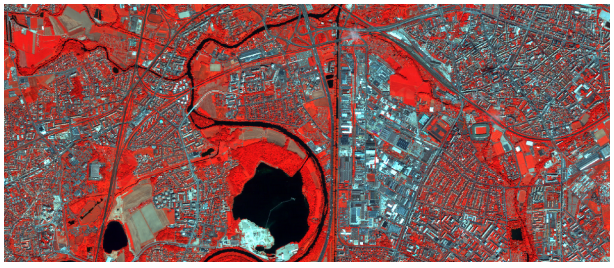
To assess the relevance of the results, we also provide global classification accuracy indexes. For each experiment, we have computed the weighted harmonic mean $\overline{\mathcal{F}}$ of the F-measures (weighted by the cardinals of the thematic

(a) Districts level (6 semantic classes).

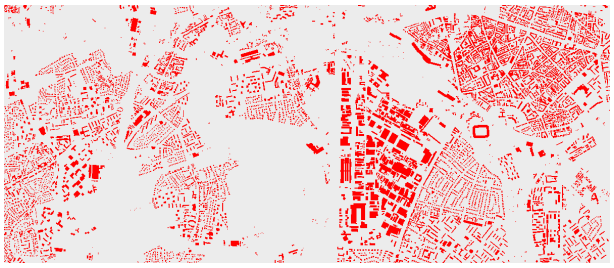
(d) MSR image - 685 × 291 pixels - four spectral bands.



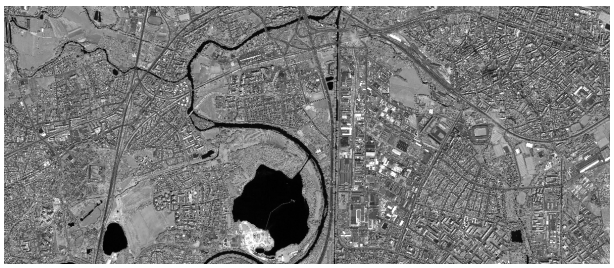
(b) Blocks level (9 semantic classes).



(e) HSR image - 2 740 × 1 164 pixels - four spectral bands.



(c) Objects level (one semantic class).



(f) VHSR image - 10 960 × 4 656 pixels - one spectral band.

Figure 12: Ground-truth maps (a, b, c) associated to the satellite images of the **Elsau** dataset (d, e, f). (a) Districts level. (b) Blocks level. (c) Objects level. (d) MSR image (Spot-5, © CNES), 1 pixel = 9.6 m × 9.6 m. (e) HSR image (QuickBird, © DigitalGlobe Inc.), 1 pixel = 2.4 m × 2.4 m. (f) VHSR image (QuickBird, © DigitalGlobe Inc.), 1 pixel = 60 cm × 60 cm. See Table 3 for more details about the colours and the semantics associated to the different classes of the ground-truth maps.

classes), and the Kappa index [48]. The Kappa index $\mathcal{K}$, which is a measure of global classification accuracy, is defined as:

$$\mathcal{K} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{15}$$

where $\Pr(a)$ is the relative agreement among the observers, and $\Pr(e)$ is the hypothetical probability of chance agreement. The Kappa index takes value in $[0, 1]$ and decreases as the classification is in disagreement with the ground-truth map. We have computed this index as follows. The approach consists of considering all point couples $(\mathbf{x_1}, \mathbf{x_2}) = ((x_1, y_1), (x_2, y_2))$ and see the configuration of these two points in each partition (the clustering result and the ground-truth). There are four possible configurations; for each one, a counter is associated and incremented each time a configuration appears:

1. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition both in the clustering and in the reference map (counter s_s);
2. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition in the clustering but not in the reference map (counter s_d);
3. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition in the reference map but not in the clustering (counter d_s);
4. $\mathbf{x_1}$ and $\mathbf{x_2}$ belong to the same partition neither in the reference map nor in the clustering (counter d_d).

Table 4: Evaluation measures.

Symbol	Evaluation measure	Type
$\mathcal{P}$	Precision index	Per-class accuracy
$\mathcal{R}$	Recall index	
$\mathcal{F}$	F-measure	
$\mathcal{K}$	Kappa index	Global accuracy
$\overline{\mathcal{F}}$	Weighted harmonic mean of $\mathcal{F}$	

Thus, the Kappa index is computed with:

$$\Pr(a) = \frac{s_s + d_d}{s_s + s_d + d_s + d_d} \tag{16}$$

and

$$\Pr(e) = \frac{(s_s + s_d) \cdot (s_s + d_s) + (s_d + d_d) \cdot (d_s + d_d)}{(s_s + s_d + d_s + d_d)^2} \tag{17}$$

5.4.3. Results analysis

These different indexes have been used to evaluate the segmentation/clustering results provided by this methodology. First, we detail an impact-study of the proportion η_{ex} of tree-cut examples provided by the user. Second, we discuss about the impact of the number u of centroids, on the quality of the results. Finally, we present a global evaluation of the relevance of the results and we compare them to some results provided by other related methods (on other data).

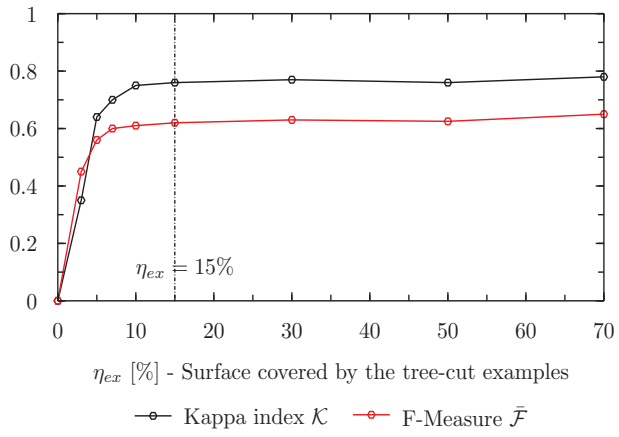
Impact of the proportion η_{ex} of tree-cut examples. We discuss hereafter the impact of the proportion of tree-cut examples provided by the user on the example-based segmentation approach. The results of this impact-study are presented in Figures 13(a, b, c).

From these graphs, one can see that from 0% to 15% of tree-cut examples provided by the user, the quality and the relevance of the u centroids (extracted during the learning step) increases. The automatic reproduction step (and in particular, the climbing algorithm) will be directly affected by the relevance of these u centroids, which model the scale examples. After 15% of examples provided by the user, the quality of these centroids does not increase anymore and remains constant (in terms of Kappa index $\mathcal{K}$ and F-Measure $\overline{\mathcal{F}}$).

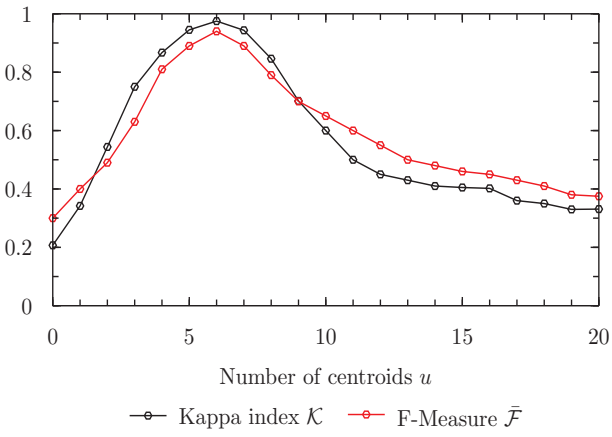
Nevertheless, providing too much tree-cut examples can be time-consuming for the user. Experiments have shown that $\eta_{ex} = 5\%$ is a good balance between time-consuming and the accuracy of the segmentation results of the MSR images (Figure 13(a)). Providing 5% of examples among the k regions to segment (2 examples in average per cluster) enables to obtain an effective automatic reproduction step without costing to much time for the user (30 seconds per example). When processing HSR images, $\eta_{ex} = 10\%$ seems sufficient to obtain relevant results (Figure 13(b)) while when processing VHSR images, η_{ex} can be set to $\eta_{ex} = 7\%$ (Figure 13(c)). Moreover, the user is free to modify η_{ex} if the accuracy of the extracted patterns does not fit with his/her current application.

Impact of the number u of centroids. Experiments have shown that the number u of centroids has an actual importance on the quality of the extracted patterns.

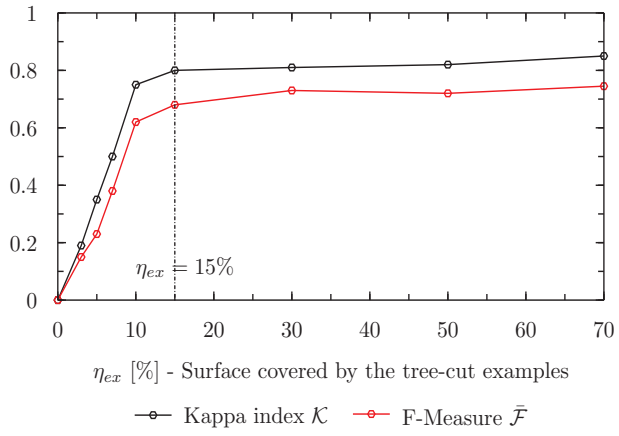
Figure 13(d, e, f) presents the evolution of the $\mathcal{K}$ and $\overline{\mathcal{F}}$ indexes by varying the value of u when processing images with different spatial resolutions. The best results have been obtained when the number of centroids u was set to $u = 7$ for the MSR images (Figure 13(d)), $u = 10$ for the HSR ones (Figure 13(e)) and $u = 6$ for the VHSR ones (Figure 13(f)). Experiments have shown that when processing MSR and VHSR images, it is required to use a lower number of centroids u than when processing HSR ones. This behaviour can be justified by assuming that the number of materials separable using HSR images is higher than when using MSR images (because of the lower spatial resolution) and VHSR ones (because of the lower spectral resolution). Furthermore, the number u of centroids can



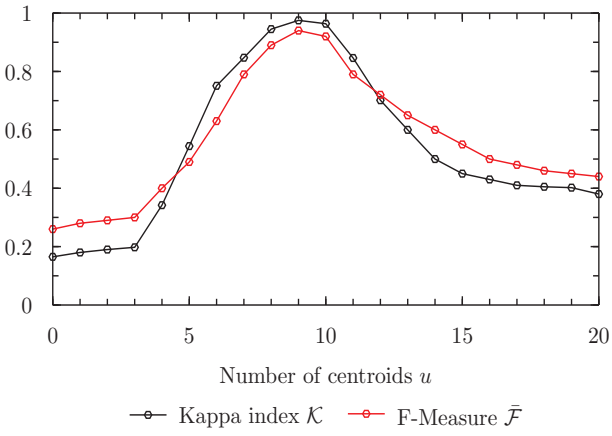
(a) MSR images.



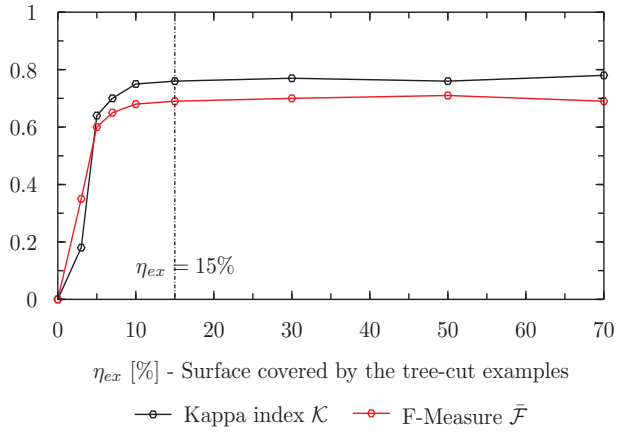
(d) MSR images.



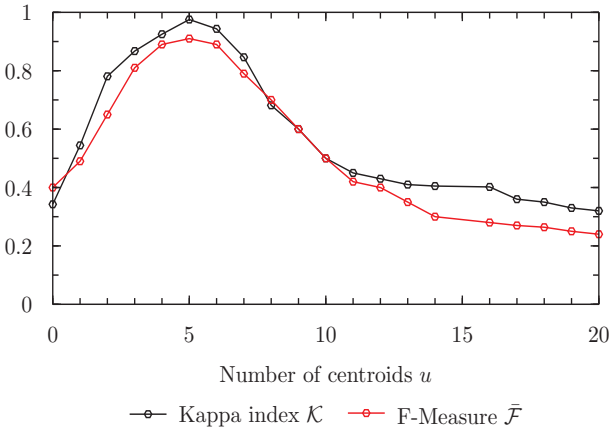
(b) HSR images.



(e) HSR images.



(c) VHSR images.



(f) VHSR images.

Figure 13: Results of the different impact studies. (a-c) Impact of the proportion η_{ex} of tree-cut examples provided by the user on the accuracy of the final clustering results. (d-f) Impact of the number u of centroids on the accuracy of the final clustering results. The evolution of the $\mathcal{K}$ index (resp. the $\overline{\mathcal{F}}$ index) is depicted in black (resp. in red). (a, d) MSR images. (b ,e) HSR images. (c, f) VHSR images.

Table 5: Results evaluation of the **Elsau**, **Illkirch**, **Hautepierre** datasets. The symbol (•) means that the thematic class is not available in the current dataset.

		Elsau					**Illkirch**					**Hautepierre**				
Level	Class	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{K}$	$\overline{\mathcal{F}}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{K}$	$\overline{\mathcal{F}}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{K}$	$\overline{\mathcal{F}}$
Districts	(1) Housing urban districts	0.72	0.77	0.74	0.68	0.56	0.59	0.80	0.68	0.74	0.59	0.61	0.86	0.70	0.72	0.67
	(2) Specific urban districts	0.56	0.47	0.51			0.43	0.54	0.48			0.42	0.59	0.49		
	(3) Water surfaces	0.69	0.91	0.78			0.39	0.91	0.54			•	•	•		
	(4) Roads	0.05	0.29	0.08			0.10	0.42	0.16			0.52	0.60	0.56		
	(5) Agricultural surfaces	0.67	0.73	0.70			0.70	0.45	0.55			0.89	0.83	0.86		
	(6) Vegetation areas	0.26	0.72	0.37			0.64	0.83	0.72			0.37	0.62	0.46		
Blocks	(1) Collective housing urban blocks	0.66	0.61	0.62	0.78	0.69	0.60	0.67	0.64	0.77	0.67	0.55	0.62	0.58	0.75	0.64
	(2) Individual housing urban blocks	0.72	0.75	0.74			0.68	0.42	0.52			0.57	0.64	0.60		
	(3) Dense urban blocks	0.64	0.63	0.64			•	•	•			•	•	•		
	(4) Industrial urban blocks	0.55	0.52	0.53			0.65	0.72	0.68			0.59	0.42	0.49		
	(5) Water surfaces	0.75	0.84	0.79			0.78	0.87	0.82			•	•	•		
	(6) Roads	0.31	0.28	0.29			0.33	0.31	0.32			0.28	0.54	0.37		
	(7) Agricultural zones	0.71	0.76	0.72			0.68	0.75	0.71			0.55	0.49	0.52		
	(8) Urban vegetation	0.21	0.63	0.31			0.35	0.77	0.48			0.12	0.27	0.17		
	(9) Forest areas	0.25	0.49	0.33			0.46	0.53	0.48			0.35	0.62	0.45		
Objects	(1) Buildings	0.83	0.64	0.72	0.76	0.72	0.82	0.72	0.77	0.75	0.77	0.74	0.63	0.68	0.69	0.68

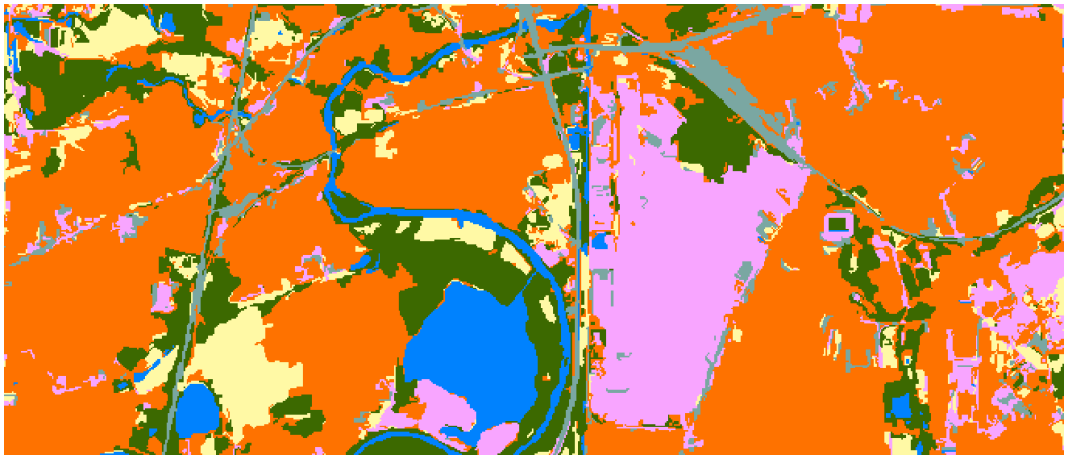
be adapted by the expert which knows, in advance, how many kinds of different structures could compose the current refined regions.

Global evaluation of the results. We have then set the different parameters to the values discussed previously and run the presented methodology on the three datasets. We present hereafter the best segmentation/clustering results which enable to globally assess the relevance of the method. Table 5 summarises the evaluation scores obtained.

Step 1 has been applied on the MSR images to separate the largest structures of the scene (*e.g.*, urban districts, forest areas, water surfaces). The obtained segmentation/clustering results are presented in Figures 14(a), 15(a), 16(a). After classification, the comparisons between the classified resulting regions and the ground-truth maps have shown Kappa values and F-Measures of ($\mathcal{K} = 0.68$ and $\overline{\mathcal{F}} = 0.56$) for the **Elsau** dataset, ($\mathcal{K} = 0.74$ and $\overline{\mathcal{F}} = 0.59$) for the **Illkirch** dataset, ($\mathcal{K} = 0.72$ and $\overline{\mathcal{F}} = 0.67$) for the **Hautepierre** dataset. The best extracted thematic classes are those of housing urban districts (1), specific urban districts (2) and agricultural surfaces (5). However, the methodology does not extract directly, at this scale, the road class (4). Indeed, the values of F-Measure associated to this class are low in two datasets ($\mathcal{F} = 0.08$ for the **Elsau** dataset and $\mathcal{F} = 0.16$ for the **Illkirch** dataset).

Step 2 has been applied on the HSR images to split the urban districts (extracted at the previous step) into different large regions corresponding to: mixed urban districts, commercial or industrial sub-districts, housing blocks, *etc*. The obtained segmentation/clustering results are presented in Figures 14(b), 15(b), 16(b). The comparisons between the classified resulting regions and the ground-truth maps have shown Kappa values and F-Measures of ($\mathcal{K} = 0.78$ and $\overline{\mathcal{F}} = 0.69$) for the **Elsau** dataset, ($\mathcal{K} = 0.77$ and $\overline{\mathcal{F}} = 0.67$) for the **Illkirch** dataset, ($\mathcal{K} = 0.75$ and $\overline{\mathcal{F}} = 0.64$) for the **Hautepierre** dataset. The classes related to the housing and industrial urban blocks (classes (1–4)) are fairly well extracted. The values of F-Measure associated to these classes are close to 0.60 in all the datasets. However, one can observe that some of the partition results are composed of several regions matching with urban patterns and numerous tiny regions forming linear structures and covering vegetation areas. These over-segmentation problems are probably due to the spatial criteria used by the algorithm (the elongation one) which does not consider the vegetation areas.

Step 3 has then been performed on the VHSR images to extract "basic" urban objects (*e.g.*, individual/collective houses, vegetations, streets/car parks, shadows) from the urban blocks extracted previously. The obtained segmentation/clustering results are presented in Figures 14(c), 15(c), 16(c). Due to the unavailability of all the "class" information for the ground-truth maps corresponding to the VHSR images, we have only assessed the building class. The comparisons between the classified resulting regions and the ground-truth maps have shown Kappa values and F-Measures of ($\mathcal{K} = 0.76$ and $\mathcal{F} = 0.72$) for the **Elsau** dataset, ($\mathcal{K} = 0.75$ and $\mathcal{F} = 0.77$) for the **Illkirch** dataset, ($\mathcal{K} = 0.69$ and $\mathcal{F} = 0.68$) for the **Hautepierre** dataset. Moreover, visual validations have shown that the other kinds

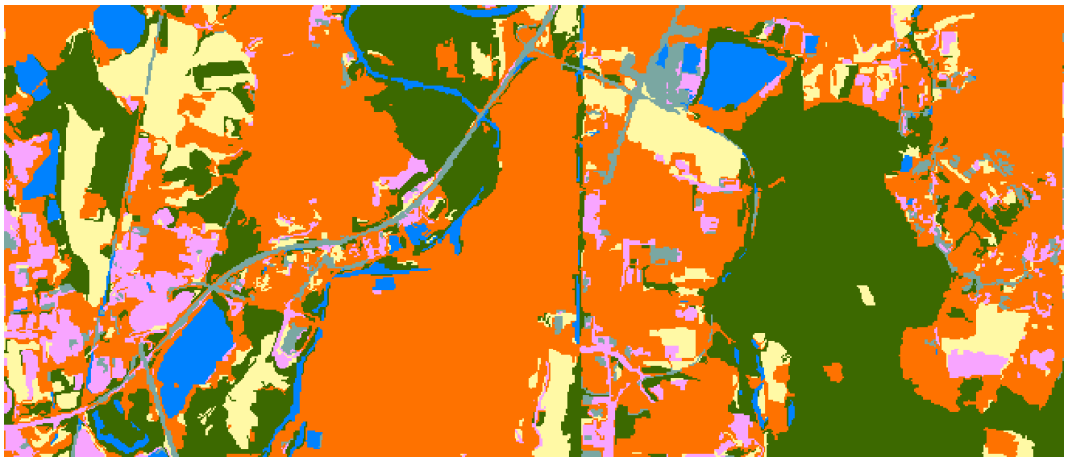


(a) MSR segmentation - districts level.

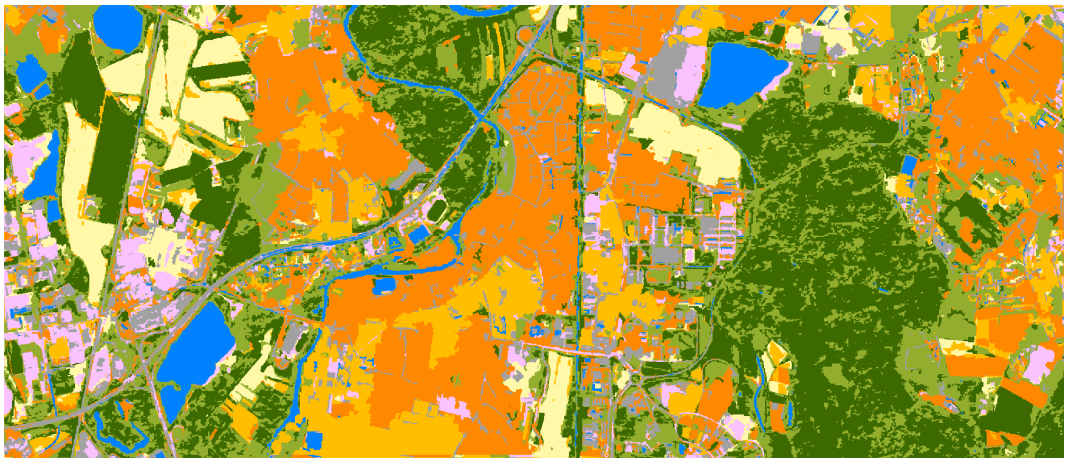
(b) HSR segmentation - blocks level.

(c) VHSR segmentation - objects level.

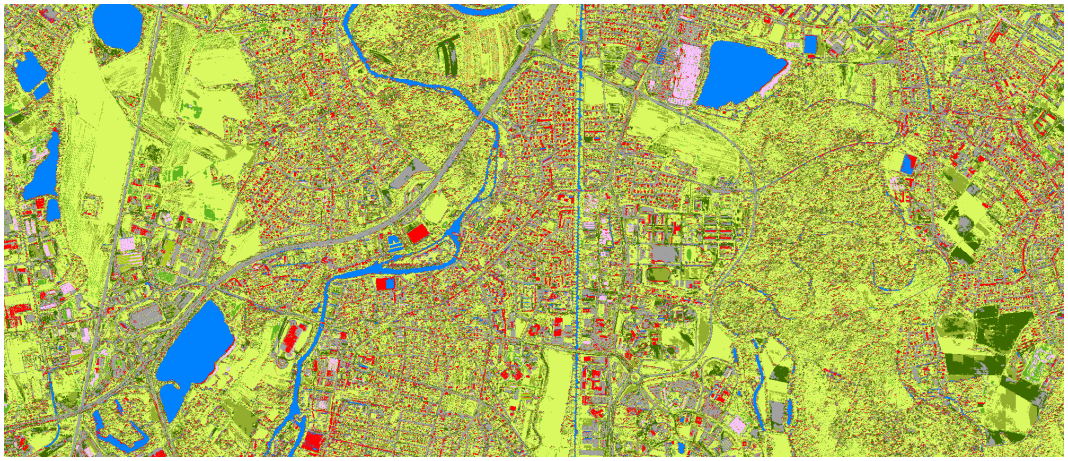
Figure 14: Segmentation results of the **Elsau** dataset (6 576 m × 2 793 m). The colours of the clusters were chosen to correspond to those defined for the groundtruth maps. (a) MSR segmentation, districts level. (b) HSR segmentation, blocks level. (c) VHSR segmentation, objects level.



(a) MSR segmentation - districts level.

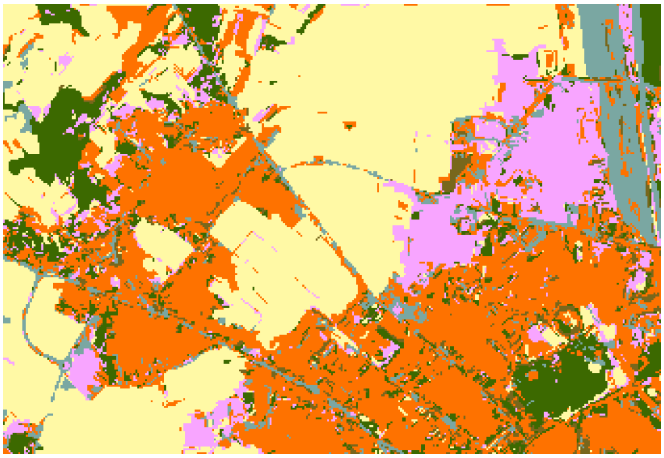


(b) HSR segmentation - blocks level.

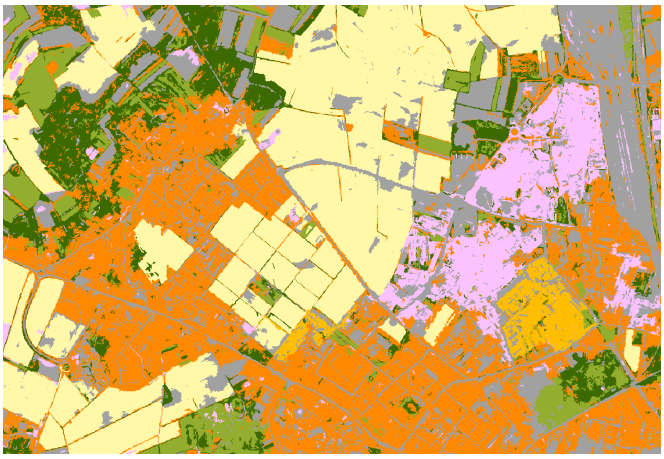


(c) VHSR segmentation - objects level.

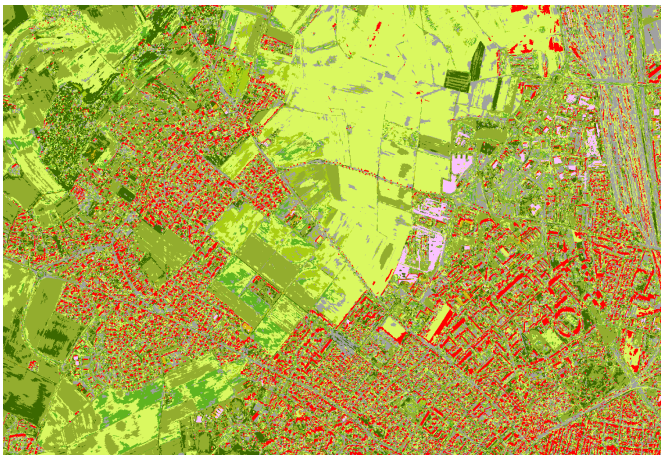
Figure 15: Segmentation results of the **Illkirch** dataset (6 576 m × 2 803 m). The colours of the clusters were chosen to correspond to those defined for the groundtruth maps. (a) MSR segmentation, districts level. (b) HSR segmentation, blocks level. (c) VHSR segmentation, objects level.



(a) MSR segmentation - districts level.



(b) HSR segmentation - blocks level.



(c) VHSR segmentation - objects level.

Figure 16: Segmentation results of the **Hautepierre** dataset (3 293 m × 2 246 m). The colours of the clusters were chosen to correspond to those defined for the groundtruth maps. (a) MSR segmentation, districts level. (b) HSR segmentation, blocks level. (c) VHSR segmentation, objects level.

of "basic" urban objects in the urban blocks are well extracted even if the boundaries are not perfect.

For each of these segmentations, one may note the progressive refinement of the initial partitions which corresponds to additional information provided by the higher resolutions.

Comparison with other related methods. Most of the methods proposed in the literature rely on spectral homogeneity properties [49, 50] (sometimes also using hierarchical multiscale approaches [7]). Such methods are efficient to extract/classify basic urban objects, but may fail to detect larger and/or more complex patterns, which is the main strength of the proposed methodology. Then, we chose to compare separately the efficiency of the extraction of basic objects (*e.g.*, building/roofs, streets, shadows) and the efficiency of the extraction of more complex ones (*e.g.*, urban blocks, urban districts).

Moreover, the methods to which the proposed one could be compared, and in particular those considered here (namely, [49, 7, 50] for the extraction of basic objects, and [6] for the extraction of complex objects) can not be directly compared on the datasets described above, especially due to the unavailability of the corresponding softwares and the difficulty to correctly re-implement them by only using the publication details. Consequently, it has been chosen to compare these methods with the proposed one based on the accuracy scores provided in the related articles.

The evaluation of the results obtained on the three considered datasets has shown that the percentages of basic urban objects well recognized/extracted are comparable to those obtained in [50] where six different approaches (dedicated to extract basic urban objects) are presented. Indeed, the proposed multiresolution methodology enables to extract, in average, 73% of the basic urban objects of interest ($\overline{\mathcal{K}} = 0.73$ and $\overline{\mathcal{F}} = 0.72$) while the percentages of individual buildings well extracted reaches 78% in [50] where similar accuracy measures are used. This difference of accuracy can be explained by the fact that the six approaches presented in this article have been designed for *ad hoc* purposes, and require a significant amount of supervision (ranging from setting several thresholds to manually positioning markers on buildings). We have also compared the accuracy of the proposed method to another hierarchical one [7]. This approach has been applied to extract different kinds of basic urban objects (*e.g.* roads, buildings, urban vegetations). The scores obtained for the urban buildings class are ($\overline{\mathcal{P}} = 0.69$ and $\overline{\mathcal{R}} = 0.77$) which is also comparable to the ones obtained with the proposed methodology ($\overline{\mathcal{P}} = 0.79$ and $\overline{\mathcal{R}} = 0.66$).

Concerning the extraction of complex urban structures such as urban blocks and urban districts, we have compared the obtained results to those presented in [6], where a bottom-up segmentation approach dealing with multiresolution images is proposed. This approach has been applied to extract complex urban objects (*e.g.*, large urban blocks, parkings, residential areas, *etc.*) from couples of HSR/VHSR Ikonos images. The scores of the results obtained using this methodology reach $\overline{\mathcal{K}} = 0.73$ which is also similar to the ones obtained with the presented methodology ($\overline{\mathcal{K}} = 0.77$ and $\overline{\mathcal{F}} = 0.67$ for the urban blocks, and $\overline{\mathcal{K}} = 0.71$ and $\overline{\mathcal{F}} = 0.61$ for the urban districts).

To conclude on this study, one can note that the proposed multiresolution segmentation methodology provides comparable results as several other related approaches. In addition, this methodology enables to extract different levels of objects of interest while most of the other related ones enable only the extraction at a single semantic level.

5.5. Stability evaluation

To evaluate the degree of determinism of the proposed methodology and, in particular, the determinism of the example-based segmentation approach, we have applied the following protocol.

The example-based segmentation approach has been run k times by varying the tree-cut examples provided by the user. To process, for a specific family of k sub-images to segment ($k = 20$ in the current experiment), we have exhaustively chosen one BPT among the k BPTs to cut. This BPT was then proposed to the user for the interactive tree-cut operation. For each experiment, the cut obtained was used to learn the segmentation process. After the automatic reproduction step, the result was evaluated and the process has been repeated using another example among the k possible ones.

Results obtained are presented in Figure 17. From this graph, one can see that the impact of the choice of the region which is used as tree-cut example does not have a strong importance. Indeed, the values of the Kappa index and of the F-Measure index remain similar through the choice of the tree-cut examples ($\mathcal{K} = 0.79 \pm 0.02$, $\overline{\mathcal{F}} = 0.70 \pm 0.04$). This behaviour can be explained by the similar properties of the k sub-images to segment which have been previously gathered into the same cluster by the multiresolution clustering step. Choosing one or another of them to make the tree-cut example does not actually affect the determinism of the proposed methodology.

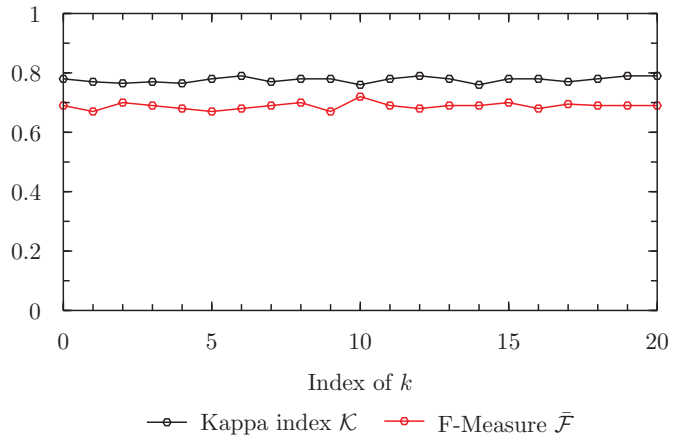


Figure 17: Impact of the choice of the tree-cut example among the $k = 20$ possibilities. The evolution of the $\mathcal{K}$ index (resp. the $\overline{\mathcal{F}}$ index) is depicted in black (resp. in red).

Table 6: Runtime and memory usage for the segmentation of the images contained in the studied datasets.

Dataset	Image (Size - pixels)	Extract Runtime	Memory (RAM)
Elsau	MSR (685 × 291)	46.2 s	67 MB
	HSR (2 740 × 1 164)	5 min 27 s	569 MB
	VHSR (10 960 × 4 656)	1 h 24 min	3.51 GB
Illkirch	MSR (685 × 292)	56.1 s	69 MB
	HSR (2 740 × 1 168)	5 min 48 s	583 MB
	VHSR (10 960 × 4 672)	1 h 36 min	3.59 GB
Hautepierre	MSR (343 × 234)	33.1 s	54 MB
	HSR (1 372 × 936)	4 min 12 s	431 MB
	VHSR (5 488 × 3 744)	53 min	2.81 GB

5.6. Computation time study

This section details the computational complexity and the runtime of the proposed methodology. As it is not possible to provide a theoretical complexity study of the proposed methodology, we present hereafter an experimental evaluation of the complexity. We also provide a runtime comparison between the computation of a classical BPT, the application of the top-down strategy on the **Elsau** dataset and the application of a similar top-down multiresolution segmentation approach [38] (on data different from the ones presented in this article).

Experimental complexity study. Table 6 provides the runtime and the memory usages for the segmentation of the images contained in the three studied datasets. Experiments have been run on an Intel® Core™2 Quad running at 2.4 GHz with 8 GB of RAM.

As shown by the third column of Table 6, the extraction process is linear with the size of the images. For instance, a HSR image which contains 16 times more pixels than a MSR one, requires 16 times more operations and time to be processed than a MSR one. Since the multiresolution clustering approach (which is mainly based on a partitioning clustering) is linear through the data, we can assume that the example-based segmentation approach is also linear through the data.

However, one can see that the memory consumption remains significant when processing the VHSR images (fourth column of Table 6).

Table 7: Memory and time comparison between a classical BPT computation and the top-down strategy. The images are those of the **Elsau** dataset.

Dataset	Image (size)	Classical BPT			Top-down approach		
		Extract Runtime	BPT size (Number of nodes)	Memory (RAM)	Extract runtime	BPT size (Number of nodes)	Memory (RAM)
Elsau	MSR (685 × 291)	46.2 s	3.98×10^4	67 MB	46.2 s	3.98×10^4	67 MB
	HSR (2 740 × 1 164)	3 min 49 s	9.12×10^6	1.42 GB	5 min 27 s	4.49×10^5	569 MB
	VHSR (10 960 × 4 656)	**n/a**	**n/a**	**n/a**	1 h 24 min	2.94×10^6	3.51 GB

Runtime comparisons. Table 7 emphasises the advantage of using a multiresolution top–down model: while a classical BPT cannot handle the whole VHSR images due to memory limitations, the top-down approach can process any large images with a tractable amount of memory. Furthermore, the extract runtimes are comparable to those obtained in [38] where a construction scheme for irregular pyramids is presented to segment multiresolution histological images using a top-down strategy. For instance, this methodology required 44 min to 1 h 48 min to process large multiresolution data composed of different images with increasing spatial resolutions (*e.g.*, 625 × 625, 2 500 × 2 500, and 10 000 × 10 000 pixels).

The size of a BPT can be estimated as: $\#nodes = n + \frac{n}{2} + \frac{n}{4} + \frac{n}{8} + \cdots + 2 + 1 = 2n - 1$ where n denotes the number of leaves in the considered BPT. The maximum size of a BPT is obtained when each pixel corresponds to a leaf node. Thus, processing the VHSR image of the **Elsau** dataset (10 960 × 4 656 pixels) by using the classical BPT approach would require up to 1.02×10^8 nodes which is hardly tractable with the current computer configurations. By using our methodology, this memory issue can be addressed.

6. Conclusion and perspectives

This article has presented an interactive hierarchical segmentation methodology which is dedicated to extract complex and/or structured objects from remote sensing images. This methodology, mainly based on binary partition trees and multiresolution clustering, combines two principal contributions. Our first contribution is a top-down multiresolution approach. This multiresolution approach enables to deal with VHSR images without being convoluted by the large size and the level of details of this data. Our second contribution is an example-based segmentation approach. This partitioning approach authorises to adapt the segmentation process (and/or the segmentation parameters) to restricted areas of homogeneous classes of radiometric intensity instead of segmenting a whole image using only one segmentation parameter (*e.g.*, only one scale parameter).

This methodology has been applied on three datasets of multiresolution satellite optical images to extract hierarchies of complex urban patterns. The results obtained with this methodology have shown that the quality and the accuracy of the extracted patterns seems sufficient to further accurately perform both classification or object detection. This seems to validate (1) the relevance of the proposed method and (2) the soundness of the semi-automation of the photo-interpretation approach.

This work opens up several perspectives and different research directions. From a methodological point of view, the top-down hierarchical segmentation approach may be enhanced by correcting the borders of the extracted regions which remain affected by the resolution gaps. The classification/clustering step could also be enhanced by using a bottom-up approach. Indeed, once the current methodology terminated (at the level t), an ascendant climbing approach through the resolution, may enable to correct the clustering results of the regions at the level $t-1$. From a geographical point of view, it is also planned to enrich the proposed methodology by introducing the knowledge of the geographer expert in the segmentation and in the clustering process. We believe that exploiting and extending this property is a promising research direction.

References

[1] D. N. M. Donoghue, Remote sensing: sensors and applications, Progress in Physical Geography 24 (3) (2000) 407–414.

[2] M. Baatz, C. Hoffmann, G. Willhauck, Progressing from object-based to object-oriented image analysis, in: T. Blaschke, S. Lang, G. J. Hay (Eds.), Object-Based Image Analysis, Lecture Notes in Geoinformation and Cartography, Springer, 2008, Ch. 1.2, pp. 29–42.

[3] T. Blaschke, Object based image analysis for remote sensing, ISPRS Journal of Photogrammetry and Remote Sensing 65 (1) (2010) 2–16.
[4] A. Puissant, C. Weber, The utility of Very High Spatial Resolution images to identify urban objects, Geocarto International 17 (1) (2002) 33–44.
[5] G. Forestier, C. Wemmert, P. Gançarski, Multisource images analysis using collaborative clustering, EURASIP Journal on Advances in Signal Processing - Special issue on Machine Learning in Image Processing 2008 (1) (2008) 1–11.
[6] R. Gaetano, G. Scarpa, G. Poggi, Hierarchical texture-based segmentation of multiresolution remote-sensing images, IEEE Transactions on Geoscience and Remote Sensing 47 (7) (2009) 2129–2141.
[7] H. G. Akcay, S. Aksoy, Automatic detection of geospatial objects using multiple hierarchical segmentations, IEEE Transactions on Geoscience and Remote Sensing 46 (7) (2008) 2097–2111.
[8] W. Sun, V. Heidt, P. Gong, G. Xu, Information fusion for rural land-use classification with high-resolution satellite imagery, IEEE Transactions on Geoscience and Remote Sensing 41 (4) (2003) 883–890.
[9] M. J. Barnsley, S. L. Barr, Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system, Computers, Environment and Urban Systems 21 (3) (1997) 209–225.
[10] C. Kurtz, N. Passat, A. Puissant, P. Gançarski, Hierarchical segmentation of multiresolution remote sensing images, in: Proceedings of the 10th International Symposium on Mathematical Morphology - ISMM'11, Vol. 6671 of Lecture Notes in Computer Science, Springer, 2011, pp. 343–354.
[11] L. Guigues, H. Le Men, J. P. Cocquerez, The hierarchy of the cocoons of a graph and its application to image segmentation, Pattern Recognition Letters 24 (8) (2003) 1059–1066.
[12] M. Pietikainen, A. Rosenfeld, Image segmentation by texture using pyramid node linking, IEEE Transactions on Systems, Man and Cybernetics 11 (12) (1981) 822–825.
[13] M. Pesaresi, J. A. Benediktsson, A new approach for the morphological segmentation of high-resolution satellite imagery, IEEE Transactions on Geoscience and Remote Sensing 39 (2) (2001) 309–320.
[14] J. Inglada, J. Michel, Qualitative spatial reasoning for high-resolution remote sensing image analysis, IEEE Transactions on Geoscience and Remote Sensing 47 (2) (2009) 599–612.
[15] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 888–905.
[16] R. Goffe, G. Damiand, L. Brun, A causal extraction scheme in top-down pyramids for large images segmentation, in: International Workshop On Structural and Syntactic Pattern Recognition, Vol. 6218 of Lecture Notes in Computer Science, Springer, 2010, pp. 264–274.
[17] J. C. Tilton, Analysis of hierarchically related image segmentations, in: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, Vol. 2, 2003, pp. 60–69.
[18] M. Baatz, A. Schäpe, Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation, in: W. Verlag (Ed.), Angewandte Geographische Informations-Verarbeitung XII, Vol. 58 of Karlsruhe, 2000, pp. 12–23.
[19] J. M. Beaulieu, M. Goldberg, Hierarchy in picture segmentation: a stepwise optimization approach, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (2) (1989) 150–163.
[20] G. Scarpa, M. Haindl, J. Zerubia, A hierarchical finite-state model for texture segmentation, in: IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, 2007, pp. 1209–1212.
[21] J. C. Serra, P. Salembier, Connected operators and pyramids, in: E. R. Dougherty, P. D. Gader, J. C. Serra (Eds.), Image Algebra and Morphological Image Processing IV, Vol. 2030, SPIE, 1993, pp. 65–76.
[22] P. Salembier, M. H. F. Wilkinson, Connected operators: A review of region-based morphological image processing techniques, IEEE Signal Processing Magazine 26 (6) (2009) 136–157.
[23] P. Salembier, A. Oliveras, L. Garrido, Antiextensive connected operators for image and sequence processing, IEEE Transactions on Image Processing 7 (4) (1998) 555–570.
[24] P. Monasse, F. Guichard, Scale-space from a level lines tree, Journal of Visual Communication and Image Representation 11 (2) (2000) 224–236.
[25] P. Salembier, L. Garrido, Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval, IEEE Transactions on Image Processing 9 (4) (2000) 561–576.
[26] V. Vilaplana, F. Marques, P. Salembier, Binary partition trees for object detection, IEEE Transactions on Image Processing 17 (11) (2008) 2201–2216.
[27] S. Valero, P. Salembier, J. Chanussot, New hyperspectral data representation using binary partition tree, in: IEEE International Geoscience and Remote Sensing Symposium, Vol. 2, 2010, pp. 80–83.
[28] P. Soille, Constrained connectivity for hierarchical image decomposition and simplification, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (7) (2008) 1132–1145.
[29] P. Soille, Constrained connectivity for the processing of very-high-resolution satellite images, International Journal of Remote Sensing 31 (22) (2010) 5879–5893.
[30] S. Aksoy, H. G. Akcay, Multi-resolution segmentation and shape analysis for remote sensing image classification, in: International Conference on Recent Advances in Space Technologies, 2005, pp. 599–604.
[31] S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 674–693.
[32] P. Scheunders, J. Sijbers, Multiscale watershed segmentation of multivalued images, in: Proceedings of the International Conference on Pattern Recognition, Vol. 3, 2002, pp. 855–858.
[33] J. B. Kim, H. J. Kim, Multiresolution-based watersheds for efficient image segmentation, Pattern Recognition Letters 24 (1–3) (2003) 473–488.
[34] Y. Chibani, Selective synthetic aperture radar and panchromatic image fusion by using the à trous wavelet decomposition, EURASIP Journal on Applied Signal Processing 32 (14) (2005) 2207–2214.
[35] Y. L. Chang, L. S. Liang, C. C. Han, J. P. Fang, W. Y. Liang, K. S. Chen, Multisource data fusion for landslide classification using generalized

positive Boolean functions, IEEE Transactions on Geoscience and Remote Sensing 45 (6) (2007) 1697–1708.
[36] Z. Wang, D. Ziou, C. Armenakis, D. Li, Q. Li, A comparative analysis of image fusion methods, IEEE Transactions on Geoscience and Remote Sensing 43 (6) (2005) 1391–1402.
[37] B. Aiazzi, S. Baronti, M. Selva, Improving component substitution pansharpening through multivariate regression of MS + Pan data, IEEE Transactions on Geoscience and Remote Sensing 45 (10) (2007) 3230–3239.
[38] R. Goffe, L. Brun, G. Damiand, Tiled top–down combinatorial pyramids for large images representation, International Journal of Imaging Systems and Technology 21 (1) (2011) 28–36.
[39] S. Mallat, Wavelets for a vision, Proceedings of the IEEE 84 (4) (1996) 604–614.
[40] C. Kurtz, N. Passat, P. Gançarski, A. Puissant, Multiresolution region-based clustering for urban analysis, International Journal of Remote Sensing 31 (22) (2010) 5941–5973.
[41] L. Garrido, P. Salembier, D. Garcia, Extensive operators in partition lattices for image sequence analysis, Signal Processing 66 (2) (1998) 157–180.
[42] A. J. Plaza, J. C. Tilton, Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images, in: IEEE International Symposium on Geoscience and Remote Sensing, Vol. 7, 2005, pp. 946–4949.
[43] C. Wemmert, A. Puissant, G. Forestier, P. Gançarski, Multiresolution remote sensing image clustering, IEEE Geoscience and Remote Sensing Letters 6 (3) (2009) 533–537.
[44] F. de Lussy, P. Kubik, D. Greslou, V. Pascal, P. Gigord, J. Cantou, PLEIADES-HR image system products and geometric accuracy, in: Proceedings of the ISPRS Hannover Workshop on High-Resolution Earth Imaging for Geospatial Information, Vol. 1, 2005, pp. 50–57.
[45] S. H. Cha, S. N. Srihari, On measuring the distance between histograms, Pattern Recognition 35 (6) (2002) 1355–1370.
[46] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics, Speech and Signal Processing 26 (1) (1978) 43–49.
[47] F. Petitjean, A. Ketterlin, P. Gançarski, A global averaging method for Dynamic Time Warping, with applications to clustering, Pattern Recognition 44 (3) (2011) 678–693.
[48] R. Congalton, A review of assessing the accuracy of classifications of remotely sensed data, Remote Sensing of Environment 37 (1) (1991) 35–46.
[49] J. Inglada, Automatic recognition of man-made objects in high resolution optical remote sensing images by svm classification of geometric image features, ISPRS Journal of Photogrammetry and Remote Sensing 62 (3) (2007) 236–248.
[50] B. Özdemir, S. Aksoy, S. Eckert, M. Pesaresi, D. Ehrlich, Performance measures for object detection evaluation, Pattern Recognition Letters 31 (10) (2010) 1128–1137.