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

Camille Kurtz, Nicolas Passat, Pierre Gançarski, Anne Puissant

Hierarchical segmentation of multiresolution/multisource remote sensing images

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) images presents several challenges related to the size, accuracy and complexity of the considered data. Based on the availability of several images of a same scene at various resolutions (Medium, High and Very High Spatial Resolution), a hierarchical approach is proposed to iteratively extract segments of interest from the lowest to the highest resolution data, 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 its interaction. In order to do so, at each resolution, it requires an interactive segmentation of one sample region for each semantic class of the image, and then automatically reproduces the user's behaviour 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 objects detection and classification purpose.

Keywords: hierarchical segmentation, binary partition trees, multisource images, multiresolution, interactive/automated segmentation, 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) 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 (see Figure 1). Thus, by opposition to lower resolution images, all these patterns are not necessarily composed of homogeneous pixels (but are often hierarchically organised).

These specific properties induced by VHSR images lead to new challenges, for human experts (since the size and complexity of the images make visual analysis a time consuming and error prone task), and for image analysis tools (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 useful 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 the extraction of potentially hierarchical patterns, especially when such data are provided by different acquisition devices, providing complementary information at distinct radiometric bands (see Figure 2).

Such new segmentation tools should allow the end-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) to VHSR ones) [5] in a hierarchical strategy which enables, at a given resolution, the exploration of the whole structure of an urban scene [6, 7]. By analysing first the image content at a coarse resolution and then gradually increasing this resolution [8], it is in particular possible to detect complex patterns (which structure the scene) while avoiding the semantic noise induced by the details [9].

Based on these considerations, a hierarchical approach is proposed to iteratively extract, from multiresolution images of a same urban scene, segments of interest from the lowest to the highest resolution data (by opposition to ascendant approaches often proposed in the literature [10]), 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 its interaction. In order to do so, at each resolution, it requires an interactive segmentation of one sample region for each semantic class of the image, and then automatically reproduces the

*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)

Figure 1: Example of object of interest (a urban block) depicted in red on a panchromatic satellite image with a spatial resolution of 60 cm. (FIXME: origin + copyright?)

user's behaviour 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.

The article, which is an extended and improved version of the preliminary work described in [11], is organised as follows. Section 2 provides a state of the art on hierarchical and multiresolution segmentation dealing with (but not restricted to) remote sensing data. Section 4 describes the proposed segmentation method. Section 5 gathers experiments enabling to assess the relevance of the approach. Conclusions and perspectives will be found in Section 6.

2. Related work

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 hypotheses. This can lead to valid results for basic objects extraction from High Spatial Resolution (HSR) images [2], but not for images (*e.g.*, VHSR ones) and/or objects of higher complexity [3].

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 [10], 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 parcels by taking into account the structural properties and the relations between simple objects.) Another example of such approach can be found in [12] in which a set of particular subgraphs of a valued graph is introduced. However, two major problems are inherent to such approaches. Firstly, computing all the grouping possibilities within the (large) space of candidate segments is not actually feasible. Secondly, the capacity 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 seems unable to directly extract more complex structure at a higher semantic level.

Moreover, when dealing with HSR images, different composite objects could be merged to form new kinds of structures of interests enabling different levels of analysis. 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 emerge at various scales and be related according to some suitable criteria in a hierarchical structure. Consequently, it is needed to improve "grouping approaches" by considering hierarchical strategies to enable the (potential) extraction of more complex structure (with a higher semantic level).

To this end, some techniques providing a multiscale partitioning have been recently 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, *e.g.*, [13] for an example of pioneering work).

In the field of remote sensing (and especially for HSR images), several techniques have been proposed. In [14], 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, and 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 very high resolution images. To overcome this limitation, new approaches have been proposed. In [6], morphological profiles are enriched with neighbourhood and spectral information, while in [15], a framework is proposed to detect complex objects in HSR images by combining spectral information with structural one. 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 detail in which the segmentation at a level can be produced by merging regions at finer levels. 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 [16], where segmentation is treated as a graph partitioning problem. However, such an approach, which makes the assumption that the images contain only few objects of interests is not adapted to capture the richness and complexity of HSR images. Another approach can be found in [17], 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

(a) 1 pixel = 10 m × 10 m (b) 1 pixel = 2.5 m × 2.5 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. (b) High Spatial Resolution (HSR) image. (c) Very High Spatial Resolution (VHSR) image.
(FIXME: bands + origin + copyright?)

robust) splitting function remains an open issue. In the, more frequent, bottom-up approaches ("region merging" or "split and merges" methods), the finest segmentations are produced first, and their regions are then merged, based on similarity criteria [18]. In remote sensing, various algorithms use this principle. For instance, in [19], a hierarchical segmentation algorithm that combines spectral clustering with iterative region growing is proposed. The multiscale segmentation algorithm presented in [20] 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 [21], 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 the user-defined threshold related to local homogeneity criteria. An alternative solution is proposed in [22]. 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 [23, 24] may be used in a hierarchical segmentation fashion by using, for instance, tree data structures. Notions such as component-tree [25] and level-lines tree [26], potentially enable to perform hierarchical segmentation, by enabling the fusion of flat zones. However, such structures strongly relying on the image intensity and in particular on extremal values, the obtained segmented components may be non relevant in the case of satellite images. By opposition, the binary partition tree (BPT) [7] 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 of resolution 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 [27]. A last approach, based on the constrained connectivity paradigm, has been recently introduced in [28] and applied to process VHSR 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 images may generally have very different sizes and be formed by various heterogeneous objects. In order to cope with this variability, either the used features must be size invariant, or the image must be processed at different resolutions. As the resolution gets coarser from that of the original image, larger (and thus, complex) structures that provide the general image context can be represented without being convoluted with the details. 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 has been proposed using the wavelet transform [30, 31], which provides 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 set of images (*e.g.*, MSR, HSR, VHSR). It is then 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 consists of fusing 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 [5] 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. The 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 process is also similar to the strategy used by the human vision system, already considered in [9, 39] to create thematic map from HSR and MSR images.

2.4. Purpose

Based on these considerations, we propose a hierarchical top-down segmentation method, extending the BPT to deal with multiresolution images. 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 is based on interactive tree-cut segmentation (based on the skills of the end-user), defined by the user on a part of the images, and automatically reproduced on the whole data. 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. It aims, in particular, at understanding the scene in the same way as the human vision system.

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

Let $\mathfrak{S} = \{R_i\}_{i=1}^t$ be a partition of E, associated to an image $\mathcal{I} : E \to V$, and $\mathcal{I}_{\mathfrak{S}} : E \to [\![1, t]\!]$ be the image induced by $\mathfrak{S}$. A clustering[1] of $\mathcal{I}$ into u classes is provided by the definition of a map $C : [\![1, t]\!] \to [\![1, u]\!]$ which, to each one of the t regions R_i, associates one of the u classes $C(i)$. 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 class. 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.

[1] This definition which enables to conveniently formalise object-based clustering, also enable 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.

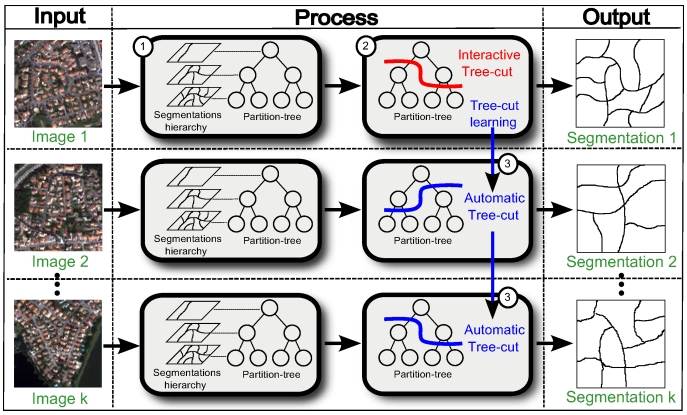


Figure 3: Interactive segmentation method (see Section 4.1). In green: input/output. In red: user interactions. In blue: automatic processing. FIXME: complete the figure accordingly.

3.5. *Binary partition tree*

Let $\mathcal{I} : E \to V$ be an image. A binary partition tree (BPT) [7] 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 $\varphi : \mathcal{N} \backslash \{E\} \to \mathcal{N}$ is a function verifying the following properties:

(P) 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 $N_1 = \varphi(N_2)$, then N_1 (resp. N_2) is the "father" (resp. a "child") of N_2 (resp. N_1). The node E is the root of the BPT. The nodes of $\mathcal{N} \setminus \varphi(\mathcal{N} \setminus \{E\})$, *i.e.*, those how have 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 classical 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, is performs n successive steps (one step for each resolution), each step being iteratively composed of:

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

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 interactive segmentation approach (i) is an original strategy, which constitutes another contribution of this article. It is then 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*

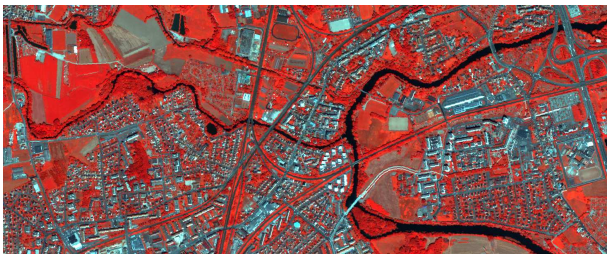
One of the main steps of the proposed segmentation methodology consists of a segmentation method visually summarised in Figure 3.

This method takes as input $k \geq 2$ images of different scenes, having the same resolution and semantics, and 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 proposes a segmentation, by performing a cut in the image BPT. interactively performs a segmentation, by performing a cut in its BPT. 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. 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).

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



(a) HSR image

(b) Elongation map

Figure 4: Elongation map computation (see text). (a) High Spatial Resolution (HSR) image. FIXME: copyright (b) Corresponding elongation map (elongated structures in light grey, non-elongated ones in dark grey).

has not been defined yet. (In the context of image segmentation, such couples of nodes are generally chosen as spatially adjacent ones, thus leading only 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). In the sequel, the chosen region model and merging criterion are defined. It has to 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 its segmentation example.

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

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

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, respectively. During the merging process, the region model of two merged regions R_i and R_j is then provided by

$$M_r(R_i \cup R_j) = \langle (\min\{v_b^-(R_i), v_b^-(R_j)\}, \max\{v_b^+(R_i), v_b^+(R_j)\}) \rangle_{b=1}^s$$
$$M_g(R_i \cup R_j) = (e(R_i \cup R_j), a(R_i) + A(R_j))$$

By opposition to M_r and a, the computation of which 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:

(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;
(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 obtained thanks to this strategy.

Merging criterion. The basic merging criterion used in most of image segmentation approaches is generally radiometric homogeneity [**?**]. However, when dealing with (V)HSR images, geometrical details also have to be taken in consideration. Consequently, as indicated in the above region model, 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. This leads to the following merging criteria

$$O_r(R_i, R_j) = \frac{1}{s} \sum_{b=1}^s \max\{v_b^+(R_i), v_b^+(R_j)\} - \min\{v_b^-(R_i), v_b^-(R_j)\}$$
$$O_g(R_i, R_j) = \frac{1}{2}(e(R_i \cup R_j) + a(R_i \cup R_j))$$

FIXME: normalisation of e *vs.* a? The similarity measure between two neighbouring regions R_i and R_j can then be computed as

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

with $\alpha \in [0, 1]$. FIXME: normalisation of O_r *vs.* O_g? In practice, the closest 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

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

provides a satisfactory behaviour of the merging function O. (Note in particular that the user may be free to tune this function, by introducing parameters in $\alpha(O_r)$ enabling to control,

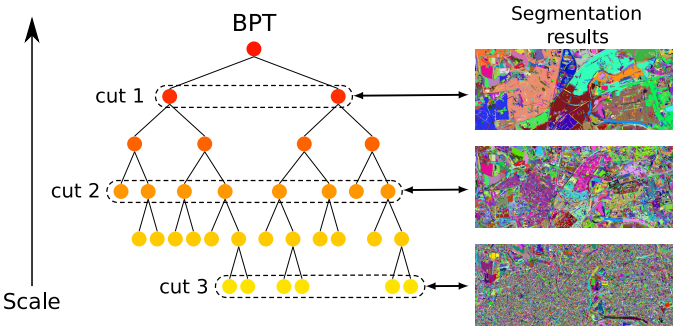


Figure 5: An example of BPT associated to the HSR image $\mathcal{I} : E \to V$ presented in Figure 4(a) (the number of nodes is significantly reduced, for the sake of readability). The nodes of $\mathcal{N}$ are depicted by colour disks (the root E is the highest node). The function φ is modelled by 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. FIXME: figure to modify.

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.

4.1.2. Learning of the cut example

By opposition to other strategies devoted to automatically extract cuts from BPTs? [6, 41], 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 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 elements of knowledge which characterise it and then enable its reproduction.

In previous works [11], the cuts in these BPTs were straightforwardly obtained by performing a thresholding on the similarity measure (called *energy* in the sequel) FIXME: formally define the energy related to the O function, 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.

Let $C \subseteq \mathcal{N}$ be the cut defined in the BPT interactively processed by the user. This cut C is first partitioned into two subsets C_e and $C_{\overline{e}}$, corresponding to the nodes/regions being elongated and non-elongated, respectively. 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 $C_{\overline{e}}$, which correspond to the areas "bounded" by the linear elements of C_e.

A clustering process which one? is then carried out on the regions of $C_{\overline{e}}$. This clustering is based on the histogram of each region, *i.e.*, for each region $R \in C_{\overline{e}}$ of $\mathcal{I} : E \to V$ (with $R \subseteq E$), the criterion characterising R is its (normalised) histogram $\mathcal{H}_{\mathcal{I},X} : V \to \mathbb{N}$ (with $\sum_{v \in V} \mathcal{H}_{\mathcal{I},X}(v) = 1$). FIXME how is determined u? FIXME how are computed the centroids? 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 $\mathcal{H}_i : V \to \mathbb{N}$ being actually an "averaged" (normalised) histogram of the cluster K_i. FIXME is it true?

FIXME: unclear (to discuss together) Centroid-based clustering algorithms require a distance (and an averaging method) to compare each object to classify. The classical distance used to compare two histograms is the Euclidean one. However, this distance suffers from the problem of the *shuffling invariance* property. This property is not desirable in the distance between the histograms of ordinal type measurements. Levels cannot be permuted by definition of ordering in levels. To deal with this issue, a solution consists in using the Dynamic Time Warping distance [42] which enables small distorsions on the radiometric axis. Associated to this distance, an averaging method was introduced in [43].

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. These centroids then have to be involved in the automatic segmentation of the $k-1$ other images.

This can be conveniently done by finding, for each one of the $k-1$ images $\mathcal{I}_j$ ($j \in [\![1, k-1]\!]$), a cut C_j in the BPT of $\mathcal{I}_j$, minimising an inertia FIXME: why "inertia"? 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 $\{\mathcal{H}_{\mathcal{I}_j,X}\}_{X \in C_j}$). The inertia $I(C_j)$ associated to a cut C_j, with respect to the set of clusters $\{\mathcal{H}_i\}_{i=1}^{u}$ can be defined as

$$I(C_j) = \sum_{i=1}^{u} \frac{|\bigcup_{X \in C_j^i} X|}{|\bigcup_{X \in C_j} X|} \cdot d^2(\overline{\mathcal{H}_{i,j}}, \mathcal{H}_i)$$

FIXME what is d? FIXME why "2" in d^2? where $C_j^i \subseteq C_j$ 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.*

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

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 [2] as

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

[2] 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 $\mathcal{N}_{j,\overline{e}} \subseteq \mathcal{I}_j$ corresponding to 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 $C_{\overline{e}}$, 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 hat the considered tree is no longer a binary one, since a node may have

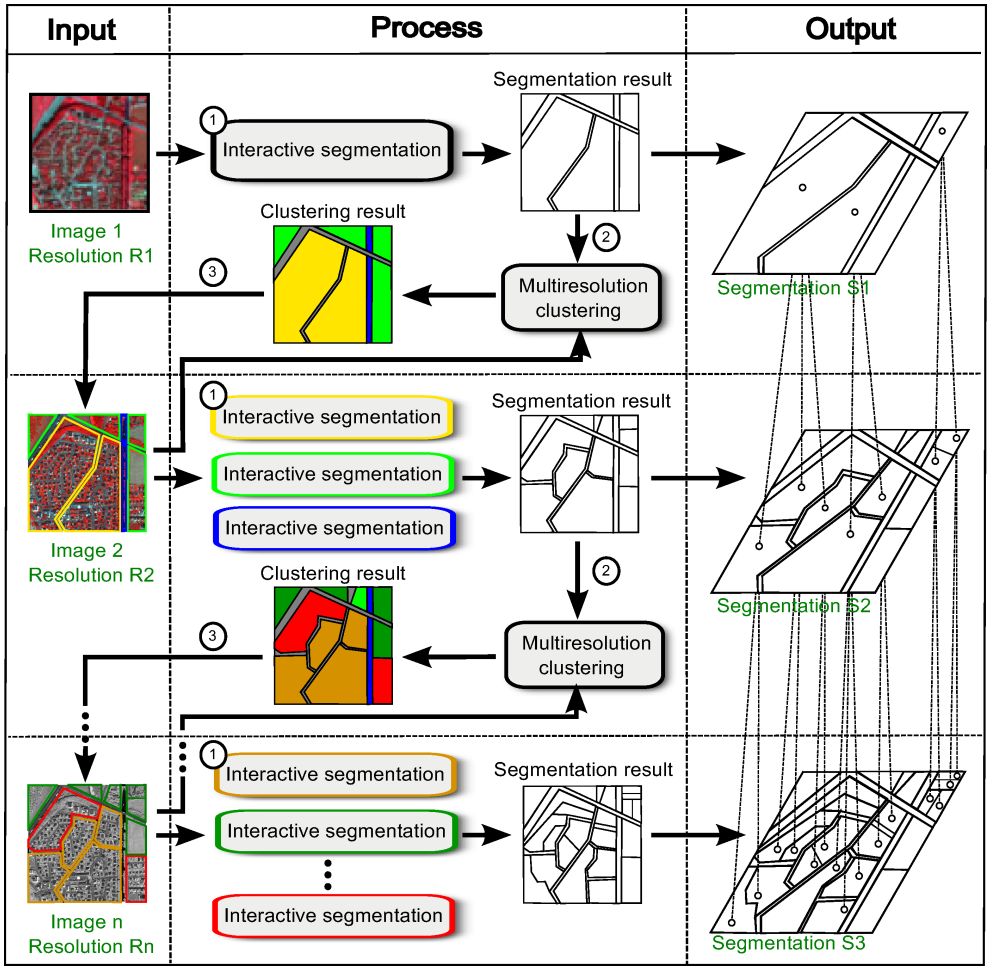


Figure 6: Work-flow overview (see Sec. 4.2). In green: input/output.

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

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

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

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

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 being close to the segmentation example provided by the user.

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 $\{\mathcal{I}_{\mathfrak{S},i}\}_{i=1}^n$ of segmented images (one per considered image/resolution), hierarchically linked, enabling different scales of interpretation.

Parameters: tolerance, neighbourhood radius, and orientations in $e(.)$, $a, b \in \alpha(.)$, number u of centroids.

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 k-th step, the image $\mathcal{I}_k$, is considered (it is then assumed that the images $\mathcal{I}_i$ ($i \in [\![1, k-1]\!]$) have already been processed).

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.

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 6 to visually follow the description of the methodology.

4.2.1. *Image segmentation*

Thanks to the previous processing of $\mathcal{I}_{k-1}$, a clustering of $\mathcal{I}_k : E_k \rightarrow V_k$ into u_{k-1} clusters FIXME: how is u_{k-1} determined? is already available[3]. (For instance, in Figure 6, a clustering of $\mathcal{I}_2$ into three (blue–water, green–urban vegetation and yellow–urban areas) clusters is available.) These clusters enable to divide $\mathcal{I}_k$ into u_{k-1} semantic classes corresponding of the level of details of the (lower) resolution of $\mathcal{I}_{k-1}$.

Each one of these classes may then be decomposed into new classes corresponding to the level of details of $\mathcal{I}_k$. In order to do so, it is necessary to perform a segmentation of the part of the image corresponding to each one of the u_{k-1} semantic classes, *i.e.* to segment the subimage $\mathcal{I}_{k,i} : K_i \rightarrow V_k$ of $\mathcal{I}_k$ defined on the cluster $K_i \subseteq E_k$ for any $i \in [\![1, u_{k-1}]\!]$ (note that the user may however choose to restrict his study to only certain of these classes, thus leading to a partial analysis of the images).

For each considered semantic class $i \in [\![1, u_{k-1}]\!]$, the segmentation of $\mathcal{I}_{k,i} : K_i \rightarrow V_k$ 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 $\mathcal{I}_{k,i}$ of same resolution and semantics, 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 (Sections 4.1.1 and 4.1.2), and this section is then automatically reproduced in all the other subimages (Sections 4.1.2 and 4.1.3).

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

FIXME: topological correction?

4.2.2. *(Multiresolution) image clustering*

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

This clustering relies on a multiresolution approach fully described in [40], and briefly recalled hereafter. The reader may also refer to Figure 7 for a visual outline of this approach.

This approach takes as input the image $\mathcal{I}_{k-1} : E_{k-1} \rightarrow V_{k-1}$, namely the image to be clustered, the segmentation $\mathcal{I}_{\mathfrak{S},k-1}$, which provides a partition $\mathfrak{S}$ of E_{k-1}, and the "next" image $\mathcal{I}_k : E_k \rightarrow V_k$. The main idea is to fuse the information provided by (1) the analysis of the "low" resolution regions of $\mathfrak{S}$ and (2) the "high" resolution semantic clustering of $\mathcal{I}_k$ (provided by a classical clustering method directly applied on the radiometric values of the pixels), to obtain a final clustering result corresponding to an 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. The final clustering result is computed by classifying (in an unsupervised way) the regions of the lowest resolution segmented image using these composition histograms.

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

[3] 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.

Table 1: Typologies and levels used by end-users to map urban areas at different scales.

1:100, 000-1:25, 000 – Urban areas level	1:10, 000 – Urban blocks level	1:5, 000 – Urban objects level
• 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

5. Experimental studies

5.1. *Applicative context*

In the domain of urban planning and management a wide range of object nomenclatures has been defined. For instance, the Corine Land Cover nomenclature has been defined for Landsat images (30 m spatial resolution), whereas the SPOT Thema nomenclature has been defined for Spot images (5–20 m). These existing products enable mapping of urban areas, respectively, from 1:100 000 (Corine Land Cover nomenclature) to 1:50 000 and 1:25 000 (SPOT Thema). With high-spatial-resolution (HSR) (1–5 m) satellite images, it is possible to extract urban objects (*e.g.*, house, garden, and road). This allows one to map individual objects at scale from 1:10 000 to 1:5000.

5.2. *Software*

provide details.

5.3. *Experiments and results*

This section describes the experiments carried out with the proposed multiresolution framework in the context of the segmentation of urban patterns from MSR and HSR images. Subsection 5.3.1 presents the data which were used to perform the method. Experiments and parametrisation are described in Subsection 5.3.3. The results of the multiresolution method devoted to extract urban elements are then presented and analysed in Subsection 5.3.2. Finally, a computation time study is performed in Subsection 5.3.5.

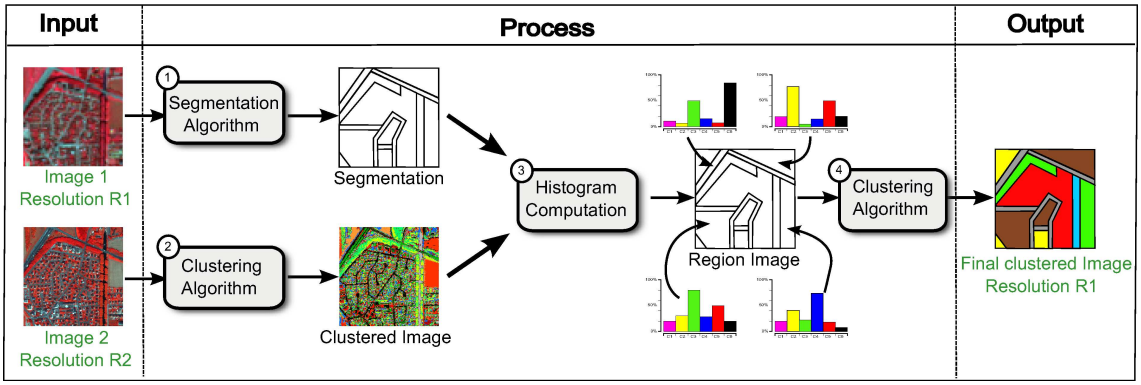


Figure 7: Multiresolution clustering approach (see Sec. 4.2.2). In green: input/output.

5.3.1. *Images*

Experiments have been performed on two sets of images called Strasbourg dataset and Toulouse dataset.

The Strasbourg dataset is composed by three multispectral images with different spatial resolutions (2.8 m, 10 m and 20 m) acquired by the QuickBird, Spot-5 and Spot-4 satellites (respectively in May 2001, August 2002 and July 2001). The Spot-5 and Spot-4 multispectral images (Fig. (a, b)) have three spectral bands (green, red, near-infrared). The QuickBird multispectral image (Fig. (c)) is available in four spectral bands (blue, green, red and near-infrared). All the data are georeferenced in the same local cartographic projection (Lambert I). These images present a part (1, 500 m $\times$ 2, 100 m) of the urban area of Strasbourg (France) which is a typical suburban area with water surfaces (in black, centre of the image), forest area (in red, bottom left of the image), industrial areas (in grey, upper right of the image), individual or collective housing blocks (in red, black and white textured on the MSR image, in red, blue and white textured in the HSR image), agricultural zones with different spectral responses due to the seasons (bare soil at the end of spring on the HSR image can appear in red in summer on the MSR image).

The Toulouse dataset is composed by three multispectral images[4] with approximately the same spatial resolutions as above (2.5 m, 10 m and 20 m). The HSR image (2.5 m) was acquired by the Spot-5 satellite in September 2003 (Fig. (c)). This image is a result of a fusion between the panchromatic image at 2.5 m and XS bands at 5 m. The resulting HSR image has four spectral bands (green, blue, red, near-infrared). The MSR images (20 m and 10 m) were simulated from the HSR one by a degradation process (Fig. (a, b)). This degradation process transforms HSR images into MSR ones by simulating the "physical" properties of such sensors. These images are georeferenced in the same local cartographic projection (Lambert III) and present a part (1, 600 m $\times$ 2, 100 m) corresponding to the South West of the city of Toulouse (France) which is also a typical suburban area.

[4]The authors would like to acknowledge the support of the Centre National d'Études Spatiales (CNES) which provided the images of the Toulouse dataset. We are grateful to Jordi Inglada for his assistance in providing and processing these data.

Table 2: Satellite Data used for the experiments.

Image	Resolution	Size	Memory
Strasbourg VHSR	60cm	15 000 $\times$ 15 000	1,4 Gb
Strasbourg HSR	2,4m		
Strasbourg MSR	10m		
Toulouse HSR	2,4m		
Toulouse MSR	10m		
Toulouse MSR	20m		

5.3.2. *Validation*

Results produced by the method have been assessed by quantitative comparisons with a land cover map. In order to compare clustering results to land cover reference maps, we have computed:

- the Kappa index (κ);
- the global accuracy;
- the percentage of false positive;
- the percentage of false negative;

The Kappa index (κ) is a measure of global classification accuracy [44] and is defined by:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (1)$$

where $\Pr(a)$ is the relative agreement among raters, and $\Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly choosing each category. The Kappa takes value in $[0, 1]$ and decreases as the classification is in disagreement with the ground-truth map. There are many ways to compute this index. When data is not labelled (as it is the case with clustering), this computation consists of taking all point couples $(p_1 , p_2) = ((x_1, y_1) , (x_2, y_2))$ 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. p_1 and p_2 belong to the same partition both in the clustering and in reference map (counter ss);

2. p_1 and p_2 belong to the same partition in the clustering but not in the reference map (counter sd);
3. p_1 and p_2 belong to the same partition in the reference map but not in the clustering (counter ds);
4. p_1 and p_2 belong to the same partition neither in the reference map nor in the clustering (counter dd).

Thus, the Kappa index can be computed with:

$$\Pr(a) = \frac{ss + dd}{ss + sd + ds + dd} \quad (2)$$

and

$$\Pr(e) = \frac{(ss + sd) \times (ss + ds) \; + \; (sd + dd) \times (ds + dd)}{(ss + sd + ds + dd)^2} \quad (3)$$

5.3.3. Experiments and parametrisation

5.3.4. Results

5.3.5. Comparative study, Computation time study

5.4. Discussion

6. Conclusion and perspectives

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 identifyurban objects, Geocarto International 17 (1) (2002) 33–44.

[5] 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.

[6] 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.

[7] 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.

[8] S. Mallat, Wavelets for a vision, Proceedings of the IEEE 84 (4) (1996) 604–614.

[9] 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.

[10] M. J. Barnsley, S. L. Barr, Distinguishing urban land-use categories in fine spatial resolutionland-cover data using a graph-based, structural pattern recognition system, Computers, Environment and Urban Systems 21 (3) (1997) 209–225.

[11] C. Kurtz, N. Passat, A. Puissant, P. Gançarski, Hierarchical segmentation of multiresolution remote sensing images, in: FIXME (Ed.), International Symposium on Mathematical Morphology - ISMM'11, 10th International Symposium, Proceedings, Vol. FIXME of Lecture Notes in Computer Science, Springer, 2011, pp. FIXME–FIXME.

[12] 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.

[13] M. Pietikainen, A. Rosenfeld, Image segmentation by texture using pyramid node linking, IEEE Transactions on Systems, Man and Cybernetics 11 (12) (1981) 822–825.

[14] 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.

[15] 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.

[16] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 888–905.

[17] 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.

[18] T. Pavlidis, Structural pattern recognition, Springer Verlag, 1977.

[19] 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.

[20] M. Baatz, A. Schäpe, Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation, in: W. Verlag (Ed.), Strobl, J., Blaschke, T., Griesebner, G. (Eds.), Angewandte Geographische Informations-Verarbeitung XII, Vol. 58 of Karlsruhe, 2000, pp. 12–23.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] P. Salembier, A. Oliveras, L. Garrido, Antiextensive connected operators for image and sequence processing, IEEE Transactions on Image Processing 7 (4) (1998) 555–570.

[26] P. Monasse, F. Guichard, Scale-space from a level lines tree, Journal of Visual Communication and Image Representation 11 (2) (2000) 224–236.

[27] V. Vilaplana, F. Marques, P. Salembier, Binary partition trees for object detection, IEEE Transactions on Image Processing 17 (11) (2008) 2201–2216.

[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. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 674–693.

[31] 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.

[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 generalizedpositive 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, Top-down combinatorial pyramids for

large images representation, International Journal of Imaging Systems and Technology FIXME (FIXME) (2011) FIXME–FIXME.

[39] 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.

[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] 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.

[42] 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.

[43] 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.

[44] R. Congalton, A review of assessing the accuracy of classifications of remotely sensed data, Remote Sensing of Environment 37 (1) (1991) 35–46.