Segmentation of Multimodal Images Based on Hierarchies of Partitions

Guillaume Tochon[1(✉)], Mauro Dalla Mura[1], and Jocelyn Chanussot[1,2]

[1] GIPSA-lab, Grenoble Institute of Technology, Saint Martin d'Hères, Grenoble, France
{guillaume.tochon,mauro.dalla-mura,
jocelyn.chanusshot}@gipsa-lab.grenoble-inp.fr
[2] Department of Electrical and Computer Engeneering, University of Iceland, Reykjavik, Iceland

Abstract. Hierarchies of partitions are widely used in the context of image segmentation, but when it comes to multimodal images, the fusion of multiple hierarchies remains a challenge. Recently, braids of partitions have been proposed as a possible solution to this issue, but have never been implemented in a practical case. In this paper, we propose a new methodology to achieve multimodal segmentation based on this notion of braids of partitions. We apply this new method in a practical example, namely the segmentation of hyperspectral and LiDAR data. Obtained results confirm the potential of the proposed method.

Keywords: Image segmentation · Multimodal image · Hierarchy of partitions · Energy minimization

1 Introduction

Multimodality is nowadays increasingly used in signal and image processing. In fact, multimodal data (i.e., data of a physical phenomenon collected from different sensors/locations, each of these showing a particular aspect of this phenomenon) allow to take advantage of both the correlation and complementarity between each mode (i.e., data collected by one particular sensor) to better understand the underlying physical phenomenon of the source. However, there is a good number of challenges that still must be faced in order to fully exploit the nature of multimodal data [8]. One talks in particular of multimodal images when several images of the same scene have been acquired by different sensors. This multimodality phenomenon occurs in several fields of image processing, such as medical imaging [1] or remote sensing [3,4]. However, the design of adapted tools to process multimodal images remains a challenge, notably due to the diverse physical meanings and contents of images produced by all possible imaging sensors. Image segmentation is a particular process that would surely benefit from the development of such multimodal tools, since it aims at partitioning an image into regions that "make sense" with respect to some underlying goal. The segmentation of a multimodal image should benefit from the complementarity

J.A. Benediktsson et al. (Eds.): ISMM 2015, LNCS 9082, pp. 241–252, 2015.
DOI: 10.1007/978-3-319-18720-4_21

of its modes to ensure a more accurate delineation of its regions, in particular when those regions share similar features in one mode but not in the other ones.

Image segmentation constitutes an ill-posed problem since a given image can often be properly segmented at various levels of detail, and the precise level to choose depends on the underlying application (an optimal level might not exist). A potential solution to this intrinsic multiscale nature issue is to use a hierarchy of segmentations, which organizes in its structure all the potential scales of interest in a nested way. The hierarchy can be built once for a given image regardless of the application, and its level of exploration can then be tuned afterwards to produce the desired segmentation [13]. In [5] for example, this tuning relies on some energy minimization process over all the possible segmentations that can be extracted from the hierarchy. The *optimal* scale thus depends on the definition of the energy. However, handling the case of a multimodal image (and thus of several hierarchies) still remains an open question. Recently, the concept of braids of partitions has been introduced [6] as a potential tool to tackle this issue. We define in this paper a strategy of energy minimization for segmenting hierarchies of segmentations issued from different modalities, based on this concept of braids of partitions.

In Section 2, we summarize the works of [5, 7] and [6] about energy minimization over hierarchies and braids of partitions, respectively. In Section 3, we introduce a new methodology to achieve multimodal segmentation, based on energy minimization over braids of partitions. Section 4 features the application of the proposed methodology in a practical case, namely the joint segmentation of hyperspectral and LiDAR data, and presents some results. Conclusion and future work are drawn in Section 5.

2 Segmentation by Energy Minimization

We first define the notations used throughout the paper, before quickly recalling the notions of energy minimization over hierarchies and braids. The words segmentation and partition are used interchangeably in the following.

2.1 Definitions and Notations

Let $\mathcal{I} : E \rightarrow V$, $E \subseteq \mathbb{Z}^2$, $V \subseteq \mathbb{R}^n$, be a generic image, of elements (pixels) $\mathbf{x}_i \in E$. A partition of E, denoted π, is a collection of regions $\{\mathcal{R}_i \subseteq E\}$ (also called classes) of E such that $\mathcal{R}_i \cap \mathcal{R}_{j \neq i} = \emptyset$ and $\bigcup_i \mathcal{R}_i = E$. The set of all possible partitions of E is denoted Π_E. For any two partitions $\pi_i, \pi_j \in \Pi_E$, $\pi_i \leq \pi_j$ when for each region $\mathcal{R}_i \in \pi_i$, there exists a region $\mathcal{R}_j \in \pi_j$ such that $\mathcal{R}_i \subseteq \mathcal{R}_j$. π_i is said to refine π_j in such case. Π_E is a complete lattice for the refinement ordering $\leq$. Minimizing some energy function over Π_E requires first the definition of a regional energy, i.e., a function $\mathcal{E}$ that maps any region $\mathcal{R} \subseteq E$ to $\mathbb{R}^+$, and the definition of some operator $\mathfrak{D}$ (such as $\sum$, $\prod$ or $\bigvee$ for instance) to express the energy of a partition as a composition of the energies of its regions:

$$\mathcal{E}(\pi = \{\mathcal{R}_i\}) = \underset{\mathcal{R}_i \in \pi}{\mathfrak{D}} \mathcal{E}(\mathcal{R}_i). \tag{1}$$

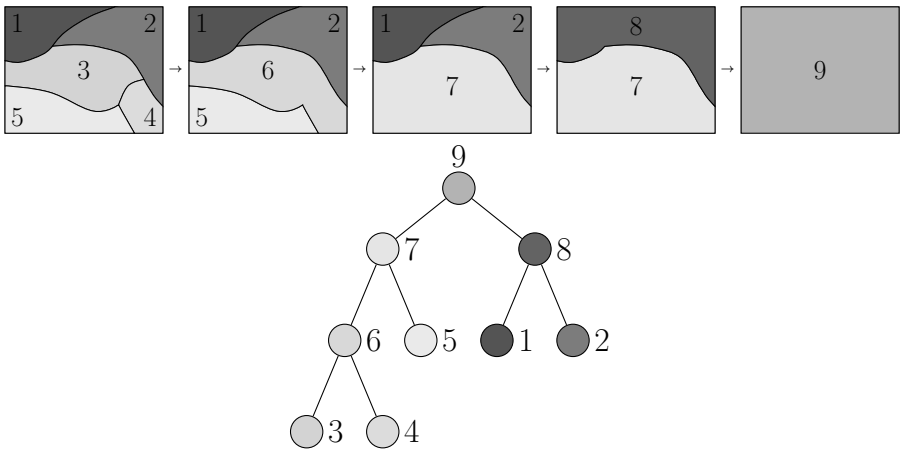


Fig. 1. Example of hierarchy of partitions (binary partition tree [14])

Well-known methods to perform image segmentation by energy minimization non-exhaustively include the Mumford-Shah functional [10], graph cuts [2] or Markov random fields [9]. However, finding the optimal partition that minimizes a given energy remains a difficult task, mainly due to the huge cardinality of Π_E (a 5×5 image can be partitioned in more than 4.6×10^{18} different ways). Hierachies, by restraining the space of possible partitions, are an appealing tool to minimize the energy on.

2.2 Minimization over a Hierarchy

A hierarchy of segmentations of E is a collection $H = \{\mathcal{R} \subseteq E\}$ such that $\emptyset \notin H$, $E \in H$ and $\forall\, \mathcal{R}_i, \mathcal{R}_j \in H$, $\mathcal{R}_i \cap \mathcal{R}_j \in \{\emptyset, \mathcal{R}_i, \mathcal{R}_j\}$. In other words, any two regions belonging to a hierarchy are either disjoint or nested. The most common way to obtain a hierarchical decomposition of an image is to start from an initial partition π_0 and to iteratively merge its regions until the whole image support is reached [12,14], resulting in a sequence of partitions $\pi_0 \leq \pi_1 \leq \cdots \leq \pi_n = \{E\}$, as displayed in Figure 1. Regions of π_0 are called *leaves*, $\pi_n = \{E\}$ is called the *root* of the hierarchy, and each non leaf node $\mathcal{R}$ contains a set of $\mathsf{S}(\mathcal{R})$ children nodes. A *cut* of H is a partition π of E whose regions belong to H. The set of all cuts of a hierarchy H built over the image $\mathcal{I}$ is denoted $\Pi_E(H)$, and is a sub-lattice of Π_E. $H(\mathcal{R})$ denotes the sub-hierarchy of H rooted at $\mathcal{R}$. Any cut of the sub-hierarchy $H(\mathcal{R})$ is called a *partial partition* of $\mathcal{R}$ following [11], and is denoted $\pi(\mathcal{R})$. The cut of H that is minimal (i.e., optimal) with respect to the energy $\mathcal{E}$ is defined as:

$$\pi^\star = \underset{\pi \in \Pi_E(H)}{\operatorname{argmin}}\ \mathcal{E}(\pi) \tag{2}$$

Assumptions on $\mathcal{E}$ under which it is easy to retrieve the minimal cut $\pi^\star$ have been studied in [5] in the context of separable energies (i.e., $\mathcal{E}(\pi) = \sum_{\mathcal{R} \in \pi} \mathcal{E}(\mathcal{R})$) and later generalized in [7] to wider classes of composition laws $\mathfrak{D}$, namely *h-increasing energies*. An energy $\mathcal{E}$ is said to be h-increasing when given any two

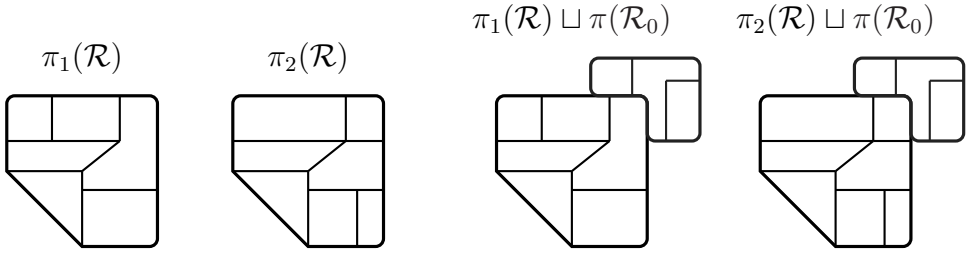


Fig. 2. Example of a h-increasing energy

$\mathcal{R}, \mathcal{R}_0 \in H$ disjoint, given partial partitions $\pi_1(\mathcal{R})$, $\pi_2(\mathcal{R})$ and $\pi(\mathcal{R}_0)$, then $\mathcal{E}(\pi_1(\mathcal{R})) \leq \mathcal{E}(\pi_2(\mathcal{R})) \Rightarrow \mathcal{E}(\pi_1(\mathcal{R}) \sqcup \pi(\mathcal{R}_0)) \leq \mathcal{E}(\pi_2(\mathcal{R}) \sqcup \pi(\mathcal{R}_0))$, with $\sqcup$ denoting disjoint union (concatenation). An example of h-increasing energy is depicted in Figure 2. In that case, it is possible to find the minimal cut of H by solving for each node $\mathcal{R}$ the following dynamic program:

$$\mathcal{E}^\star(\mathcal{R}) = \min\left\{\mathcal{E}(\mathcal{R}), \bigoplus_{r \in \mathsf{S}(\mathcal{R})} \mathcal{E}(\pi^\star(r))\right\} \tag{3}$$

$$\pi^\star(\mathcal{R}) = \begin{cases} \{\mathcal{R}\} & \text{if } \mathcal{E}(\mathcal{R}) \leq \bigoplus\limits_{r \in \mathsf{S}(\mathcal{R})} \mathcal{E}(\pi^\star(r)) \\ \bigsqcup\limits_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r) & \text{otherwise} \end{cases} \tag{4}$$

The optimal cut of $\mathcal{R}$ is given by comparing the energy of $\mathcal{R}$ and the energy of the disjoint union of the optimal cuts of its children, and by picking the smallest of the two. The optimal cut of the whole hierarchy is the one the root node, and is reached by scanning all nodes in the hierarchy in one ascending pass [5].

Energies in the literature often depend in practice on a positive real-valued parameter λ that acts as a trade-off between simplicity (i.e., favoring under-segmentation) and a good data fitting of the segmentation (i.e., leading to over-segmentation). These energies $\mathcal{E}_\lambda$ generate sequences of optimal cuts $\{\pi^\star_\lambda\}$ in turn indexed by this parameter λ. The behavior of $\pi^\star_\lambda$ with respect to λ has been studied in [7], which introduced in particular the property of *scale-increasingness*: $\mathcal{E}_\lambda$ is scale-increasing if for any $\mathcal{R} \in H$, any of its partial partition $\pi(\mathcal{R})$, and any $0 \leq \lambda_1 \leq \lambda_2$, $\mathcal{E}_{\lambda_1}(\mathcal{R}) \leq \mathcal{E}_{\lambda_1}(\pi(\mathcal{R})) \Rightarrow \mathcal{E}_{\lambda_2}(\mathcal{R}) \leq \mathcal{E}_{\lambda_2}(\pi(\mathcal{R}))$.
In the case where the energy is h-increasing for any λ and scale-increasing with respect to λ, the family $\{\pi^\star_\lambda\}$ of optimal cuts is hierarchically organized, that is

$$\lambda_1 \leq \lambda_2 \Rightarrow \pi^\star_{\lambda_1} \leq \pi^\star_{\lambda_2}. \tag{5}$$

In such case, it is possible to transform some hierarchy H into an optimal version $H^\star$, composed of all the optimal cuts $\pi^\star_\lambda$ of H when λ spans $\mathbb{R}^+$. In practice, the energy $\mathcal{E}_\lambda$ is seen as a function of λ, and (3) is conducted over

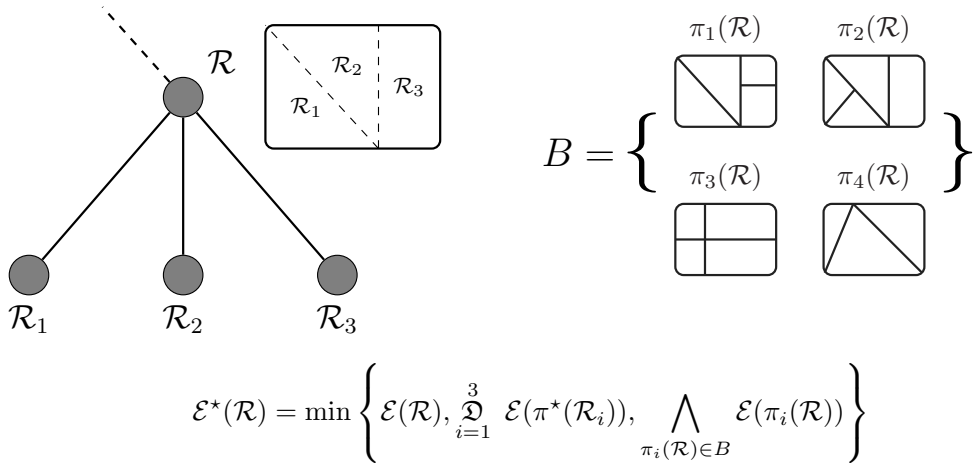


Fig. 3. Illustration of a step of the dynamic program (7) applied to a braid structure: one has to choose between $\{\mathcal{R}\}$, $\bigsqcup \pi^\star(\mathcal{R}_i)$ or any other $\pi_i(\mathcal{R}) \in B$. Note however that $\mathcal{R} \neq E$, otherwise B would not be a braid since $\pi_3(\mathcal{R}) \vee \pi_4(\mathcal{R}) = \mathcal{R}$.

the space of such functions. The output of the dynamic program is no longer some optimal cut for a given value of λ, but some partition of $\mathbb{R}^+$ into intervals $[0, \lambda_1[\cup [\lambda_1, \lambda_2[\cup \cdots \cup [\lambda_p, +\infty[$ where all λ values within a given interval $[\lambda_i, \lambda_{i+1}[$ are leading to the same optimal cut $\pi^\star_{\lambda_i}$. The reader is referred to [5] for more practical implementation details.

2.3 Minimization over a Braid

Braids of partitions have been recently introduced in [6] as a potential tool to combine multiple hierarchies and thus tackle segmentation of multimodal images, but these have not been yet investigated in practice in multimodal data fusion. Braids of partitions are defined as follows: a family of partitions $B = \{\pi_i\}$ is called a *braid* whenever there exists some hierarchy H_m, called *monitor hierarchy*, such that:

$$\forall\, \pi_i, \pi_j \in B, \pi_i \vee \pi_{j \neq i} \in \Pi_E(H_m) \backslash \{E\} \tag{6}$$

where $\pi_i \vee \pi_j$ denotes the refinement supremum, i.e. the smallest partition that is refined by both π_i and π_j. In other words, a braid is a family of partitions such that the refinement suprema of any pair of different partitions of the family are hierarchically organized, even though the partitions composing the braid might not be. For this reason, braids of partitions are more general than hierarchies of partitions: while hierarchies are braids, the converse is not necessarily true. It is also worth noting that the refinement supremum of any two partitions must differ from the whole image $\{E\}$ in (6). Otherwise, any family of arbitrary partitions would form a braid with $\{E\}$ as a supremum, thus loosing any interesting structure. The optimal cut of a braid of partitions is reached by solving the dynamic program (3) for every node $\mathcal{R}$ of the monitor hierarchy H_m, with a slight modification:

$$\mathcal{E}^\star(\mathcal{R}) = \min \left\{ \mathcal{E}(\mathcal{R}), \bigoplus_{r \in \mathsf{S}(\mathcal{R})} \mathcal{E}(\pi^\star(r)), \bigwedge_{\pi_i(\mathcal{R}) \in B} \mathcal{E}(\pi_i(\mathcal{R})) \right\} \tag{7}$$

In addition to comparing the node energy with respect to the optimal energy of its children, one has also to consider all other partial partitions of $\mathcal{R}$ that can be contained in the braid, since $\mathcal{R}$ represents the refinement supremum of some regions in the braid, and not those regions themselves. The optimal cut of $\mathcal{R}$ is then given by $\{\mathcal{R}\}$, the disjoint union of the optimal cuts of its children or some other partial partition of $\mathcal{R}$ contained in the braid, depending on which has the lowest energy. A step of this dynamic program is illustrated by Figure 3. Notice that the optimal cut of a braid B is obtained through an energy minimization procedure conducted on its monitor hierarchy H_m. However, this optimal cut may be composed of regions that are solely contained in the braid and therefore not supported by nodes of the monitor hierarchy (it would be the case in the example depicted by Figure 3 if $\pi_4(\mathcal{R})$ were for instance chosen to be the optimal cut of $\mathcal{R}$).

3 Proposed Methodology

3.1 Generation of a Braid from Multiple Hierarchies

The refinement supremum of two cuts of a hierarchy remains a cut of this hierarchy. For this reason, it is straightforward to compose a braid with cuts coming from the same hierarchy since any family of such cuts is a braid. It also implies in that case that the regions composing the corresponding monitor hierarchy are a subset of the regions composing the initial hierarchy. However, this guarantee is lost when one wants to compose a braid from cuts coming from multiple hierarchies: all those cuts must be sufficiently related to ensure that all their pairwise refinement suprema are hierarchically organized. As an example, let $B = \{H_1 = \{\pi_1^1 \geq \pi_1^2\}, H_2 = \{\pi_2^1 \geq \pi_2^2\}\}$ be some family of partitions composed of two supposedly independent hierarchies H_1 and H_2, both composed of two ordered cuts. B being a braid implies that all pairwise refinements suprema are hierarchically organized. In particular, this must be true for $\pi_1^1 \vee \pi_1^2 = \pi_1^1$ and $\pi_2^1 \vee \pi_2^2 = \pi_2^1$, which were initally assumed to come from independent hierarchies. Thus, the partitions composing B cannot be chosen arbitrarily. This leads us to introduce the property of *h-equivalence* (h standing here for *hierarchical*): two partitions π_a and π_b are said to be h-equivalent, and one notes $\pi_a \overset{h}{\simeq} \pi_b$ if and only if $\forall \mathcal{R}_a \in \pi_a$, $\forall \mathcal{R}_b \in \pi_b, \mathcal{R}_a \cap \mathcal{R}_b \in \{\emptyset, \mathcal{R}_a, \mathcal{R}_b\}$. In other words, π_a and π_b may not be globally comparable, but they are locally comparable (for instance, $\pi_1(\mathcal{R})$ and $\pi_2(\mathcal{R})$ of Figure 3 are not globally comparable, but they locally are). In particular, given a hierarchy H, $\forall \pi_1, \pi_2 \in \Pi_E(H)$, $\pi_1 \overset{h}{\simeq} \pi_2$: all cuts of a hierarchy are h-equivalent. $\overset{h}{\simeq}$ is a tolerance relation: it is reflexive and symmetric, but not transitive. Given some hierarchy H and a partition $\pi_* \in \Pi_E$, we denote by $H \overset{h}{\simeq} \pi_*$

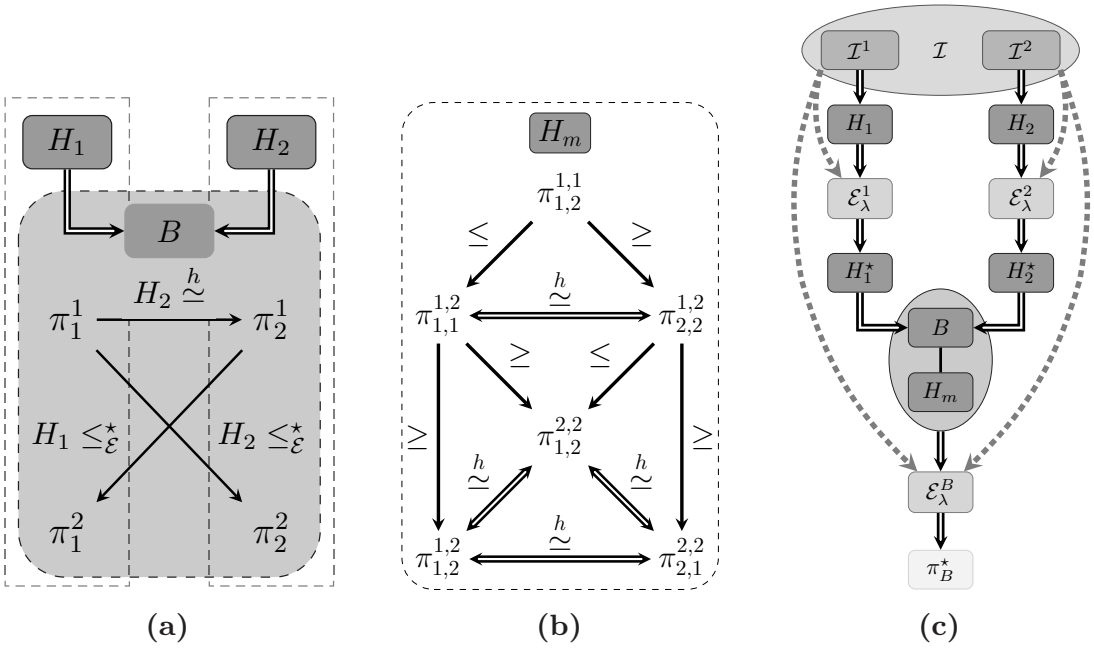


Fig. 4. a) Possible composition of a braid B with cuts from two hierarchies H_1 and H_2, b) cuts of the corresponding monitor hierarchy H_m, and c) workflow of proposed multimodal segmentation methodology.

the set of cuts of H that are h-equivalent to π_*. Obviously, $H \overset{h}{\simeq} \pi_* \subseteq \Pi_E(H)$ with equality if and only if $\pi_* \in \Pi_E(H)$. Provided some hierarchy H, some h-increasing and scale-increasing energy $\mathcal{E}_\lambda$ and some partition $\pi_* \in \Pi_E$, we also define $H \leq_{\mathcal{E}}^{\star} \pi_* = \{\pi \in \Pi_E(H^\star) \text{ s.t } \pi \leq \pi_*\}$ as the set of optimal cuts of H with respect to $\mathcal{E}_\lambda$ that are refinements of π_*. Following, it is possible to compose a braid B with cuts extracted from two hierarchies H_1 and H_2 using these two relations, as depicted in figure 4a: Given $\pi_1^1 \in \Pi_E(H_1)$, take some $\pi_2^1 \in H_2 \overset{h}{\simeq} \pi_1^1$. Then, π_1^2 and π_2^2 are taken in $H_1 \leq_{\mathcal{E}}^{\star} \pi_2^1$ and $H_2 \leq_{\mathcal{E}}^{\star} \pi_1^1$, respectively. In practice, we choose $\pi_2^1 = \bigvee\{H_2 \overset{h}{\simeq} \pi_1^1 \backslash \{E\}\}$, $\pi_1^2 = \bigvee\{H_1 \leq_{\mathcal{E}}^{\star} \pi_2^1\}$ and $\pi_2^2 = \bigvee\{H_2 \leq_{\mathcal{E}}^{\star} \pi_1^1\}$. Under this configuration, it is guaranteed that $B = \{\pi_1^1, \pi_1^2, \pi_2^1, \pi_2^2\}$ forms a braid with monitor hierarchy H_m whose cuts $\pi_{i,j}^{k,l} = \pi_i^k \vee \pi_j^l$ are organized as displayed by figure 4b. Other configurations for the composition of B may work as well.

3.2 Methodology

We now propose a methodology to perform multimodal image segmentation, using the previously introduced concept of braids of partitions to fuse the output of several hierarchies. The proposed method is illustrated by the workflow in figure 4c, detailed step by step in the following. Let $\mathcal{I} = \{\mathcal{I}^1, \mathcal{I}^2\}$ be a multimodal image, assumed to be composed of two modes $\mathcal{I}^1$ and $\mathcal{I}^2$ having the same spatial support E, for a matter of clarity (the extension to a greater number of modes follows the same scheme). First, two hierarchies H_1 and H_2 are built on $\mathcal{I}^1$

and $\mathcal{I}^2$, respectively. Two energies $\mathcal{E}_\lambda^1$ and $\mathcal{E}_\lambda^2$ are defined as piecewise constant Mumford-Shah energies [10] whose goodness-of-fit (GOF) term acts with respect to each mode $\mathcal{I}^1$ and $\mathcal{I}^2$, and whose regularization term is half the length of the region perimeter:

$$\mathcal{E}_\lambda^i(\pi) = \sum_{R\in\pi} \left(\Xi_i(\mathcal{R}) + \frac{\lambda}{2}|\partial\mathcal{R}| \right) \tag{8}$$

with $\Xi_i(\mathcal{R}) = \sum_{\mathbf{x}\in\mathcal{R}} \|\mathcal{I}^i(\mathbf{x}) - \boldsymbol{\mu}_i(\mathcal{R})\|_2^2$ being the GOF term acting on mode $\mathcal{I}^i$ and $\boldsymbol{\mu}_i(\mathcal{R})$ is the mean value/vector in mode $\mathcal{I}^i$ of pixels belonging to region $\mathcal{R}$. Piecewise constant Mumford-Shah energies are a popular choice when it comes to minimizing some energy function because of their ability to produce consistent segmentations. However, other types of energies could be investigated as well, depending on the underlying application. The only constraint here is that the energies $\mathcal{E}_\lambda^1$ and $\mathcal{E}_\lambda^2$ must be h-increasing and scale-increasing. It is known to be the case for Mumford-Shah energies [7]. Following, the two optimal hierarchies $H_1^\star$ and $H_2^\star$ are generated from the optimal cuts of H_1 and H_2 with respect to $\mathcal{E}_\lambda^1$ and $\mathcal{E}_\lambda^2$. The braid B is composed as described previously in subsection 3.1 and by figure 4a: a first partition $\pi_1^{1\star}$ is extracted from $H_1^\star$, and is used to extract two partitions $\pi_2^{1\star}$ and $\pi_2^{2\star}$ from $H_2^\star$ following the relations $\overset{h}{\simeq}$ and $\leq_{\mathcal{E}}^\star$, respectively. A second partition $\pi_1^{2\star}$ is finally extracted from $H_1^\star$ using $\leq_{\mathcal{E}}^\star$ and $\pi_2^{1\star}$. Eventually, B is composed of 4 partitions $\{\pi_1^{1\star}, \pi_1^{2\star}, \pi_2^{1\star}, \pi_2^{2\star}\}$ extracted from the two hierarchies $H_1^\star$ and $H_2^\star$, and the braid structure is guaranteed, allowing to construct the monitor hierarchy H_m. A last energy term $\mathcal{E}_\lambda^B$ is defined, relying on both modes of the multimodal image $\mathcal{I}$:

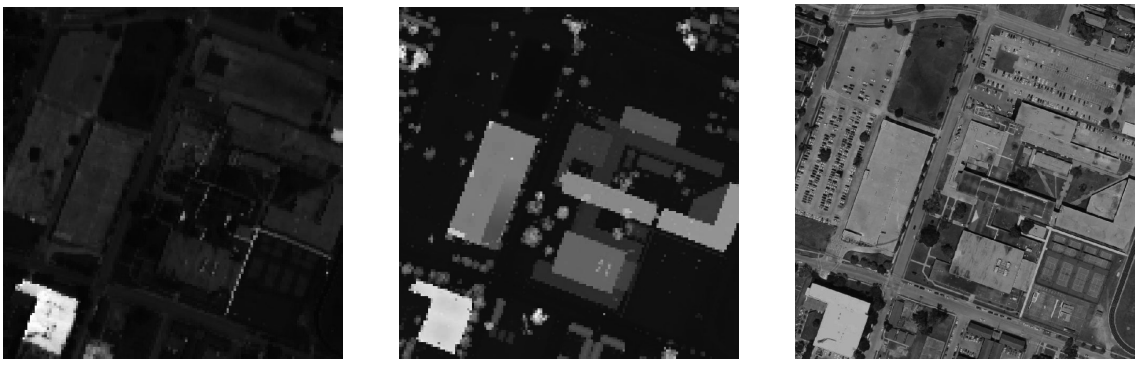
$$\mathcal{E}_\lambda^B(\pi) = \sum_{\mathcal{R}\in\pi} \left(\max\left(\frac{\Xi_1(\mathcal{R})}{\Xi_1(\mathcal{I}^1)}, \frac{\Xi_2(\mathcal{R})}{\Xi_2(\mathcal{I}^2)} \right) + \frac{\lambda}{2}|\partial\mathcal{R}| \right) \tag{9}$$

The GOF term of each region $\mathcal{R}$ is now defined as the maximum with respect to both modes of the normalized GOFs. The normalization allows both GOF terms to be in the same dynamical range. $\mathcal{E}_\lambda^B$ is also a h-increasing and scale-increasing energy. Its minimization over H_m and B following the dynamic program (7) gives some optimal segmentation $\pi_B^\star$ of $\mathcal{I}$, which should contain salient regions shared by both modes as well as regions exclusively expressed by $\mathcal{I}^1$ and $\mathcal{I}^2$.

4 Results

4.1 Conducted Experiments

We apply the proposed methodology on the multimodal data set described in [4] composed of a hypespectral (HS) image $\mathcal{I}^1$ of 144 spectral bands evenly spaced between 380 nm and 1050 nm, and a LiDAR-derived digital surface model (DSM) $\mathcal{I}^2$, with the same ground-sampling distance of 2.5 m. Data were acquired over the University of Houston campus. The study site features an urban area with several buildings of various heights and made of different materials, some parking



(a) **(b)** **(c)**

Fig. 5. a) RGB composition of the hyperspectral image, b) corresponding LiDAR-derived DSM, c) very-high resolution RGB image of the same site

lots, tennis courts, roads and some portions of grass and trees. A RGB composition of the hyperspectral image is displayed in figure 5a, and the corresponding LiDAR-derived DSM is shown in figure 5b. It is also shown for visualization purpose a very-high resolution RGB image of the scene in figure 5c[1].

Hierarchies H_1 and H_2 are obtained by building two binary partition trees [14] on $\mathcal{I}^1$ and $\mathcal{I}^2$ with standard parameters (mean spectrum and spectral angle for the region model and merging criterion of the HS mode, mean value and Euclidean distance for the DSM). Both hierarchies are built on the same inital partition π_0, obtained as the refinement infimum of two mean shift clustering procedures conducted on the RGB composition of the HS and on the DSM. The braid B is constructed following the procedure exposed in figure 4a: $\pi_1^{1\star}$ is the first cut extracted from $H_1^\star$ and contains around 125 regions. It is used to extract $\pi_2^{1\star}$ and $\pi_2^{2\star}$ from $H_2^\star$, which comprise 342 and 349 regions, respectively. Finally, $\pi_1^{2\star}$ is extracted from $H_1^\star$ using $\pi_2^{1\star}$ and contains 379 regions. The four partitions composing B generate $\binom{4}{2} = 6$ cuts of the monitor hierarchy H_m, which is built by re-organizing those cuts in a hierarchical manner. The leaf partition of H_m, denoted π_0^B, is obtained as $\bigwedge\{\pi_i \vee \pi_{j\neq i}, \pi_i, \pi_j \in B\}$. Finally, the minimization of $\mathcal{E}_\lambda^B$ over H_m, following (7), is conducted with λ being empirically set to 5.10^{-5}, and produces an optimal segmentation $\pi_B^\star$ of the braid composed of 302 regions. To evaluate the improvements brought by the braid structure, we propse to extract from $H_1^\star$ and $H_2^\star$ the two optimal cuts $\pi_1^\star$ and $\pi_2^\star$ that have the same (or a close) number of regions as $\pi_B^\star$ (in practice, $\pi_1^\star$ and $\pi_2^\star$ have 301 and 302 regions, respectively). This should allow a fair visual comparison since all three partitions should feature regions of similar scales. In addition, we compute for the partitions $\pi_0, \pi_0^B, \pi_1^\star, \pi_2^\star$ and $\pi_B^\star$ their average GOF with respect to both modes $\mathcal{I}^1$ and $\mathcal{I}^2$ as follows:

$$\epsilon(\pi|\mathcal{I}^i) = \frac{1}{|E|} \sum_{\mathcal{R}\in\pi} |\mathcal{R}| \times \Xi_i(\mathcal{R}) \tag{10}$$

[1] https://goo.gl/maps/VVXE6

with $|\mathcal{R}|$ denoting the number of pixels in region $\mathcal{R}$, and $\Xi_i(\mathcal{R})$ is the Mumford-Shah GOF term defined in equation (8).

4.2 Results

Table 1 presents the number of regions as well as the average GOF of leaf partitions π_0 and π_0^B, and of optimal partitions $\pi_1^\star, \pi_2^\star$ and $\pi_B^\star$ with respect to both

Table 1. Number of regions and average GOF of leaf partitions π_0, π_0^B and optimal partitions $\pi_1^\star, \pi_2^\star, \pi_B^\star$ with respect to both modes $\mathcal{I}^1$ and $\mathcal{I}^2$

	π_0	π_0^B	$\pi_1^\star$	$\pi_2^\star$	$\pi_B^\star$
$\lvert\pi\rvert$	416	354	301	302	302
$\epsilon(\pi\lvert\mathcal{I}^1)$	13.2	16.6	16.0	57.2	19.8
$\epsilon(\pi\lvert\mathcal{I}^2)$	262.2	297.7	611.7	413.0	358.8

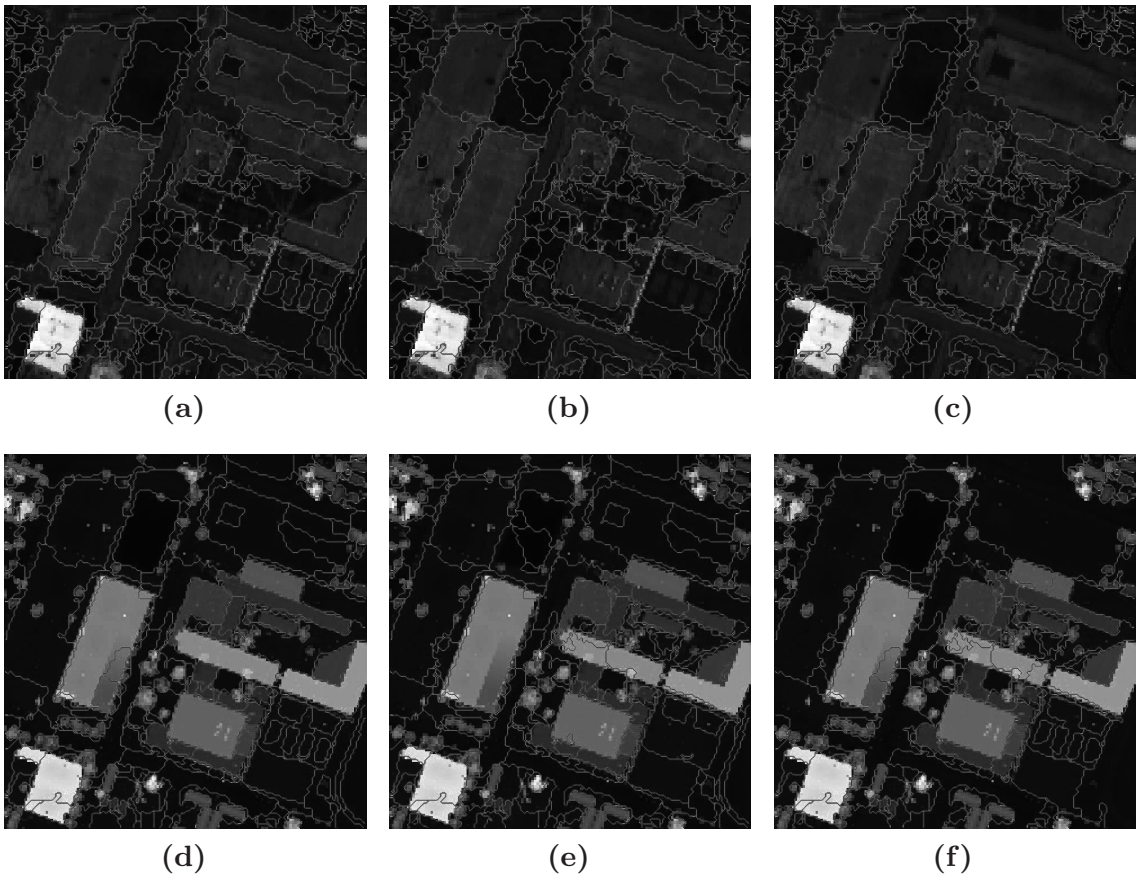


Fig. 6. Top row: optimal partitions a) $\pi_B^\star$, b) $\pi_1^\star$ (optimal with respect to the HS image) and c) $\pi_2^\star$ (optimal with respect to the DSM) superimposed over the HS image. Bottom row: optimal partitions d) $\pi_B^\star$, e) $\pi_1^\star$ and f) $\pi_2^\star$ superimposed over the DSM image.

modes $\mathcal{I}^1$ and $\mathcal{I}^2$. Its analysis demonstrates the interest of the proposed methodology using the braid structure. One can indeed remark, not surprinsigly, that $\pi_1^\star$ and $\pi_2^\star$ score a low average GOF value with respect to their corresponding mode, but a greater average GOF with respect to the complementary mode. On the other hand, $\pi_B^\star$ outperforms $\pi_1^\star$ with respect to $\mathcal{I}^2$ while scoring a similar value for $\mathcal{I}^1$, and outperforms $\pi_2^\star$ both with respect to $\mathcal{I}^1$ and $\mathcal{I}^2$. Thus, $\pi_B^\star$ better fits both modes of the multimodal image at the same time. In addition, it contains fewer regions than π_0 and π_0^B while not increasing the average GOF too much. Therefore, $\pi_B^\star$ decreases over-segmentation compared to the two leaf partitions while maintaining comparable GOF values. Figure 6 shows the optimal partitions $\pi_B^\star$, $\pi_1^\star$ and $\pi_2^\star$ superimposed over the RGB composition of the HS image (top row, from figure 6a to 6c) and over the DSM image (bottom row, from figure 6d to 6f). The qualitative analysis of figure 6 leads to similar conclusions : while $\pi_1^\star$ tends to under-segment regions featuring the same spectral properties but not the same elevation (typically, the buildings in the center of the scene), those regions are correctly segmented in $\pi_B^\star$. Similary, regions at the same elevation are often under-segmented in $\pi_2^\star$ even if they are made of different materials (parking lots, roads and grass for instance) but correctly delineated in $\pi_B^\star$. This demonstrates how the construction of the braid and associated monitor hierarchy as well as the following energy minimization were able to fuse the information contained in both modes to produce a more accurate segmentation of the multimodal image.

5 Conclusion

In conclusion, we presented in this article a new method to perform multimodal segmentation, based on the hierarchical minimization of some energy function. In particular, we used the recently introduced concept of braids of partitions and associated monitor hierarchies and we adapted to them the dynamic program procedure conducted to perform energy minimization over hierarchies. The proposed framework was investigated over a multimodal image composed of a hyperspectral and a LiDAR mode. Results demonstrated, quantitatively and qualitatively, the ability of the proposed approach to produce a segmentation that not only retains salient regions shared by both modes, but also regions appearing in only one mode of the multimodal image.

Future work include a deeper investigation on the way to compose a braid with cuts coming from several hierarchies, and a more thorough assessment analysis of the improvements brought by the proposed methodology.

Acknowledgement. This work was partially funded through the ERC CHESS project, ERC-12-AdG-320684-CHESS.

References

1. Bießmann, F., Plis, S., Meinecke, F., Eichele, T., Muller, K.: Analysis of multimodal neuroimaging data. QIEEE Reviews in Biomedical Engineering 4, 26–58 (2011)

2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11), 1222–1239 (2001)
3. Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J.: Challenges and opportunities of multimodality and data fusion in remote sensing. In: 2013 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), pp. 106–110. IEEE (2014)
4. Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Bellens, W.L.R., Pizurica, A., Gautama, S., Philips, W., Prasad, S., Du, Q., Pacifici, F.: Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6), 2405–2418 (2014)
5. Guigues, L., Cocquerez, J., Le Men, H.: Scale-sets image analysis. International Journal of Computer Vision 68(3), 289–317 (2006)
6. Kiran, B.: Energetic-Lattice based optimization. PhD thesis, Université Paris-Est, Paris (October 2014)
7. Kiran, B., Serra, J.: Global–local optimizations by hierarchical cuts and climbing energies. Pattern Recognition 47(1), 12–24 (2014)
8. Lahat, D., Adaly, T., Jutten, C.: Challenges in multimodal data fusion. In: 2013 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), pp. 101–105. IEEE (2014)
9. Li, S.: Markov random field modeling in computer vision. Springer-Verlag New York, Inc. (1995)
10. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics 42(5), 577–685 (1989)
11. Ronse, C.: Partial partitions, partial connections and connective segmentation. Journal of Mathematical Imaging and Vision 32(2), 97–125 (2008)
12. Soille, P.: Constrained connectivity for hierarchical image partitioning and simplification. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(7), 1132–1145 (2008)
13. Tarabalka, Y., Tilton, J., Benediktsson, J., Chanussot, J.: A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(1), 262–272 (2012)
14. Valero, S., Salembier, P., Chanussot, J.: Hyperspectral image representation and processing with binary partition trees. IEEE Transactions on Image Processing 22(4), 1430–1443 (2013)

Multi-image Segmentation: A Collaborative Approach Based on Binary Partition Trees

Jimmy Francky Randrianasoa[1(✉)], Camille Kurtz[2], Éric Desjardin[1], and Nicolas Passat[1]

[1] Université de Reims Champagne-Ardenne, CReSTIC, France
jimmy.randrianasoa@univ-reims.fr
[2] Université Paris-Descartes, LIPADE, France

Abstract. Image segmentation is generally performed in a "one image, one algorithm" paradigm. However, it is sometimes required to consider several images of a same scene, or to carry out several (or several occurrences of a same) algorithm(s) to fully capture relevant information. To solve the induced segmentation fusion issues, various strategies have been already investigated for allowing a consensus between several segmentation outputs. This article proposes a contribution to segmentation fusion, with a specific focus on the "*n* images" part of the paradigm. Its main originality is to act on the segmentation research space, *i.e.*, to work at an earlier stage than standard segmentation fusion approaches. To this end, an algorithmic framework is developed to build a binary partition tree in a collaborative fashion, from several images, thus allowing to obtain a unified hierarchical segmentation space. This framework is, in particular, designed to embed consensus policies inherited from the machine learning domain. Application examples proposed in remote sensing emphasise the potential usefulness of our approach for satellite image processing.

Keywords: Segmentation fusion · Morphological hierarchies · Multi-image · Collaborative strategies · Binary partition tree · Remote sensing

1 Introduction

In image processing / analysis, segmentation is a crucial task. The concept of segmentation is also quite generic, in terms of semantics (from low-level definition of homogeneous areas to high-level extraction of specific objects), in terms of definition (object versus background, or total partition of the image support), and in terms of algorithmics.

The principal invariant of segmentation is the "one image, one algorithm" paradigm. Indeed, for a given application, a specific algorithm is generally chosen (or designed) with respect to its adequacy with the considered image processing / analysis problem; parametrized with respect to the physical / semantic properties of the targeted images; and then applied (once) on a given image, or on each image of a given dataset.

Since segmentation is an ill-posed problem, it is plain that the result of a segmentation algorithm applied on an image cannot be completely satisfactory. Based on this

This research was partially funded by the French *Agence Nationale de la Recherche* (Grant Agreements ANR-10-BLAN-0205 and ANR-12-MONU-0001).

J.A. Benediktsson et al. (Eds.): ISMM 2015, LNCS 9082, pp. 253–264, 2015.
DOI: 10.1007/978-3-319-18720-4_22

assertion, it is sometimes relevant to relax the "one image, one algorithm" paradigm, either by operating one algorithm on several images of a same scene – in order to enrich / improve the information provided as input – or by applying several algorithms, or the same algorithm with several parameter sets – in order to enrich / improve the information provided as output.

In the literature, these "*n* images, one algorithm" and "one image, *n* algorithms" paradigms are dealt with under the common terminology of *segmentation fusion*. Most of the time, the strategy consists of computing *n* segmentation maps, and to develop a consensus strategy to gather and unify these *n* results into a single, assuming that an accurate global result will emerge from a plurality of less accurate ones.

This fusion strategy has also been proposed – and intensively studied – in the research field of machine learning, especially in clustering [1]. Intrinsically, the problems are the same, with the only – but important – difference that image segmentation, by contrast with general data clustering, implies the handling of the spatial organisation of the processed data.

In this article, we propose a new approach for segmentation fusion, that focuses on the "*n* images, one algorithm" paradigm. The relevance and novelty of this approach, that lies in the framework of morphological hierarchies and connected operators [2], derives from the early stage where the fusion occurs. By contrast with most segmentation fusion approaches, we do not intend to fuse several segmentation maps obtained from several images, but we directly develop a consensus when defining the research space, *i.e.*, during the construction of a hierarchy that models the *n* images.

This strategy, that acts at the research space level, induces two important side effects. First, it allows us to rely on hierarchical approaches for segmentation, and thus to propose a rich and versatile segmentation framework, that can be easily instantiated according to the application field. Second, by operating the fusion on the internal data-structures involved in the algorithmic construction of the hierarchy, instead of spatial regions of the segmentation maps, we can directly benefit from all the strategies devoted to "non-spatial" data fusion, previously proposed by the machine learning community.

Our approach relies on Binary Partition Trees (BPTs) [3] as hierarchical model. This choice is motivated both by the possibility to tune a BPT construction, by contrast with other tree structures more deterministically deriving from the images (namely, component-trees or trees of shapes). It is also justified by the frequent and successful application of BPTs in remote sensing, where the use of multiple images of a same scene is frequent, such as the use of fusion strategies for clustering issues.

This article is organized as follows. In Section 2, a brief overview of related works is proposed, describing previous segmentation fusion strategies, and applications of BPTs and hierarchies for analysing multiple remote sensing images. In Section 3, the general data-structure / algorithmic framework for building a BPT from several images is described, and various consensus strategies for instantiating this framework are discussed. In Section 4, experiments are performed on remote sensing images, in order to illustrate the relevance and usefulness of BPT segmentation from several images, by considering two application cases, namely mono-date multi-imaging on urban areas, and multi-date imaging on agriculture areas. In Section 5, a discussion concludes this article by emphasising perspectives and potential further works.

2 Related Works

2.1 Segmentation Fusion

Segmentation fusion consists of establishing a consensus between several segmentation maps. The taxonomy of segmentation fusion is directly mapped on that of segmentation.

On the one hand, segmentation can be viewed as a process that aims to extract one structure (the object) versus the remainder of the image (the background). Typical examples of such segmentation strategies are deformable models, graph-cuts, *etc.* In this context, segmentation fusion can be interpreted as a geometrical problem of shape or contour averaging / interpolation (see, *e.g.*, [4,5]). This subfamily of segmentation fusion methods, mainly used in medical imaging, is out of the scope of our study.

On the other hand, segmentation can be viewed as a process that aims to define a partition of the whole image, in order to extract meaningful homogeneous areas. Examples of such segmentation strategies are watersheds, split-and-merge and, more generally, connected operators. Clustering methods also enter in this category, with the difference that they generally do not take into account the spatial organisation of the image data.

Following a machine learning vision, some methods handled segmentation fusion via clustering ensemble [1]; a comparative study can be found in [6]. It was also proposed to interpret the information gathered at each pixel in the various segmentation maps as feature vectors then involved in optimization procedures. In particular, a two-stage K-means approach was considered in [7], or probabilistic frameworks in [8,9].

Other approaches explicitly took into account the spatial organisation of the segmentation maps. In pioneering works [10], images of a same scene obtained from various modalities were merged and optimized based on edge information. Later, statistical analysis of the co-occurrence probability of neighbour pixels was considered to improve the accuracy of a partition from several versions with slightly disturbed borders [11].

More recently, connected operators were also involved. Stochastic watersheds [12] were introduced as a solution for improving the robustness of marker-based watersheds, by fusing the results obtained from different initial seeds. This approach further inspired a stochastic minimum spanning forest approach [13] for segmentation / classification of hyperspectral remote sensing images. Random walkers [14] were also considered for segmenting a graph generated from the information derived from the degree of accordance between different segmentation maps.

2.2 Morphological Hierarchies, Multi-images and Remote Sensing

Morphological hierarchies associated to connected operators [2] have been successfully involved in image segmentation in the “one image, one algorithm” paradigm. In the associated tree-structures, the nodes model homogeneous regions in the image whereas the edges represent their inclusion relations.

Classical trees, as component-trees [15] or trees of shapes [16], allow us to perform hierarchical segmentation by fusion of flat zones. These structures provide as output partial partitions of an image with tunable levels of details. However, they strongly rely on the image intensity, which is not compliant with the specificities of satellite images.

A first solution to deal with this issue relies on the constrained connectivity [17]. The connectivity relation generates a partition of the image domain; fine to coarse partition

hierarchies are then produced by varying a threshold value associated with each connectivity constraint. In a different manner, the BPT [3] reflects a (chosen) similarity measure between neighbouring regions, and models the hierarchy between these regions. The BPTs were used to segment various types of satellite images [18,19]. However, objects in satellite images appear often too much heterogeneous to be correctly segmented from a single image. It appears then relevant to consider n images of a same scene to enrich the data space and improve the capability of segmentation hierarchies.

In this context, efforts were conducted to extend morphological hierarchies for handling n images. In [20] an extension of the BPT model was proposed to deal with multiresolution satellite images by considering one hierarchy per resolution image. In a similar vein, an approach based on multiple morphological hierarchies was developed in [21] to segment a multispectral image. The originality of this approach was to build independently one hierarchy per radiometric band and combine them to select meaningful segments, optimizing a mixed spectral / connectivity measure.

These recent works show the interest of considering a multi-image paradigm to enhance the hierarchical segmentation of complex structures from remote sensing images. Multi-image fusion segmentation methods take advantage of the complementarity of available data, and can be adapted for morphological hierarchies by interpreting the "n images, one algorithm" as "n images, one hierarchy". By performing collaborations where the fusions occur relatively to the different image contents, it is then possible to gradually build a unique consensual hierarchy, which can be used to detect complex patterns while avoiding (spectral, semantic, ...) noise appearing in a single image.

Based on these considerations, we propose in the next section a hierarchical collaborative segmentation method, extending the BPT to deal with multi-images. By contrast with classical segmentation fusion approaches, we do not intend to fuse several segmentation maps, but we directly develop a consensus during the construction of the hierarchy that models the n images. To build this consensual hierarchy, we propose different algorithmic consensus strategies (inspired from ensemble clustering strategies previously developed by the machine learning community) that seek widespread agreement among the content of different images of a same scene.

3 Building a Binary Partition Tree from Several Images

We first recall the BPT construction algorithm. In particular, we focus on the data-structure point of view, which is the cornerstone of our contribution. Then, we describe our generalization of this algorithm to deal with several images. Various families of consensus strategies for instantiating this framework are finally discussed.

3.1 The Standard BPT Construction

Summary of the Algorithm [3]. A BPT is a hierarchical representation of an image. More precisely, it is a binary tree, whose each node is a connected region. Each of these nodes is either a leaf – then corresponding to an "elementary" region – or an internal node, modelling the union of the regions of its two children nodes. The root is the node

corresponding to the support of the image. Practically, a BPT is built from its leaves – provided by an initial partition of the image support – to its root, in a bottom-up fashion, by iteratively choosing and merging two adjacent regions which minimize a merging criterion that reflects, *e.g.*, the spectral and / or geometrical likenesses of the regions.

Structural Description of the Algorithm. An image is a function $I : \Omega \to V$ that associates to each point x of the finite set Ω a value $I(x)$ of the set V. A crucial point for building a BPT – and more generally for performing connected filtering – is to take into account the structure of Ω. More precisely, it is mandatory to model the fact that two points x and y of Ω are neighbours. This is done by defining an adjacency A_Ω, *i.e.*, an irreflexive, symmetric, binary relation on Ω. In other words, (Ω, A_Ω) is a graph that models the structure of the image space.

Let us now consider an initial partition $\mathcal{L}$ of Ω. (Each node $L \subseteq \Omega$ of $\mathcal{L}$ is generally assumed to be connected with respect to A_Ω.) This partition $\mathcal{L}$ defines the set of the leaves of the BPT we are going to build (*e.g.*, $\mathcal{L}$ can be the set of the image flat zones).

For any partition $\mathcal{P}$ of Ω (and in particular for $\mathcal{L}$) we can define an adjacency inherited from that of Ω. More precisely, we say that two distinct nodes $N_1, N_2 \in \mathcal{P}$ are adjacent if there exist $x_1 \in N_1$ and $x_2 \in N_2$ such that (x_1, x_2) is an edge of A_Ω, *i.e.*, x_1 and x_2 are adjacent in (Ω, A_Ω). This new adjacency relation $A_\mathcal{P}$ is also irreflexive and symmetric. In the case of $\mathcal{L}$, it allows us to define a graph $\mathfrak{G}_\mathcal{L} = (\mathcal{L}, A_\mathcal{L})$ that models the structure of the partition of the image I.

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

From a structural point of view, the BPT $\mathfrak{T}$ is the Hasse diagram of the partially ordered set $(\bigcup_{i=0}^n \mathcal{N}_i, \subseteq)$. From an algorithmic point of view, $\mathfrak{T}$ is built in parallel to the progressive collapse from $\mathfrak{G}_0$ to $\mathfrak{G}_n$; in other words, $\mathfrak{T}$ stores the node fusion history. More precisely, we define a sequence $(\mathfrak{T}_i)_{i=0}^n$ as follows. We set $\mathfrak{T}_0 = (\mathcal{N}_0, \emptyset) = (\mathcal{L}, \emptyset)$. Then, for each i from 1 to n, we build $\mathfrak{T}_i$ from $\mathfrak{T}_{i-1}$ by adding the new node $N_{i-1} \cup N'_{i-1}$, and the two edges $(N_{i-1} \cup N'_{i-1}, N_{i-1})$ and $(N_{i-1} \cup N'_{i-1}, N'_{i-1})$. The BPT $\mathfrak{T}$ is finally defined as $\mathfrak{T}_n$.

Remark. The classical – image and application-oriented – description of the BPT construction algorithm considers as input: the image I (*i.e.*, the geometrical embedding of Ω, and the value associated to each point of Ω); a region model, that allows us to "describe" the nodes; and a merging criterion, that allows us to quantify the homogeneousness of nodes before and after a putative fusion. These information are important from an applicative point of view. However, from an algorithmic point of view, their only use is to define a valuation on the edges that allows us to *choose* which nodes to fuse at any

given step. In the sequel, we will then consider – without loss of correctness – that a BPT is fully defined by only two input information: (1) the graph $\mathfrak{G}_{\mathcal{L}} = (\mathcal{L}, A_{\mathcal{L}})$ that models the initial partition of the image; (2) a valuation function $W : (2^{\Omega})^2 \times V^{\Omega} \to \mathbb{R}$ that allows us to choose, at each step of the process, the next pair of nodes to be merged.

Data-Structures. The above description of the BPT construction algorithm implies to define and update, during the whole process, several data-structures, namely: the graph $\mathfrak{G}$, that allows us to know what nodes remain to be merged and what are their adjacency links; and the tree $\mathfrak{T}$ that is progressively built. In order to efficiently compute the valuation W, it is also important to associate each node of $\mathfrak{G}$ to the corresponding part of the image I, *e.g.*, via a mapping between $\mathfrak{G}$ (actually, $\mathcal{N}$) and Ω.

The last – but not least – required data-structure is a sorted list $\mathcal{W}$ that gathers the valuations of each remaining edge of $\mathfrak{G}$. This list contains the information that will authorise, at each of the n iterative steps of the process, to choose the couple of nodes to be merged. This choice is made in constant time $O(1)$, since $\mathcal{W}$ is sorted. After the merging operation, $\mathcal{W}$ has to be updated: (1) to remove the edge between the two nodes; (2) to update the edges affected by the merging operation; and (3) to re-order these updated edges. Operation (1) is carried out in constant time $O(1)$. Operation (2) is carried out in $O(\alpha.T_W)$, where T_W is the cost of the computation of W for an edge, and α is the number of neighbours of the merged nodes (α is generally bounded by a low constant value). Operation (3) is carried out in $O(\alpha.\log_2|\mathcal{W}|)$.

3.2 Generalizing the BPT Construction to Several Images

Let us now consider $k > 1$ images I_j, instead of one. Assuming that they correspond to a same scene, we consider – up to resampling – that all are defined on the same support. We then have a set of images $\{I_j : \Omega \to V_j\}_{j=1}^{k}$. The purpose is now to build a BPT from these k images, by generalizing the algorithmic framework described in Section 3.1. A step of this algorithmic process is illustrated in Figure 1.

Structural Evolutions. As stated above, we first need a graph that models the initial partition $\mathcal{L}$ of the image(s). Since all the I_j share the same support Ω, such a graph can still be obtained easily, either by subdividing Ω into one-point singleton sets – the induced graph $\mathfrak{G}_{\mathcal{L}}$ is then isomorphic to (Ω, A_{Ω}) – or by considering flat zones, *e.g.*, maximal connected sets of constant value with respect to the Cartesian space $\Pi_{j=1}^{k} V_j$.

The k images I_j share the same support, but they take their values in different sets V_j. As a consequence, following the standard BPT construction paradigm, each of them is associated with a specific valuation function $W_j : (2^{\Omega})^2 \times V_j^{\Omega} \to \mathbb{R}$ that is defined in particular with respect to the value set V_j.

From a data-structure point of view, the generalized BPT construction algorithm will still handle one graph $\mathfrak{G}$, that will be progressively collapsed; and one tree $\mathfrak{T}$ that will be built to finally provide the BPT. A unique mapping between $\mathcal{N}$ and Ω will still allow to have access to the values of a node for the k images. The main difference now lies

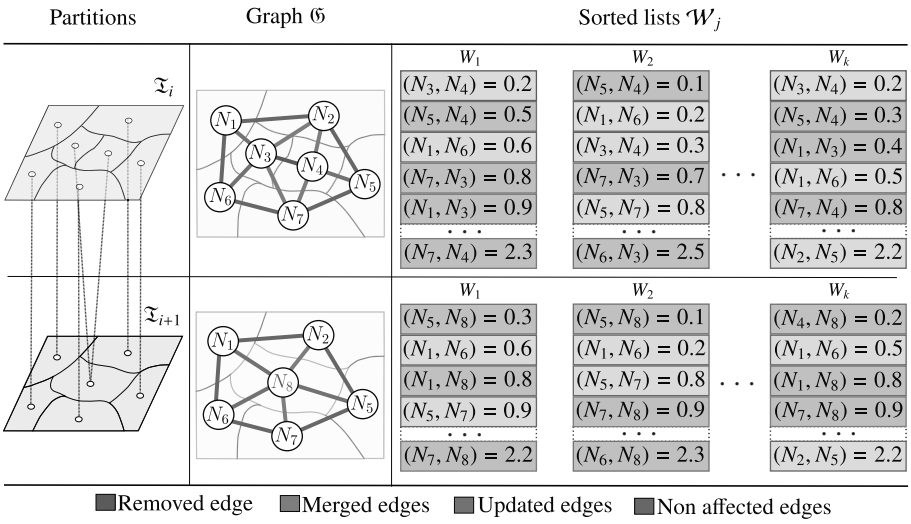


Fig. 1. One step of the building of a BPT from k images. Left: the partition of Ω before and after the fusion of two nodes. Center: the associated graph $\mathfrak{G}$, before and after the fusion of N_3 and N_4, forming the new node N_8. The red edge is removed. The blue and orange edges are updated, *e.g.*, (N_1, N_3) becomes (N_1, N_8); the orange are merged by pairs, *e.g.*, (N_7, N_3) and (N_7, N_4) become (N_7, N_8). The green edges are not affected. Right: the k lists $\mathcal{W}_j$ each corresponding to an image. The red cells are removed, as the edge (N_3, N_4) is suppressed; this edge had been chosen according to a given consensus policy, due to its "high" position in the k lists. The scores of blue and orange cells are updated with respect to N_8; the orange cells are merged by pairs. The positions of the blue and orange cells are updated with respect to their new scores. The scores of the green cells are not affected.

into the fact that each function W_j induces – in first approximation – a specific sorted list $\mathcal{W}_j$ to gather the valuation of the remaining edges of $\mathfrak{G}$, with respect to W_j.

Algorithmic Consequences. From an algorithmic point of view, each iteration of the construction process preserves the same structure. An edge is chosen and the two incident nodes of the graph are merged. This operation leads to update the nodes and edges of $\mathfrak{G}$, and adds a new node plus two edges in $\mathfrak{T}$. The main differences are that: (*i*) several sorted lists then have to be updated instead of one; and (*ii*) the choice of the optimal edge has to be made with respect to the information carried by these k sorted lists instead of only one, for a standard BPT.

From a computational point of view, choosing the edge to remove is no longer a constant time operation, but will depend on the way information are used and compared. Afterwards, operations (1–3) described in the standard BPT construction algorithm, for the sorted list maintenance, have to be duplicated for each list. These operations are then carried out in $O(k)$, $O(k.\alpha.T_{W_\star})$ and $O(k.\alpha.\log_2|\mathcal{W}_\star|)$, respectively.

However, this initial generalization of the BPT construction algorithm can be refined by studying more precisely the policies that are considered to choose an edge, with respect to the information carried by the W_j valuation functions and / or the $\mathcal{W}_j$ sorted lists.

3.3 Consensus Strategies

At each iteration, the choice of the optimal edge to remove, leading to the fusion of its two incident nodes, depends on a consensus between the information of the k images. For each image I_j, useful information are carried, on the one hand, by the valuation function $W_j : (2^{\Omega})^2 \times V^{\Omega} \rightarrow \mathbb{R}$ that gives an *absolute* value to each edge and, on the other hand, by the sorted list $\mathcal{W}_j$, that gives a *relative* information on edges, induced by their ordering with respect to W_j. These information are of distinct natures; we study their relevance according to the kinds of considered consensus policies.

Absolute Information Consensus. Let us consider that the consensus policy consists of choosing the edge of lowest mean valuation among the k images, or the edge of minimal valuation among all images. The first consensus (namely *min of mean*) is defined by a linear formulation: $\arg_{(N,N')\in\mathcal{N}} \min \sum_{j=1}^{k} W_j((N,N'))$, while the second (namely *min of min*) is defined by a non-linear formulation: $\arg_{(N,N')\in\mathcal{N}} \min \min_{j=1}^{k} W_j((N,N'))$. However, in both cases the decision is made by considering the absolute information carried by the edges. In other words, it is sufficient to know the k values of each point of Ω with respect to the images I_j. Then, the k sorted lists $\mathcal{W}_j$ are useless, and a single sorted list $\mathcal{W}$ that contains the information of these – linear or non-linear – formulations is indeed sufficient. The construction of a BPT from k images is then equivalent to that from one image defined as $I : \Omega \rightarrow \Pi_{j=1}^{k} V_j$.

Relative Local Information Consensus. Let us now consider that the consensus policy consists of choosing the edge that is the most often in first position in the k sorted lists $\mathcal{W}_j$, or the most frequently present in the $r \ll |\mathcal{W}_\star|$ first positions in the k sorted lists $\mathcal{W}_j$. These consensus (namely, *majority vote* and *most frequent*, potentially weighted) policies do not act on the absolute valuations of the edges, but on their relative positions in the lists. In such case, it is then mandatory to maintain k sorted lists. However, the decision process does not require to explicitly access the whole lists, but it can be restricted to the first (or the first r) element(s) of each, leading to a *local* decision process.

Relative Global Information Consensus. Let us finally consider that the consensus policy consists of choosing the edge that has the best global ranking among the k sorted lists $\mathcal{W}_j$. Such consensus (*e.g.*, *best average*, or *best median ranking*) policy, also acts on the relative positions of the edges in the lists. By contrast with the above case, the decision process requires to explicitly access to the whole content of all these lists, leading to a *global* decision process of high computation cost. (Such cost may be reduced by maintaining, in favourable cases, a $(k+1)$-th list that summarises the global information, and / or by adopting heuristic strategies that update the lists only after a given number of steps).

Algorithmic and Structural Consequences. The choice of a consensus strategy is strongly application-dependent. As a consequence, it is important to consider a trade-off between the structural and computational cost of the approach versus the benefits in terms of results accuracy. In particular, these costs are summarized in Table 1.

Table 1. Space and time cost of the BPT construction for various families of consensus policies. For the sake of readability, r, α and $T_{W_\star}$, which are practically bounded by low constant values have been omitted here.

Consensus policies	$\#\mathcal{W}_\star$	Edge choice	Edge removal	Edges update	Edges sorting
Absolute information	1	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$	$\Theta(\log_2 \lvert\mathcal{W}_\star\rvert)$
Relative local inf.	k	$\Omega(k)$	$\Theta(k)$	$\Theta(k)$	$\Theta(k.\log_2 \lvert\mathcal{W}_\star\rvert)$
Relative global inf.	k	$\Omega(k.\lvert\mathcal{W}_\star\rvert)$	$\Theta(k)$	$\Theta(k)$	$\Theta(k.\log_2 \lvert\mathcal{W}_\star\rvert)$

4 Experiments

To experiment our framework, two applications have been considered in the context of remote sensing: one-time, one-sensor, several (noisy) images, to assess the ability to retrieve information despite image degradation; and multi-time, one-sensor, one image per date, to assess the ability to capture time-independent and redundant information.

The BPT construction and segmentation approaches were voluntarily chosen as very simple, in order to avoid any bias related to these choices, thus better focusing on the actual structural effects of multi-image BPT versus standard BPT.

At this stage, these experiments have to be considered as toy-examples, since neither quantitative validation nor fine parameter tuning were carried out. Our purpose is mainly to give the intuition of potential uses of such BPTs in the field of remote sensing.

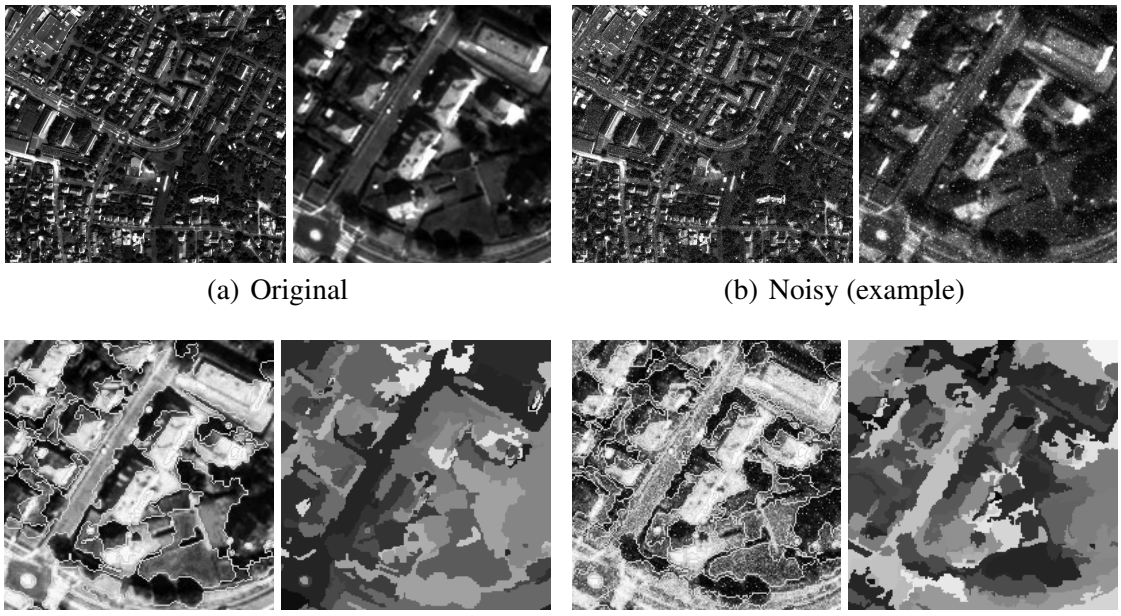
4.1 Urban Noisy Images

Data. The dataset used here was sensed over the town of Strasbourg (France). The original sample (Figure 2(a)) is an urban image (1024 × 1024 pixels) acquired by the Pléiades satellite in 2012 (courtesy LIVE, UMR CNRS 7263). It is a pansharpened image at a spatial resolution of 60 cm with four bands (NIR, R, G, B). From this image, a series of 7 noisy images was generated by adding Gaussian and speckle noise (Figure 2(b)).

Method and Results. The BPTs are built from the trivial partition $\mathcal{L}$ composed by all singleton sets, *i.e.*, one pixel per region. The valuation function $W_\star : (2^\Omega)^2 \times V^\Omega \to \mathbb{R}$ is defined as the increase of the ranges of the intensity values (for each radiometric band), potentially induced by the fusion of the incident regions. In the case of multi-images, the relative local information consensus policy *most-frequent*, weighted according to the position of the edges within the lists is applied for the first 10% of the lists $\mathcal{W}_\star$.

The non-noisy image of Figure 2(a) is first segmented by extracting a user-defined horizontal cut from its "standard" BPT (Figure 2(c)). According to a same number of regions, a cut is then selected from the multi-images BPT leading to a comparable segmentation result (Figure 2(d)).

From these figures, we observe that the results obtained from noisy images are slightly degraded, but of comparable quality, with respect to Figure 2(c). This tends to confirm



(a) Original (b) Noisy (example)

(c) Segmentation of (a) (d) Segmentation of (b) + other noisy images

Fig. 2. (a) Initial image, PLÉIADES, 2012; and a zoomed sample, 200×200 pixels. (b) An example of noisy image generated from (a), Gaussian ($\sigma = 10\%$) and speckle noise (5%), and a zoomed sample, 200×200 pixels. (c) Segmentation from the mono-image BPT of (a). (d) Segmentation from the multi-image BPT of (b) plus 6 other noisy images.

the ability of the multi-image BPT-based segmentation to generate accurate results, by discriminating relevant information from noise thanks to the consensus operated between the various images, even in the case of low signal-to-noise ratio.

4.2 Agricultural Image Time Series

Data. The dataset used here is a time series of agricultural images (1000×1000 pixels) of an area located near Toulouse (France). Images were acquired by the FORMOSAT-2 satellite over the 2007 cultural year, see Figure 3(a–c). They were ortho-rectified and have a spatial resolution of 8 m, with four spectral bands (NIR, R, G, B).

Method and Results. The BPTs are built from the same partition $\mathcal{L}$ and valuation function $W_\star : (2^\Omega)^2 \times V^\Omega \to \mathbb{R}$ as in Section 4.1. They are also segmented in the same way.

The segmented results, depicted in Figure 3(e), provide regions that are not the same as those obtained from a standard mono-image BPT, computed from one of the images of the series (Figure 3(d)). On the one hand, some segmentation effects deriving from semantic noise in mono-image segmentation are sometimes corrected by the redundant information obtained from multi-images. On the other hand, the multi-image BPT focuses on time-specific details that are only accessible via a temporal analysis, providing a potentially useful tool for such data.

(a) May (b) July (c) August (d) Seg. mono-im. (e) Seg. multi-im.

Fig. 3. (a–c) 3 samples (200 × 200 pixels) of the image time series, FORMOSAT-2, 2007. (d) Segmentation obtained from the mono-image BPT from (a). (e) Segmentation obtained from the multi-image BPT from (a–c).

5 Conclusion

This article has presented a data-structure / algorithmic framework and different fusion consensus strategies for building a unique BPT from several images. This contribution is, to the best of our knowledge, the first attempt to handle segmentation fusion in the framework of (morphological) hierarchies, in the "n images, one algorithm" paradigm.

The experiments carried out on satellite multi-image datasets have shown that the quality of the induced morphological hierarchies are sufficient to further perform improved segmentation, *e.g.*, from noisy or multi-temporal images of a same scene. The consensus strategies considered in this study remain, however, mostly basic (most-frequent, majority vote, *etc.*). Integrating higher-level consensus may then allow us to improve the quality of the hierarchies and the induced segmentation.

In the case of mono-date images, the fusion decisions underlying to the consensus strategies could also be guided by semantic information. Recent advances concerning the hierarchical modelling of such semantic information, in the context of classified remote sensing data [22], may facilitate such approaches.

In the case of multi-date images, we may handle the land-cover evolutions of the observed territories by considering an adequate region model and merging criterion [23]. Such spatio-temporal information could be used to follow local consensus between images leading to hyper-trees where the branches model local temporal fusion decisions.

In a next issue, the results obtained with this method will be fully assessed by quantitative comparisons (using datasets provided by different sensors) and compared to the results produced by other hierarchical and fusion-based segmentation methods.

References

1. Topchy, A., Jain, A.K., Punch, W.: Clustering ensembles: Models of consensus and weak partitions. IEEE TPAMI 27, 1866–1881 (2005)
2. Salembier, P., Wilkinson, M.H.F.: Connected operators: A review of region-based morphological image processing techniques. IEEE SPM 26, 136–157 (2009)
3. Salembier, P., Garrido, L.: Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. IEEE TIP 9, 561–576 (2000)
4. Rohlfing, T., Maurer Jr., C.R.: Shape-based averaging. IEEE TIP 16, 153–161 (2007)
5. Vidal, J., Crespo, J., Maojo, V.: A shape interpolation technique based on inclusion relationships and median sets. IVC 25, 1530–1542 (2007)
6. Franek, L., Abdala, D.D., Vega-Pons, S., Jiang, X.: Image segmentation fusion using general ensemble clustering methods. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part IV. LNCS, vol. 6495, pp. 373–384. Springer, Heidelberg (2011)
7. Mignotte, M.: Segmentation by fusion of histogram-based K-means clusters in different color spaces. IEEE TIP 17, 780–787 (2008)
8. Calderero, F., Eugenio, F., Marcello, J., Marqués, F.: Multispectral cooperative partition sequence fusion for joint classification and hierarchical segmentation. IEEE GRSL 9, 1012–1016 (2012)
9. Wang, H., Zhang, Y., Nie, R., Yang, Y., Peng, B., Li, T.: Bayesian image segmentation fusion. KBS 71, 162–168 (2014)
10. Chu, C.C., Aggarwal, J.K.: The integration of image segmentation maps using region and edge information. IEEE TPAMI 15, 72–89 (1993)
11. Cho, K., Meer, P.: Image segmentation from consensus information. CVIU 68, 72–89 (1997)
12. Angulo, J., Jeulin, D.: Stochastic watershed segmentation. In: ISMM, pp. 265–276 (2007)
13. Bernard, K., Tarabalka, Y., Angulo, J., Chanussot, J., Benediktsson, J.A.: Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach. IEEE TIP 21, 2008–2021 (2012)
14. Wattuya, P., Rothaus, K., Praßni, J.S., Jiang, X.: A random walker based approach to combining multiple segmentations. In: ICPR, pp. 1–4 (2008)
15. Salembier, P., Oliveras, A., Garrido, L.: Antiextensive connected operators for image and sequence processing. IEEE TIP 7, 555–570 (1998)
16. Monasse, P., Guichard, F.: Scale-space from a level lines tree. JVCIR 11, 224–236 (2000)
17. Soille, P.: Constrained connectivity for hierarchical image decomposition and simplification. IEEE TPAMI 30, 1132–1145 (2008)
18. Vilaplana, V., Marques, F., Salembier, P.: Binary partition trees for object detection. IEEE TIP 17, 2201–2216 (2008)
19. Benediktsson, J.A., Bruzzone, L., Chanussot, J., Dalla Mura, M., Salembier, P., Valero, S.: Hierarchical analysis of remote sensing data: Morphological attribute profiles and binary partition trees. In: Soille, P., Pesaresi, M., Ouzounis, G.K. (eds.) ISMM 2011. LNCS, vol. 6671, pp. 306–319. Springer, Heidelberg (2011)
20. Kurtz, C., Passat, N., Gançarski, P., Puissant, A.: Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology. PR 45, 685–706 (2012)
21. Akcay, H.G., Aksoy, S.: Automatic detection of geospatial objects using multiple hierarchical segmentations. IEEE TGRS 46, 2097–2111 (2008)
22. Kurtz, C., Naegel, B., Passat, N.: Connected filtering based on multivalued component-trees. IEEE TIP 23, 5152–5164 (2014)
23. Alonso-González, A., López-Martínez, C., Salembier, P.: PolSAR time series processing with binary partition trees. IEEE TGRS 52, 3553–3567 (2014)