Segmentation of Multimodal Images based on Hierarchies of Partitions

Guillaume Tochon, Mauro Dalla Mura, Jocelyn Chanussot

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, 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 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 $\mathcal{I} : E \to V$, $E \subseteq \mathbb{Z}^2$, $V \subseteq \mathbb{R}^n$, 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:

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

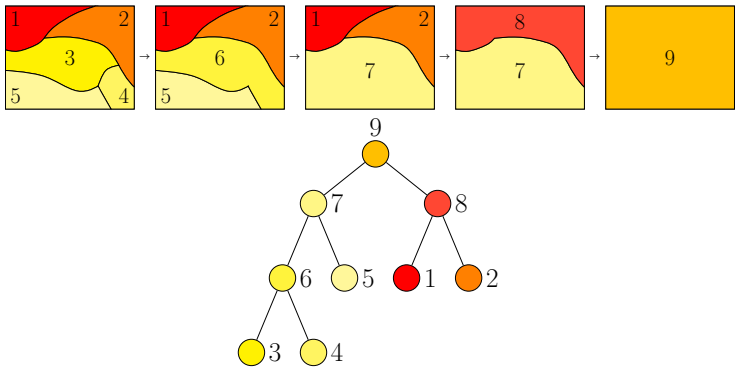


Fig. 1: Example of hierarchy of partitions (binary partition tree [14]).

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

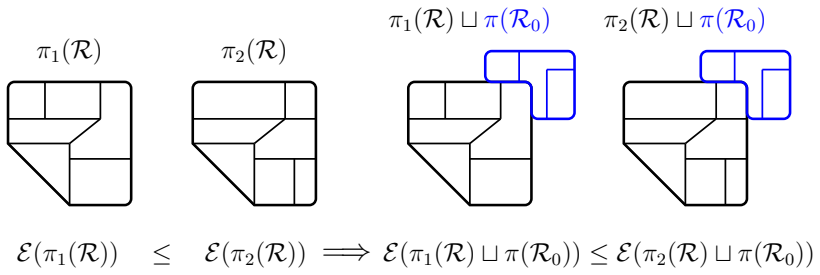


Fig. 2: Example of a h-increasing energy.

h-increasing energies. An energy $\mathcal{E}$ is said to be h-increasing when given any two $\mathcal{R}, \mathcal{R}_0 \in H$ 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:

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

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

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

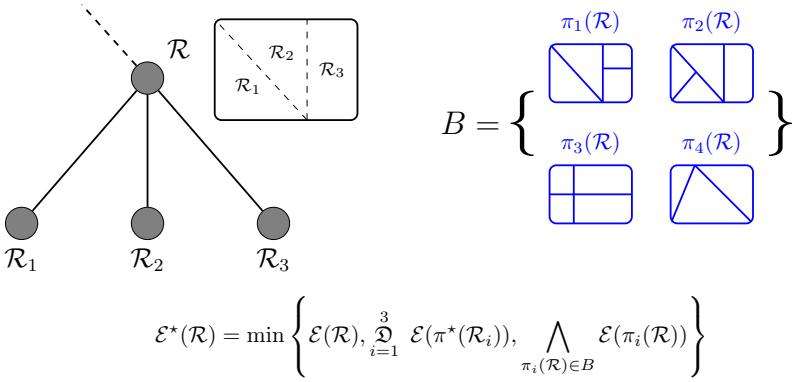


Fig. 3: Illustration of a step of the dynamic program (7) applied to a braid structure: one has to choose between $\{\mathcal{R}\}$, $\bigsqcup \pi^\star(\mathcal{R}_i)$ or any other $\pi_i(\mathcal{R}) \in B$. Note however that $\mathcal{R} \neq E$, otherwise B would not be a braid since $\pi_3(\mathcal{R})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

practice, the energy $\mathcal{E}_\lambda$ is seen as a function of λ, and (3) is conducted over 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:

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

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

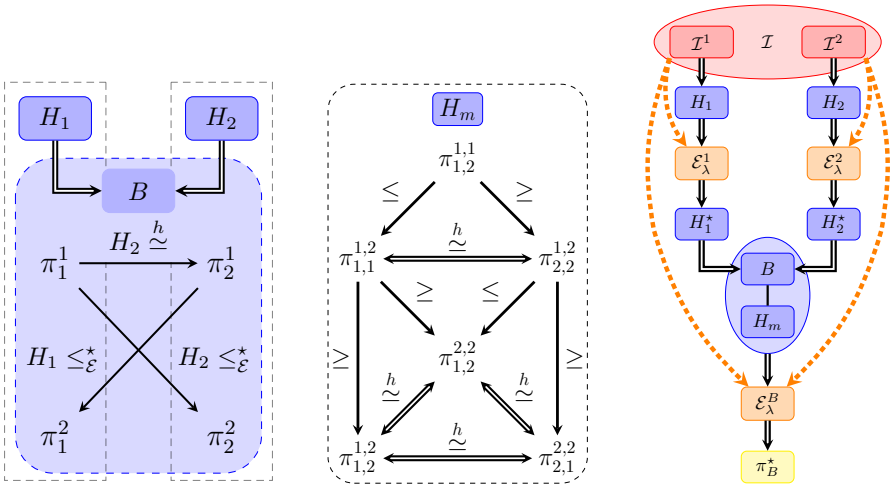


Fig. 4: a) Possible composition of a braid B with cuts from two hierarchies H_1 and H_2, b) cuts of the corresponding monitor hierarchy H_m, and c) workflow of proposed multimodal segmentation methodology.

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 $H \overset{h}{\simeq} \pi_*$ 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 $\mathcal{I}$:

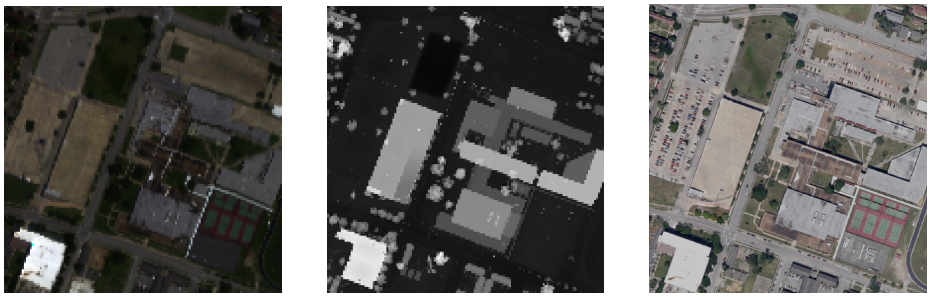
$$\mathcal{E}_\lambda^B(\pi) = \sum_{\mathcal{R} \in \pi} \left(\max\left(\frac{\Xi_1(\mathcal{R})}{\Xi_1(\mathcal{I}^1)}, \frac{\Xi_2(\mathcal{R})}{\Xi_2(\mathcal{I}^2)} \right) + \frac{\lambda}{2} |\partial \mathcal{R}| \right) \tag{9}$$

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)



(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.

$\mathcal{I}^2$, 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 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[3].
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

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

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

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:

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

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 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.

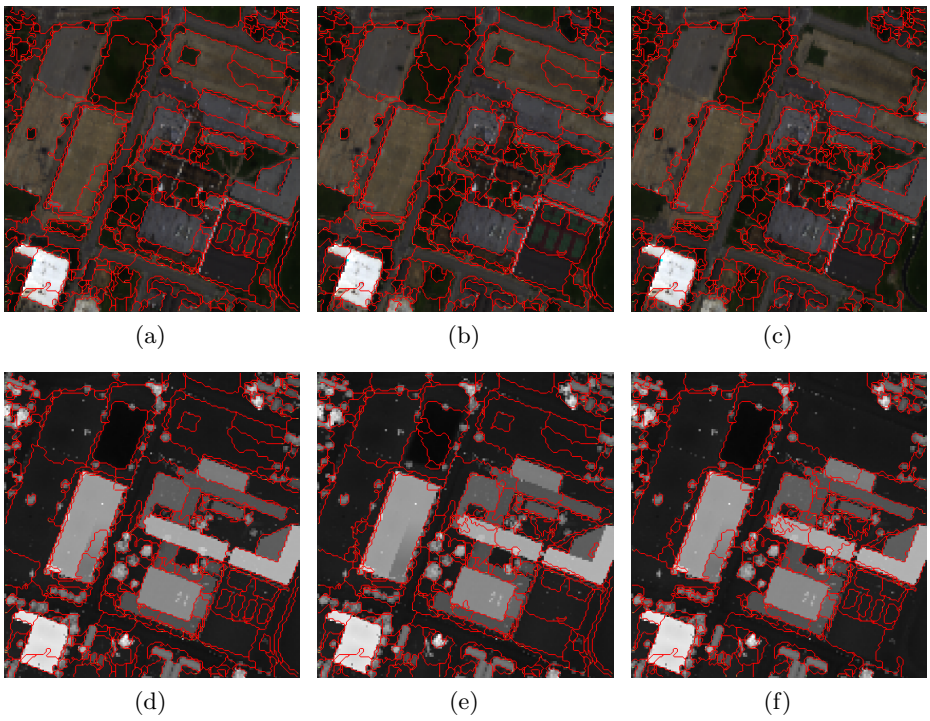


Fig. 6: Top row: optimal partitions a) $\pi_B^\star$, b) $\pi_1^\star$ (optimal with respect to the HS image) and c) $\pi_2^\star$ (optimal with respect to the DSM) superimposed over the HS image. Bottom row: optimal partitions d) $\pi_B^\star$, e) $\pi_1^\star$ and f) $\pi_2^\star$ superimposed over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

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. F. Bießmann, S. Plis, F. Meinecke, T. Eichele, and K. Muller. Analysis of multimodal neuroimaging data. *Biomedical Engineering, IEEE Reviews in*, 4:26–58, 2011.
2. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.
3. M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, and J. Chanussot. Challenges and opportunities of multimodality and data fusion in remote sensing. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European*, pages 106–110. IEEE, 2014.
4. C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. L., R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici. Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 7(6):2405–2418, June 2014.
5. L. Guigues, J. Cocquerez, and H. Le Men. Scale-sets image analysis. *International Journal of Computer Vision*, 68(3):289–317, 2006.
6. B. Kiran. *Energetic-Lattice based optimization.* PhD thesis, Universit Paris-Est, Paris, Oct. 2014.
7. B. Kiran and J. Serra. Global–local optimizations by hierarchical cuts and climbing energies. *Pattern Recognition*, 47(1):12–24, 2014.
8. D. Lahat, T. Adaly, and C. Jutten. Challenges in multimodal data fusion. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European*, pages 101–105. IEEE, 2014.
9. S. Li. *Markov random field modeling in computer vision.* Springer-Verlag New York, Inc., 1995.
10. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.
11. C. Ronse. Partial partitions, partial connections and connective segmentation. *Journal of Mathematical Imaging and Vision*, 32(2):97–125, 2008.
12. P. Soille. Constrained connectivity for hierarchical image partitioning and simplification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(7):1132–1145, 2008.
13. Y. Tarabalka, J. Tilton, J. Benediktsson, and J. Chanussot. A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(1):262–272, 2012.
14. S. Valero, P. Salembier, and J. Chanussot. Hyperspectral image representation and processing with binary partition trees. *Image Processing, IEEE Transactions on*, 22(4):1430–1443, 2013.