Pattern Recognition
journal homepage: www.elsevier.com

Braids of partitions for the hierarchical representation and segmentation of multimodal images

Guillaume Tochon[a], Mauro Dalla Mura[b], Miguel Angel Veganzones[b], Thierry Géraud[a], Jocelyn Chanussot[b,c]

[a] *EPITA Research and Development Laboratory (LRDE), 14-16, rue Voltaire, FR-94276 Le Kremlin-Bicêtre, France*
[b] *GIPSA-Lab, 11 rue des Mathématiques, 38400 Saint Martin d'Hères, France.*
[c] *Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavík, Iceland.*

ABSTRACT

Hierarchical data representations are powerful tools to analyze images and have found numerous applications in image processing. When it comes to multimodal images however, the fusion of multiple hierarchies remains an open question. Recently, the concept of braids of partitions has been proposed as a theoretical tool and possible solution to this issue. In this paper, we demonstrate the relevance of the braid structure for the hierarchical representation of multimodal images. We first propose a fully operable procedure to build a braid of partitions from two hierarchical representations. We then derive a framework for multimodal image segmentation, relying on an energetic minimization scheme conducted on the braid structure. The proposed approach is investigated on different multimodal images scenarios, and the obtained results confirm its ability to efficiently handle the multimodal information to produce more accurate segmentation outputs.

1. Introduction

The notion of scale of analysis is a key paradigm in image processing. A given image can be analyzed at different scales, *i.e.*, different levels of details, depending on the pursued goal. For low-level applications such as image denoising, algorithms mostly work at very fine scales, where objects of interest are either defined as pixels or small groups of pixels. On the other hand, high-level image understanding and simplification applications such as object recognition or image segmentation focus on coarser scales of analysis, where the handled objects of interest are defined as large groups of pixels conveying some semantic meaning. Hierarchical representations are a well suited tool to handle this multi-scale nature of images, since they allow to encompass all potential scales of interest in a single structure. The hierarchical representation can be constructued once and regardless of the application, and its scale of analysis can then be tuned afterward to comply with the pursued task. The component tree (also called min-tree and max-tree) [1], the inclusion tree (also called tree of shapes (ToS)) [2], the α-tree (also called the hierarchy of quasi-flat zones) [3] and the binary partition tree (BPT) [4] constitute a non-exhaustive list of the most known hierarchical representations in the mathematical morphology literature. Reviews can be found in [5; 6]. Hierarchical representations have proven to be useful for many image processing and computer vison tasks of various scales of analysis, such as image filtering [7; 8] and simplification [9], image segmentation [10; 11] as well as object detection [12] and recognition [13].

The most common framework for a given input image is to build and process a single hierarchical representation. In some cases however, it could be interesting to associate one image with multiple hierarchical representations (each one focusing on a particular feature of the image for example), or, on the contrary, to build a common hierarchical representation for multiple input images. While these largely remain open questions, some recent works have been devoted to such fusion issues. The fusion of multiple hierarchical representation constructed on a single image is for instance addressed in [14; 15], where hierarchies of watersheds (see [16]) driven by area and dynamics attributes are combined through the composition by infimum, supremum or averaging of their saliency maps. The representation of multimodal images (*i.e.* several images acquired over the same scene with different setups, such as different sensor types or acquisition dates) [17] within a single hierarchical structure is another challenge studied in the literature [18]. By ensuring a more

e-mail: guillaume.tochon@lrde.epita.fr (Guillaume Tochon)

complete and accurate representation of the recorded source, as they consider several single acquisitions of it, multimodal images are nowadays increasingly used in image processing. However, jointly integrating the redundant and complementary information featured by the various modalities in a hierarchical representation in generic manner is an arduous task, as it depends both on the nature of the handled multimodality as well as the underlying application.

Some works have been devoted to this challenging issue. The ToS structure has for instance been extended to multivariate images in [19], where univariate ToS are first computed for each individual modality and then further merged into a graph from which is derived the final multivariate ToS representation (note that a similar idea is presented in [20] to extend component trees to multivariate images, but the final result is a graph and no longer a tree-based representation). In [21], a single BPT is built over a whole video sequence by integrating motion cues during the construction stage, allowing to perform some object tracking by simply identifying nodes of interest in the resulting trajectory BPT structure. Another approach for the construction of a multi-feature BPT has been introduced in [22; 23], where all modalities of the input multimodal image cooperate in a consensus framework to allow for the construction of a single tree structure. Finally, braids of partitions were proposed in [24] as a generalization of hierarchies of partitions for a theoretical standpoint, and we actually sketched in a previous work [25] the potentiality of such braid structures to act as suited hierarchical representations of multimodal images. In [25], we proposed to build the braid structure (and its associated monitor hierarchy) by experimentally combining cuts extracted from two hierarchies of partitions, and showed the interest of the resulting structure within the framework of multimodal image segmentation. Here, we extend those preliminary results by defining a complete methodology for the hierarchical representation and segmentation of multimodal images. More precisely, we complete our previous work [25] in the following aspects:

1. We provide the formal proof that the proposed methodology to build the braid structure from two hierarchical representation mathematically satisfies the definition of the braid structure.
2. We demonstrate the relevance of the obtained braid structure for multimodal image representation by integrating it into a more general multimodal image segmentation framework. We benchmark this latter framework on two different multimodal datasets[1].
3. We conduct a sensibility analysis to the two key parameters that have to be tuned in order to operate the proposed multimodal image segmentation framework.

Note that the segmentation application should be taken as a proof of concept to demonstrate the soundness of the proposed braid framework and its adaptability to different multimodal scenarios with their respective specificities, and not as an attempt to outperform state-of-the-art multimodal segmentation techniques specialized in the segmentation of a particular multimodality.

[1] One is presented in the supplementary materials available from page 13.

The remainder of this paper is organized as follows: Section 2 introduces various definitions and properties related to hierarchical representations and hierarchical energy minimization procedures. Section 3 presents the concept of braids of partitions proposed by [24] and extends the classical energetic framework on these particular structures. Section 4 details the main contributions of this paper as stated above, while Section 5 shows the application of this methodology and discusses the obtained results. Conclusion and future work are drawn in Section 6.

2. Hierarchies of partitions

2.1. Hierarchies of partitions

Let $\mathcal{I} : E \to V$ be a generic image of elements (pixels) $x_i \in E$ belonging to the *support* space E of the image, *i.e.*, its pixel grid (in which case $E \subseteq \mathbb{Z}^2$ although there is no requirement for E to be discrete in the following), and of pixel values $\mathcal{I}(x_i) \in V \subseteq \mathbb{R}^n$. Following this definition, a $P-$multimodal image $\mathcal{I}_P$ is characterized by the joint composition of its P modalities $\{\mathcal{I}_1, \dots, \mathcal{I}_P\}$, with $\mathcal{I}_i : E_i \to V_i$, $i = 1, \dots P$. Although each domain E_i could be different for the various modalities, we restrict here to the case where all the modalities share the same domain $E_1 = \cdots = E_P \equiv E$, implying that all modalities are co-registered. On the other hand, all sets V_i are not restricted to be the same, and can be of different dimensionality.

A *region* $\mathcal{R} \subseteq E$ is some (non necessarily connected) subset of E. A *partition* π of E is a collection of regions $\{\mathcal{R}_i \subseteq E\}$ 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. The words segmentation and partition are used interchangeably in the following.

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 (partial) ordering $\leq$. In particular, it is possible to define the refinement supremum $\pi_i \vee \pi_j$ of two partitions π_i and π_j as the lowest partition of Π_E that is refined by both π_i and π_j, and the refinement infimum $\pi_i \wedge \pi_j$ as the greatest partition that refines both π_i and π_j.

A hierarchy of partitions H of E is a collection of partitions $\{\pi_i \in \Pi_E\}_{i=0}^n$ ordered by refinement, that is $H = \{\pi_0 \leq \pi_1 \leq \cdots \leq \pi_n\}$. π_0 is called the *leaf* partition (its regions are the *leaves* of H) and $\pi_n = \{E\}$ is the *root* of the hierarchy. A hierarchy of partitions is often represented as a tree graph (also called dendogram), where the nodes of the graph correspond to the various regions contained in the partitions of the sequence, and the vertices denote the inclusion between these regions. Alternatively, H can be equivalently defined as a collection of regions $H = \{\mathcal{R} \subseteq E\}$ that satisfy the following 3 properties:

1. $\emptyset \notin H$, $E \in H$.
2. $\forall \mathcal{R}_i, \mathcal{R}_j \in H$, $\mathcal{R}_i \cap \mathcal{R}_j \in \{\emptyset, \mathcal{R}_i, \mathcal{R}_j\}$. Any two regions belonging to H are either disjoint or nested.
3. $\forall \mathcal{R} \in H, \mathcal{R} \notin \pi_0 \Rightarrow \mathcal{R} = \bigcup_{r\in\pi_0}\{r | r \subset \mathcal{R}\}$. Any non leaf region $\mathcal{R}$ is exactly recovered by the union of all leaves of H that are included in $\mathcal{R}$.

Considering only items 1 and 2 allows to define tree-based representations such as the ToS, but item 3 is mandatory to define hierarchies or partitions.

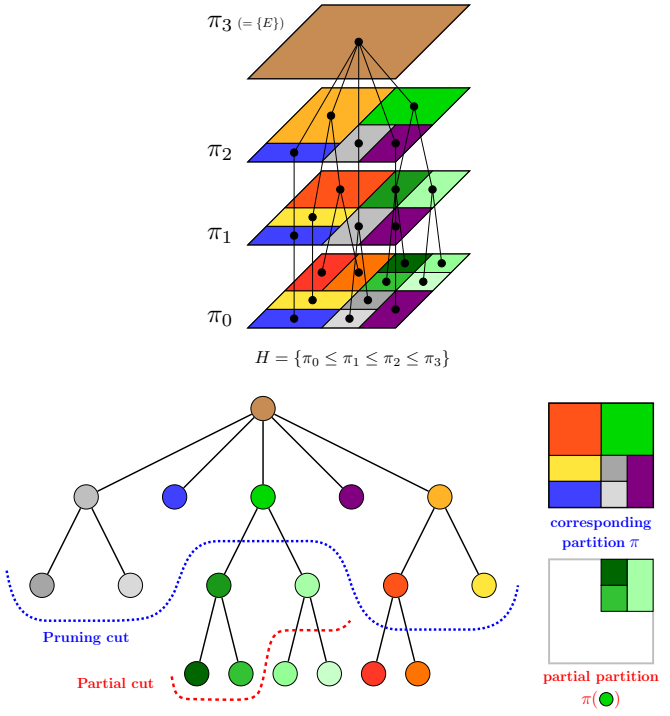

Fig. 1: Summary of presented notions related to hierarchies of partitions.

A (pruning) *cut* of H is a partition π of E whose regions belong to H, and $\Pi_E(H)$ denotes the set of all such cuts. $H(\mathcal{R})$ stands for 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 [26], and is denoted $\pi(\mathcal{R})$. Figure 1 illustrates those notions related to hierarchies of partitions.

2.2. Hierarchical energy minimization

Many image processing algorithms can be formulated in a framework where some objective (also called *energy*) function is minimized, and the resulting minimizer defines the sought result. This is for instance the case for segmentation purposes, where many algorithms seeks the best partitioning of the image with respect to some criterion (for instance, region homogeneity) and under some potential constraints (an upper bound on the total number of regions composing the partition for example). Well-known segmentation algorithms formulated as energy minimization processes include the Mumford-Shah functional [27], graph cuts [28] or Markov random fields [29].
In the following, an energy function will be defined as a mapping $\mathcal{E} : \Pi_E \to \mathbb{R}^+$ that associates to each partition $\pi \in \Pi_E$ a real positive number $\mathcal{E}(\pi)$. More specifically, the energy of a partition π can be expressed as some particular composition of the energies of the regions composing the partition:

$$\mathcal{E}(\pi) = \bigotimes_{\mathcal{R}_i \in \pi} \mathcal{E}(\mathcal{R}_i), \tag{1}$$

where $\otimes$ is a composition rule to explicit the relationship between the energy of the partition π and those of its regions $\mathcal{R}_i \in \pi$. While $\otimes$ can be arbitrarily defined, the sum composition (*i.e.* $\mathcal{E}(\pi) = \sum_{\mathcal{R}_i \in \pi} \mathcal{E}(\mathcal{R}_i)$) is the option that is classically adopted in practice. However, the minimization of such energy functions over the whole set of partitions Π_E is particularly complicated due to the huge cardinality of Π_E. Hierarchies of partitions, by restraining the space of possible partitions, are an appealing tool to minimize the energy on.
Given some hierarchy of partitions H and some energy $\mathcal{E}$, the cut of H that is minimal (*i.e.*, *optimal*) with respect to the energy $\mathcal{E}$ is defined as:

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

Note that this definition does not guarantee the uniqueness of the optimal cut, as several cuts could equally minimize the energy. For the sake of readability, *the* optimal cut will imply *the largest* of all optimal cuts in the following. The combinatorics of trees makes the exhaustive search of the optimal cut $\pi^\star$ non tractable in practice. To overcome this issue, conditions that have to be satisfied by $\mathcal{E}$ to ease the retrieval of the optimal cut were formally investigated for the first time in [30] in the context of separable energies (*i.e.*, $\otimes \equiv \sum$) and later on generalized in [31; 32] to wider classes of composition rules $\otimes$, namely *h-increasing energies*. In that case, the optimal cut of H can be found by solving for each node $\mathcal{R}$ the following dynamic program:

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

$$\pi^\star(\mathcal{R}) = \operatorname{argmin}\left\{\mathcal{E}(\mathcal{R}), \mathcal{E}\Big(\bigsqcup_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r)\Big)\right\} \tag{4}$$

with $\sqcup$ denoting disjoint union (concatenation) and $\mathsf{S}(\mathcal{R})$ being the set of children nodes of $\mathcal{R}$. The optimal cut of $\mathcal{R}$ is given by comparing the proper energy of $\mathcal{R}$ and the energy of the disjoint union of the optimal partial cuts of its children, and by picking the smallest of the two. The optimal cut of the whole hierarchy is the one of the root node, and is reached by scanning all nodes in the hierarchy in one ascending pass [30]. It was shown in [24] that all energies which can be expressed as a Minkowski expression:

$$\mathcal{E}(\pi) = \left(\sum_{\mathcal{R} \in \pi} \mathcal{E}(\mathcal{R})^\alpha\right)^{\frac{1}{\alpha}} \tag{5}$$

are h-increasing for every $\alpha \in [-\infty, +\infty]$, generalizing previously obtained results for energies composed by the sum ($\alpha = 1$) [30; 4], the supremum ($\alpha = +\infty$) [33] and the infimum ($\alpha = -\infty$) [34], notably. Thus, the optimal cut of a hierarchy for any type of Minkowski-composed energy function can be easily retrieved following equations (3) and (4).
Energies in the literature often depend in practice on a positive real-valued parameter λ that acts as a trade-off between simplicity and a good data fitting of the segmentation. In that context, there is no longer one optimal cut $\pi^\star$ for a given hierarchy H and some energy $\mathcal{E}_\lambda$ parametrized by λ, but rather a family of them $\{\pi^\star_\lambda\}$ in turn indexed by this parameter λ. It was shown in particular in [32] that under the assumption of *scale-increasingness* for $\mathcal{E}_\lambda$, the family $\{\pi^\star_\lambda\}$ of optimal cuts can be ordered by refinement, that is

$$\lambda_1 \leq \lambda_2 \Rightarrow \pi^\star_{\lambda_1} \leq \pi^\star_{\lambda_2}. \tag{6}$$

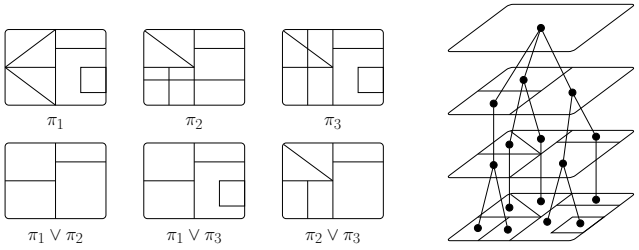

Fig. 2: Example of braid of partitions $B = \{\pi_1, \pi_2, \pi_3\}$. On the right is a monitor hierarchy of B since the pairwise refinement suprema $\pi_i \vee \pi_j, i, j \in \{1, 2, 3\}, i \neq j$ define cuts of this hierarchy different from the whole space E.

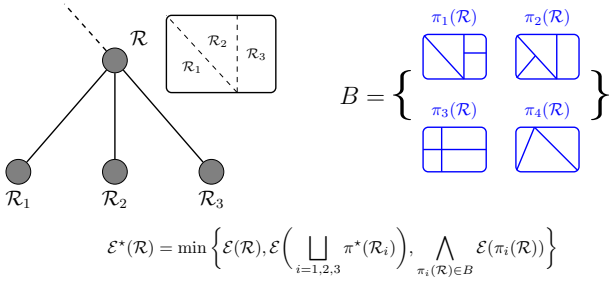

Fig. 3: A step of the dynamic program (8) applied to a braid structure: one has to choose between $\{\mathcal{R}\}$, $\bigsqcup \pi^\star(\mathcal{R}_i)$ or any other $\pi_i(\mathcal{R})$ for $\pi_i \in B$. Note however that $\mathcal{R} \neq E$, otherwise B would not be a braid since $\pi_3(\mathcal{R}) \vee \pi_4(\mathcal{R}) = \mathcal{R}$.

This property notably allows to transform some hierarchy H into its *persistent* version $H^\star$, composed of all the optimal cuts $\pi^\star_\lambda$ of H when λ spans $\mathbb{R}^+$. The reader is referred to [30] for more practical implementation details.

3. Braids of partitions

3.1. Definition of a braid

The analysis of a multimodal image by means of a hierarchical representation inevitably raises the question of the optimal exploitation of both the redundant and complementary information contained in the various modalities. Braids of partitions have been recently introduced in [24] as a potential tool to combine multiple hierarchies and thus precisely answer this question [25]. Braids of partitions are defined as follows:

Definition 1 (Braid of partitions)**.** *A family of partitions $B = \{\pi_i \in \Pi_E\}$ is called a braid of partitions whenever there exists some hierarchy H_m, called monitor hierarchy, such that:*

$$\forall\, \pi_i, \pi_j \in B, \pi_i \vee \pi_{j \neq i} \in \Pi_E(H_m) \backslash \{E\} \tag{7}$$

Braids of partitions generalize hierarchies of partitions in the sense that the refinement ordering between the partitions composing the braid no longer needs to exist, as long as all their pairwise refinement suprema are hierarchically organized. It is also worth noting that those refinement suprema must differ from the whole image $\{E\}$ in (7). Otherwise, any family of arbitrary partitions would form a braid with $\{E\}$ as a supremum, thus loosing any interesting structure. An example of braid is displayed by Figure 2.

The structure of a braid of partitions B, along with its monitor hierarchy H_m, appears well suited for the hierarchical representation of multimodal images. As it can be observed in Figure 2, the monitor hierarchy H_m encodes all regions that are common to at least two different partitions contained in B. Assuming that these partitions originate from different modalities, the monitor hierarchy therefore expresses regions that are salient across the modalities, at various scales. In other word, the monitor hierarchy can be seen as a representation of the redundant information contained in the multimodal image. On the other hand, the family B exhibits the complementary information: all regions contained in B but not in H_m belong to a single modality, and can thus be considered as complementary information. Jointly considering the braid B and its monitor hierarchy H_m therefore leads to hierarchical representation of the multimodal image that relies both on the complementary and redundant information contained in the data.

3.2. Minimizing an energy function over a braid

While any two regions belonging to a braid of partitions may no longer be either disjoint or nested, as it is the case for hierarchies of partitions, it was shown in [24] that the dynamic program structure holding on hierarchies (equations (3) and (4)) remains valid, with however a slight modification. In particular, the optimal cut of a braid is reached by solving the following dynamic program for every node $\mathcal{R}$ of the monitor hierarchy H_m:

$$\mathcal{E}^\star(\mathcal{R}) = \min\left\{\mathcal{E}(\mathcal{R}), \mathcal{E}\left(\bigsqcup_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r)\right), \bigwedge_{\pi_i \in B} \mathcal{E}(\pi_i(\mathcal{R}))\right\} \tag{8}$$

$$\pi^\star(\mathcal{R}) = \begin{cases} \{\mathcal{R}\} & \text{if } \mathcal{E}^\star(\mathcal{R}) = \mathcal{E}(\mathcal{R}) \\ \displaystyle\bigsqcup_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r) & \text{if } \mathcal{E}^\star(\mathcal{R}) = \mathcal{E}\left(\bigsqcup_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r)\right) \\ \displaystyle\operatorname*{argmin}_{\pi_i \in B} \mathcal{E}(\pi_i(\mathcal{R})) & \text{otherwise.} \end{cases} \tag{9}$$

Compared to the classical procedure over hierachies, one has also to consider all the others partial partitions of $\mathcal{R} \in H_m$ 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. Note that, although the dynamic program is conducted over the monitor hierarchy H_m, the optimal cut of the braid B may be composed of regions that do not belong to H_m (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 $\mathcal{R}$).

4. Proposed hierarchical analysis of multimodal images with braids

4.1. Constructing a braid from multiple hierarchies

As pointed out in [24], the two issues that arise when working with braids of partitions are the validation of the braid structure for a given family of partitions (that is, condition (7) is fulfilled),

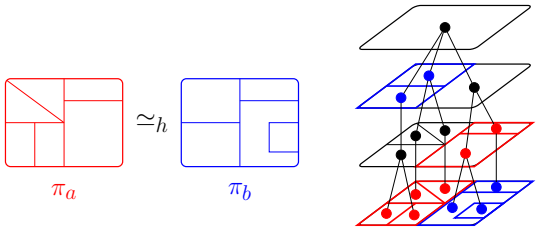

Fig. 4: Illustration of the h-equivalence relation: π_a and π_b are h-equivalent (left), they define two different cuts of a same hierarchy (right).

and the definition of general procedures that generate braids of partitions.

It is straightforward to compose a braid using a single hierarchy since the supremum of two cuts of a hierarchy also defines a cut of this hierarchy. For this reason, any set of cuts coming from a single hierarchy is a braid. However, this guarantee is lost when one wants to compose a braid from cuts coming from multiple hierarchies, or, even further, with arbitrary partitions (note in that respect that, although tempting to think so, the family of partitions generated by the stochastic watershed [35] has not a braid structure in general). As a matter of fact, all those partitions must be sufficiently related so their pairwise refinement suprema define cuts of the same monitor hierarchy H_m. To analyze the relationships which must be holding between the cuts of various hierarchies to form a braid, we introduce the property of *h-equivalence* (h standing here for *hierarchical*):

Definition 2 (h-equivalence)**.** *Two partitions π_a and π_b are said to be h-equivalent, and one notes $\pi_a \simeq_h \pi_b$ if and only if*

$$\forall \mathcal{R}_a \in \pi_a,\ \forall \mathcal{R}_b \in \pi_b, \mathcal{R}_a \cap \mathcal{R}_b \in \{\emptyset, \mathcal{R}_a, \mathcal{R}_b\}. \tag{10}$$

In other words, a region in π_a either refines or is a refinement of a region in π_b. Partitions π_a and π_b may not be globally comparable but they locally are, as displayed by Figure 4. Obviously, if $\pi_a \leq \pi_b$ or $\pi_b \leq \pi_a$, then $\pi_a \simeq_h \pi_b$. All cuts of a hierarchy H are h-equivalent: $\forall \pi_1, \pi_2 \in \Pi_E(H)$, $\pi_1 \simeq_h \pi_2$. Conversely, if two partitions are h-equivalent, they define two cuts of a same hierarchy. Despite a somewhat misleading name, $\simeq_h$ is not an equivalence relation but only a tolerance relation: it is reflexive and symmetric but not transitive in general.

Following, we aim to build a braid B using cuts extracted from several hierarchical representations. To do so, we must investigate what kind of relationship must be holding between those cuts in order to guarantee the braid structure (that is, equation (7) is satisfied. Let the family of partitions $B = \{\pi_i \in \Pi_E\}$ be a braid, and let H_m be a monitor hierarchy of B.

Proposition 1. *If there exist $\pi_i, \pi_j \in B$ such that $\pi_i \leq \pi_j$, then $\pi_j \in \Pi_E(H_m)$.*

Proof. As $\pi_i \leq \pi_j$, it follows that $\pi_i \vee \pi_j = \pi_j$. And from the definition (7) of a braid, $\pi_i \vee \pi_j \in \Pi_E(H_m)$, so $\pi_j \in \Pi_E(H_m)$. □

Thus, if the braid B has two partitions ordered by refinement (two cuts extracted from the same hierarchy for instance), the coarest of them also belongs to the set of cuts $\Pi_E(H_m)$ of the monitor hierarchy H_m.

Proposition 2. *If there exist $\pi_i, \pi_j, \pi_k, \pi_l \in B$ such that $\pi_i \leq \pi_j$ and $\pi_k \leq \pi_l$, then $\pi_j \simeq_h \pi_l$.*

Proof. Using Proposition (1) for both $\pi_i \leq \pi_j$ and $\pi_k \leq \pi_l$, it follows that $\pi_j, \pi_l \in \Pi_E(H_m)$. Thus $\pi_j \simeq_h \pi_l$ using the property of h-equivalence. □

Therefore, if the braid B has two pairs partitions ordered by refinement, the coarest of both pairs are necessarily h-equivalent to each other since they both belong to the set of cuts $\Pi_E(H_m)$ of the monitor hierarchy H_m.

Given some hierarchy H and a partition $\pi_* \in \Pi_E$, we denote by $H^{\simeq_h}(\pi_*)$ the set of cuts of H that are h-equivalent to π_*: $H^{\simeq_h}(\pi_*) \subseteq \Pi_E(H)$ with equality if and only if $\pi_* \in \Pi_E(H)$. Similarly, we denote by $H^{\leq}(\pi_*)$ the set of cuts of H that are a refinement of π_*.

Now, let H_1 and H_2 be two hierarchies of partitions built over the same space E. We aim to extract two cuts $\pi_i^1, \pi_i^2 \in \Pi_E(H_i)$ from each of those two hierarchies $H_i, i \in \{1, 2\}$ in order for the family $B = \{\pi_1^1, \pi_1^2, \pi_2^1, \pi_2^2\}$ to be a braid. For this purpose, we propose the following iterative procedure:

1. First select arbitrarily some cut $\pi_1^1 \in \Pi_E(H_1)$.
2. Then choose a cut π_2^1 in the constrained set $H_2^{\simeq_h}(\pi_1^1)\backslash\{E\}$, that is, a cut from H_2 which is h-equivalent to π_1^1 and different from the whole space $\{E\}$.
3. Finally, complete by taking a cut in each hierarchy that is a refinement of the cut previously extracted from the other hierarchy, that is $\pi_i^2 \in \Pi_E(H_i), i \in \{1, 2\}$ such that $\pi_1^2 \leq \pi_2^1$ and $\pi_2^2 \leq \pi_1^1$.

This procedure is summarized by Figure A.11.

Proposition 3. *Under this configuration, $B = \{\pi_1^1, \pi_1^2, \pi_2^1, \pi_2^2\}$ has a braid structure.*

Proof. The proof is provided as a supplementary material in Section Appendix A. □

While other configurations for the composition of B may also work, it is the first time that, to the best of our knowledge, guidelines to create a non trivial braid by composing cuts from two hierarchies are explicitly provided. We are, up to now, only able to provide those guidelines and to guarantee the braid structure when at most two cuts are extracted from those two hierarchies.

4.2. Braid-based multimodal image segmentation

From a conceptual point of view, conducting the energy minimization procedure described in section 3.2 over a braid structure is appealing to perform multimodal segmentation. As a matter of fact, if the braid is composed of partitions extracted from the hierarchies constructed on the various modalities, then the monitor hierarchy can be seen as a hierarchical representation containing the salient regions that are common to the various modalities, at all scales. Then, during the energy minimization procedure, the dynamic program has to decide whether a common salient region $\mathcal{R} \in H_m$ should be retained (that is, if $\pi^\star(\mathcal{R}) = \{\mathcal{R}\}$), or replaced either by common regions at a smaller scale ($\pi^\star(\mathcal{R}) = \bigsqcup_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r)$) or by the set of regions at

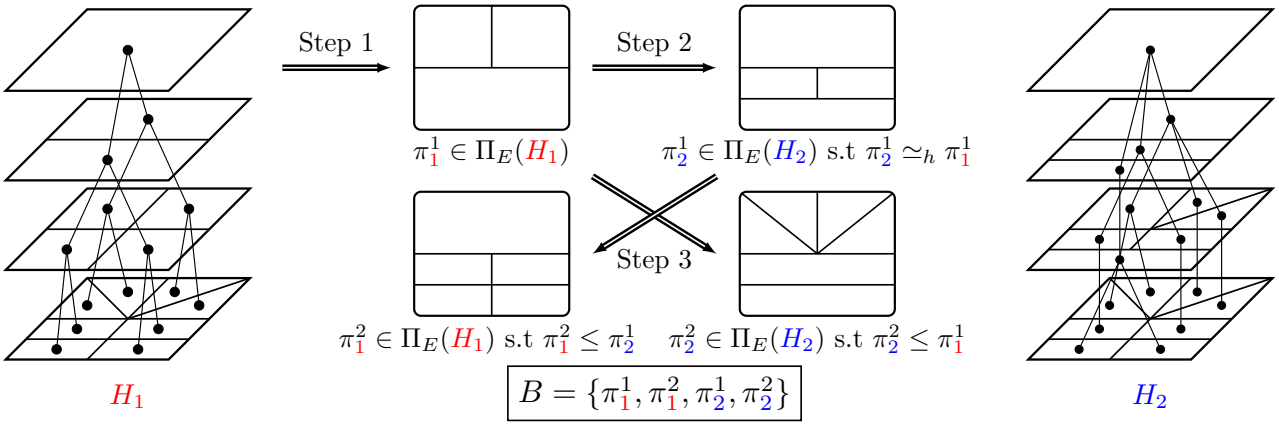

Fig. 5: Composing a braid B with cuts from two hierarchies H_1 and H_2.

a smaller scale, coming from one modality and that fit all the modalities at the same time ($\pi^\star(\mathcal{R}) = \text{argmin}_{\pi_i(\mathcal{R})\in B}\, \mathcal{E}(\pi_i(\mathcal{R}))$). Therefore, we propose a methodology to perform multimodal image segmentation based on the concept of braids of partition, as illustrated by the workflow in Figure 6.

Let $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2\}$ be a multimodal image, assumed to be composed of two modalities $\mathcal{I}_1$ and $\mathcal{I}_2$ having the same spatial support E (hence being co-registered). First, two hierarchies H_1 and H_2 are built independently on $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively. Two energy functions $\mathcal{E}_\lambda^1$ and $\mathcal{E}_\lambda^2$ are defined on their respective hierarchies, with only constraints to be h-increasing and scale-increasing in order to transform the hierarchies H_1 and H_2 into their persistent versions $H_1^\star$ and $H_2^\star$. For segmentation purposes, we propose to define the energy functions as a piece-wise constant Mumford-Shah energy [27]:

$$\mathcal{E}_\lambda^i(\pi) = \sum_{R\in\pi} \left(\Xi_i(\mathcal{R}) + \frac{\lambda}{2}|\partial\mathcal{R}|\right) \tag{11}$$

where

$$\Xi_i(\mathcal{R}) = \int_{\mathcal{R}} \|\mathcal{I}_i(x) - \boldsymbol{\mu}_i(\mathcal{R})\|_2^2 dx \tag{12}$$

with $\boldsymbol{\mu}_i(\mathcal{R})$ being the mean value/vector in modality $\mathcal{I}_i$ of pixel values belonging to region $\mathcal{R}$, and $|\partial\mathcal{R}|$ denotes the length of the boundary of $\mathcal{R}$. The first term $\Xi_i(\mathcal{R})$ is classically termed the goodness-of-fit (GOF) term and penalizes inhomogeneous regions, thus leading to fine partitions and favoring over-segmentation. The second term $|\partial\mathcal{R}|/2$ is often called the regularization term and promotes partitions with few region boundaries, therefore favoring under-segmentation. The λ coefficient achieves a trade-off to balance the effects of the GOF and regularization terms. The piece-wise constant Mumford-Shah energy function, in addition to being h-increasing and scale-increasing [32], is a popular choice when it comes to minimizing some energy function because of its ability to produce consistent segmentations [36].

The braid B is then composed following the procedure previously described in section 4.1, allowing to construct the monitor hierarchy H_m. A last energy term $\mathcal{E}_\lambda^B$ is defined as a multimodal piece-wise constant Mumford-Shah energy, relying on both modalities of the multimodal image $\mathcal{I}$:

$$\mathcal{E}_\lambda^B(\pi) = \sum_{\mathcal{R}\in\pi} \left(\max\left(\frac{\Xi_1(\mathcal{R})}{\Xi_1(\mathcal{I}_1)}, \frac{\Xi_2(\mathcal{R})}{\Xi_2(\mathcal{I}_2)}\right) + \frac{\lambda}{2}|\partial\mathcal{R}|\right) \tag{13}$$

The GOF term of each region $\mathcal{R}$ is now defined as the maximum with respect to both normalized unimodal GOFs. The maximum criterion allows to penalize a region $\mathcal{R}$ that would fit only one modality. It therefore ensures the regions of the braid optimal cut to conform both modalities at the same time. 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 thanks to the fact that the GOF term is positive, and the regularization term is the same as in the classical Mumford-Shah functional. Its minimization over H_m and B following the dynamic program (8) and (9) gives some optimal segmentation $\pi_B^\star$ of $\mathcal{I}$, which should contain salient regions shared by both modalities as well as regions exclusively expressed by $\mathcal{I}_1$ and $\mathcal{I}_2$.

4.3. Results assessment

Assessing the consistency of the hierarchical representation of an image in a generic manner is a challenging task, as it greatly depends upon a specific application. A common approach is therefore to process the hierarchy accordingly, and appraise the obtained results with respect to the application. The hierarchical model is then declared to be relevant if it leads to proper results. For standard image segmentation purposes, hierarchical segmentation results are often assessed by comparing the algorithm outputs against manually delineated reference segmentation maps [37; 10; 38]. In the case of multimodal images however, it is much more difficult to proceed similarly, as available benchmark multimodal images are scarce and come without any reference ground truth data for segmentation applications. For those reasons, the assessment of hierarchical segmentations for multimodal images is often conducted either by visually comparing the multimodal segmentation result against the marginal segmentation outputs (when each modality is processed individually) [22].

To that extent, we propose here to evaluate the ability of the

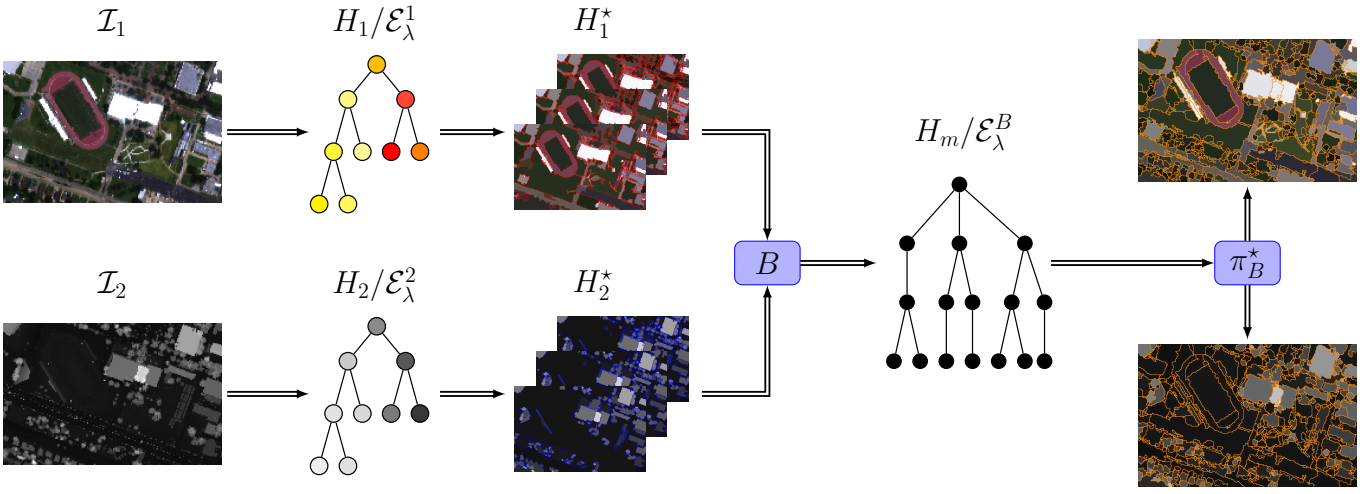

Fig. 6: Proposed braid-based multimodal segmentation methodology.

braid structure to represent multimodal images by comparing the braid optimal cut $\pi_B^\star$ against the two optimal cuts $\pi_1^\star$ and $\pi_2^\star$ extracted from $H_1^\star$ and $H_2^\star$ and containing the same (or a close) number of regions. In addition, we also compare the braid optimal cut with respect to $\pi_{[23]}^\star$, obtained following the method described in [23], where a common hierarchical representation is constructed for the various modalities of the multimodal images (more details are given in the supplementary materials page 13 and following pages). This allows a fair visual comparison since all four partitions $\pi_B^\star, \pi_1^\star, \pi_2^\star$ and $\pi_{[23]}^\star$ should feature regions of similar scales. In addition, the comparison of partitions with the same (or similar) complexity can be done by evaluating their closeness with respect to the data. For this reason, we compute the average GOF of $\pi_B^\star, \pi_1^\star, \pi_2^\star$ and $\pi_{[23]}^\star$ with respect to both modalities $\mathcal{I}_1$ and $\mathcal{I}_2$ as follows:

$$\epsilon(\pi|\mathcal{I}_i) = \frac{1}{|E|}\sum_{\mathcal{R}\in\pi}|\mathcal{R}|\times\Xi_i(\mathcal{R}) \tag{14}$$

with $|\mathcal{R}|$ denoting the cardinality of region $\mathcal{R}$, and $\Xi_i(\mathcal{R})$ is the Mumford-Shah GOF term defined in equation (12). Therefore, a consistent braid-based hierarchical representation of the multimodal image should lead to segmentation results competing with the optimal marginal segmentation $\pi_i^\star$ of each modality $\mathcal{I}_i$.

5. Experimental validation

The multimodal data set on which we investigate the proposed framework[2] is the Hyperspectral/LiDAR data set described in [39]. It is composed of a $342\times1903\times144$ hyperspectral (HS) image and a LiDAR-derived digital surface model, with the same ground-sampling distance of 2.5 m. Data were acquired over the campus of the University of Houston in 2012. The study site features a typical urban area with several houses and buildings of various shapes and heights, with roofs made of different materials, some parking lots, walkways, roads as well as portions of grass and trees. Due to the scarcity of co-registered data sets for the same given multimodal scenario, and in order to nonetheless provide some consistent results across different multimodal scenes (we recall however that conducting a fully exhaustive validation over a large multimodal data base is beyond the scope of this paper, as we aim to demonstrate here the potentiality of the braid structure for the hierarchical analysis of multimodal images), we selected 9 crops of size 150×200 (either horizontally or vertically) from the global data set and use them as a corpus of Hyperspectral/LiDAR multimodal images. Smaller image sizes also allow us to reduce the computational burden of the whole processing chain as well as having a better insight on the construction and processing of the braid structure and easing the results analysis. Figure 7 displays the Hyperspectral/LiDAR data set as well as the 9 selected crops (note the shaded portion on the right part of the HS image due to the presence of a cloud during the data acquisition). The HS/LiDAR complementarity lies in the fact that both modalities convey information of different physical nature (ground spectral reflectance for the HS modality and height above ground for the LiDAR). Thus, the modalities may be redundant in some part of the scene but well complement themselves in other parts, and the integration of this multimodal information is expected to resolve those potential errors in the optimal marginal segmentations.

5.1. Experimental Set-up

The first step of the braid-based multimodal image representation and segmentation methodology is to build the hierarchical representations of the various modalities, as shown by the workflow of Figure 6. While there is no special requirement on the chosen hierachical representation, we work in practice with the BPT, which has already proved to be very efficient for hierarchical image representation and segmentation [4; 40; 33]. The BPT representation of an image is governed by the definition of an initial partition of the image π_0, a region model $\mathcal{M}_\mathcal{R}$ and a merging criterion $O\left(\mathcal{R}_i,\mathcal{R}_j\right)$. Here, we use the mean spectrum and

[2]The same investigation is presented on a second multimodal data set in Section Appendix C of the supplementary materials

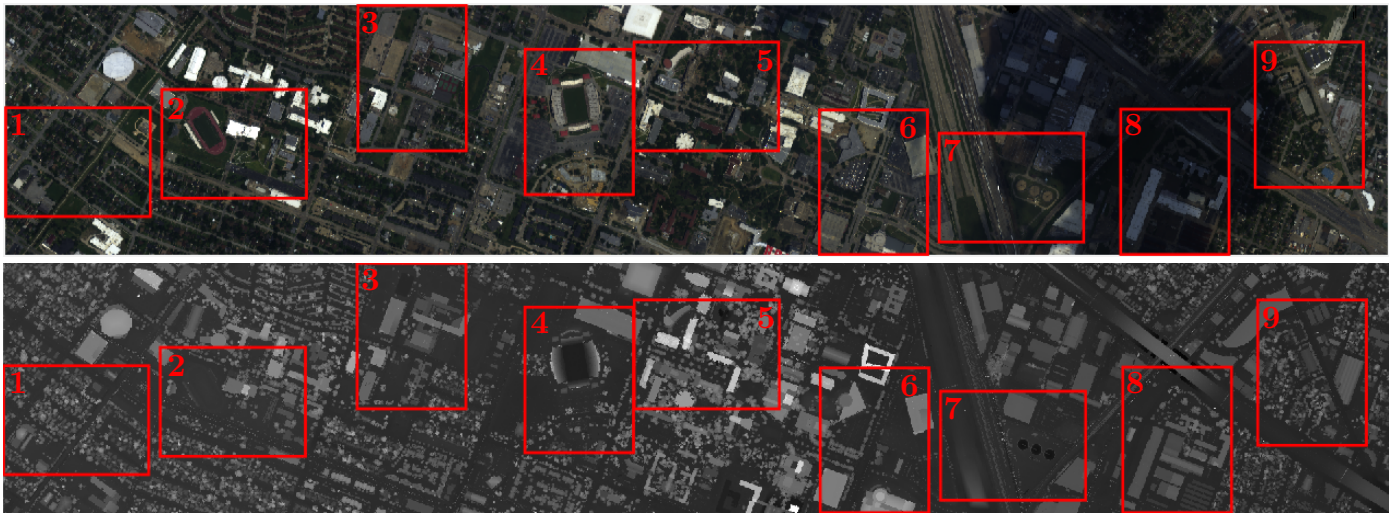

Fig. 7: Multimodal hyperspectral (top) and LiDAR (bottom) data set, with the 9 used crops (numbered from left to right).

spectral angle for the region model and merging criterion of the HS modality, and the mean value and Euclidean distance for the LiDAR modality, which can be considered as standard settings (see the aforementioned references for more details). Moreover, the two BPTs H_1 and H_2 are built on the same leaf partition π_0, which is obtained as the refinement infimum of two mean shift clustering procedures [41] conducted on each modality independently. The fact that H_1 and H_2 have the same leaf partition ensures that the three constrained set of cuts $H_2^{\widetilde{\supset}_h}(\pi_1^1)\backslash\{E\}$, $H_1^{\leq}(\pi_2^1)$ and $H_2^{\leq}(\pi_1^1)$ involved in the construction procedure are non empty since they all contain at least the leaf partition.

Constructing the braid B by following the procedure exposed in Figure A.11 raises the question from which hierarchy should the first cut should be chosen. While this is still an open question, we can provide the following empirical rule of thumb: the first cut should be extracted from the hierarchy built on the modality whose main regions of interest are the coarsest. Consequently, for all 9 crops, the first cut is extracted from the BPT built on the LiDAR modality (thus denoted $\mathcal{I}_1$ from hereon), since it contains less fine details than the HS modality (hence named $\mathcal{I}_2$).

Two parameters must be tuned to carry out the proposed workflow: the number of regions $|\pi_1^{1\star}|$ in the optimal cut $\pi_1^{1\star}$ that steers the following construction of the braid and its monitor hierarchy, and the value of the regularization parameter λ (equation (13)) that trades off between over- and under-segmentation for the braid optimal cut. Figure C.13 presents their respective influence (keeping constant the other parameter) on the number of regions $|\pi_B^\star|$ in the braid optimal cut $\pi_B^\star$ for 5 out of the 9 crops of the HSI/LiDAR multimodal data set. While the range of λ has empirically been set between 10^{-5} and 5.10^{-5}, the one of $|\pi_1^{1\star}|$ should roughly correspond to the number of expected large salient regions in $\mathcal{I}_1$. Consequently, it has been set between 100 and 300 with steps of 50.

Figure C.13 (left) not surprisingly reveals the decreasing behavior of $|\pi_B^\star|$ with respect to λ. As a matter of fact, the greater the λ, the higher the penalty on $|\pi_B^\star|$, hence the fewer the number of regions in the braid optimal cut $\pi_B^\star$. The influence of $|\pi_1^{1\star}|$ on $|\pi_B^\star|$, as displayed by Figure C.13 (right) is however much less clear, since crops #3, #5 and #7 show a slightly increasing behavior while crops #1 and #9 exhibits an minor decreasing trend. Despite $|\pi_B^\star|$ remaining relatively stable, as the nature of the regions composing $\pi_B^\star$ changes with the value of $|\pi_1^{1\star}|$, we intuited that this change should impact the average GOF value $\epsilon(\pi_B^\star|\mathcal{I}_i)$ with respect to both modalities $\mathcal{I}_1$ and $\mathcal{I}_2$. While Figure C.14 displays that it is indeed the case, the behavior of $\epsilon(\pi_B^\star|\mathcal{I}_i)$ with respect to $|\pi_1^{1\star}|$ does not show any clear trend that would give an insight on how to properly tune $|\pi_1^{1\star}|$ as it seems to depend on the content in the scene. Thus, λ and $|\pi_1^{1\star}|$ are respectively set to 3.10^{-5} and 200 in the following.

The collaborative method presented in [23] is implemented in a similar fashion: a unique BPT H [23] is built upon the same initial partition π_0, whose construction is parametrized using the same region models and merging criteria as for the marginal cases, with the additional consensus strategy being set to the *best median ranking*. The optimal cut $\pi_{[23]}^\star$ is then obtained from H [23] by minimizing energy (13) to produce the same, or a similar, number of regions than contained in $\pi_B^\star$.

5.2. Results

Table 1 presents the number of regions as well as the average GOF of optimal partitions $\pi_1^\star, \pi_2^\star, \pi_{[23]}^\star$ and $\pi_B^\star$ with respect to both modalities $\mathcal{I}_1$ (the LiDAR image) and $\mathcal{I}_2$ (the HS image), for all 9 crops. For each crop, the lowest modality-wise average GOF appears in bold. Several remarks arise from the analysis of table 1. First of all, it can be observed that most of the time, the marginal partition $\pi_i^\star$ scores the lowest average GOF with respect to its own modality $\mathcal{I}_i$ (5 out of 9 times for the LiDAR modality and even 8 out of 9 times for the HS modality). However, while $\pi_i^\star$ appears optimal with respect to its own modality, it consistently yields the worst result in respect of the other modality. This comes as no surprise since the marginal segmentation is, by construction, expected to fit only its own modality and cannot account for all the complementary features

Table 1: Size (top) and GOF with respect to $\mathcal{I}_1$ (bottom left) and $\mathcal{I}_2$ (bottom right) of optimal partitions with respect to the 9 crops, and their overall average rank (last column). For each crop, the lowest modality-wise GOF is in bold.

	1	2	3	4	5	6	7	8	9	avg rank
$\pi_1^\star$	872 653 | 56.8	522 2919 | 491.3	507 600 | 68.4	458 623 | 50.9	457 **549** | 26.2	592 **1201** | 55.2	325 **251** | 15.7	348 **520** | 8.1	702 **377** | 48.2	2.0 | 4.0
$\pi_2^\star$	868 651 | **8.9**	523 1125 | **30.2**	511 906 | **14.6**	458 1101 | 23.3	458 999 | **18.8**	589 2531 | **15.1**	325 617 | **6.3**	349 2939 | **2.2**	702 1612 | **7.8**	3.78 | **1.22**
$\pi_{[23]}^\star$	867 **413** | 9.3	526 951 | 40.5	511 556 | 17.5	457 686 | **21.9**	457 639 | 21.4	594 1742 | 19.0	325 346 | 8.4	348 571 | 3.4	702 380 | 11.0	2.33 | 2.22
$\pi_B^\star$	867 476 | 10.9	522 **851** | 32.5	509 **445**| 16.1	457 **547** | 22.7	458 573 | 25.6	589 1662 | 19.7	325 260 | **6.3**	348 698 | 4.4	702 394 | 12.4	**1.89** | 2.56

present in the other modality. The collaborative BPT partition $\pi_{[23]}^\star$ surpasses the other strategies only once for each modality and ranks second best twice with respect to $\mathcal{I}_1$ and 6 times with respect to $\mathcal{I}_2$. On the other hand, the braid optimal partition $\pi_B^\star$ outperforms $\pi_1^\star$ 3 times over $\mathcal{I}_1$ (ranking second best 4 other times) and equates $\pi_2^\star$ once (ranking second best twice). Either way, both the braid and the consensus BPT strategy are able to integrate some multimodal information within their structure, yielding optimal segmentations that better describe "in average" both modalities at the same time: the descriptive accuracy and robustness of a multimodal image are increased thanks to the complementarity (for the former) and redundancy (for the latter) of the information contained by each single modality. The behavior of both approaches appears rather similar in terms of performances, as they most of the time achieve a trade-off between the two marginal segmentations in terms of average GOF. Still, $\pi_{[23]}^\star$ seems to perform better with respect to $\mathcal{I}_2$ than $\mathcal{I}_1$ while the opposite observation holds for $\pi_B^\star$. This bias of $\pi_B^\star$ better fitting $\mathcal{I}_1$ than $\mathcal{I}_2$ might come from the fact that the first extracted partition $\pi_1^{\star\star}$ was extracted from $H_1^\star$ during the braid construction procedure. These observations are confirmed by the average ranks of all compared approaches with respect to both modalities.

Figure 10 shows the optimal LiDAR marginal partition $\pi_1^\star$, the optimal HS marginal partition $\pi_2^\star$, the optimal collaborative partition $\pi_{[23]}^\star$ and the braid optimal partition $\pi_B^\star$, represented by their mean height with respect to $\mathcal{I}_1$ and their mean RGB value with respect to $\mathcal{I}_2$. By lack of room, only crops #2, #5 and #7 are displayed. The qualitative analysis of Figure 10 leads to similar conclusions. While $\pi_1^\star$ is able to correctly segment all notable regions of the LiDAR modality for all 3 crops, it fails at segmenting regions with similar height but not made of the same materials. This is notably the case of the running track and football pitch in the center of crop #2, the roads and walkways in the left and center of crop #5 and the lawns aroud the water treatment plant in the center of crop #7. Contrarily, $\pi_2^\star$ is able to preserve the spectrally salient regions for all three crops. Regions that have close spectral signatures but not the same height are however generally mis-segmented in $\pi_2^\star$. This is in particular the case in crops #2 and #5 where several batches of trees are either grouped together, or fused with the neighboring grass (whose spectral response is rather close). A slightly different issue appears in crop #7, where the gradient of the sloped grassy area in the bottom left corner is not well preserved, since this information is obviously transparent to the HS modality. While it is visually complicated to distinguish notable differences between $\pi_{[23]}^\star$ and $\pi_B^\star$, a careful inspection reveals that the latter seems to better retain small details (such as isolated trees or walkways) than the former, possibly due to the fact that the consensus strategy implemented in order to obtain $\pi_{[23]}^\star$ averages out small features. However, the impossibility of creating reliable ground truth images makes really challenging the fair comparison of both hierarchical multimodal approaches. Nevertheless, all erroneously segmented regions of the marginal partitions $\pi_1^\star$ and $\pi_2^\star$ appear correctly delineated in both the collaborative partition $\pi_{[23]}^\star$ and the braid optimal cut $\pi_B^\star$.

6. Conclusion

In conclusion, we presented in this article a novel methodology for the hierarchical representation and segmentation of multimodal images by taking advantage of the newly introduced concept of braids of partitions. We showed that such structures were well suited to describe the inherent redundant and complementary information contained within multimodal images, and were thus relevant hierarchical representations for such images. Because of the lack of clear guidelines to check the validity of such structure given a family of partitions, we proposed here an iterative procedure to extract two cuts from two different and supposedly unrelated hierarchies and guarantee that they form a braid. Following, we endowed the resulting braid structure with an energy minimization framework in order to obtain an optimal partition of the multimodal data. In particular, we extended the classical piece-wise constant Mumford-Shah energy function to multimodal images for segmentation purposes. We investigated the proposed methodology on 9 different crops extracted from a Hyperspectral/LiDAR multimodal data set and 9 multimodal RGB/Depth images from the Middleburry Stereo data set (presented in supplementary materials). In particular, we conducted a sensitivity analysis to the two parameters from which the proposed braid-based multimodal segmentation framework depends. While the impact of the regularization parameter occuring in the energy minimization process is well understood, the influence of the nature of the first extracted cut steering the whole braid construction procedure will required deeper investigations. Nevertheless, the obtained results quantitatively and qualitatively

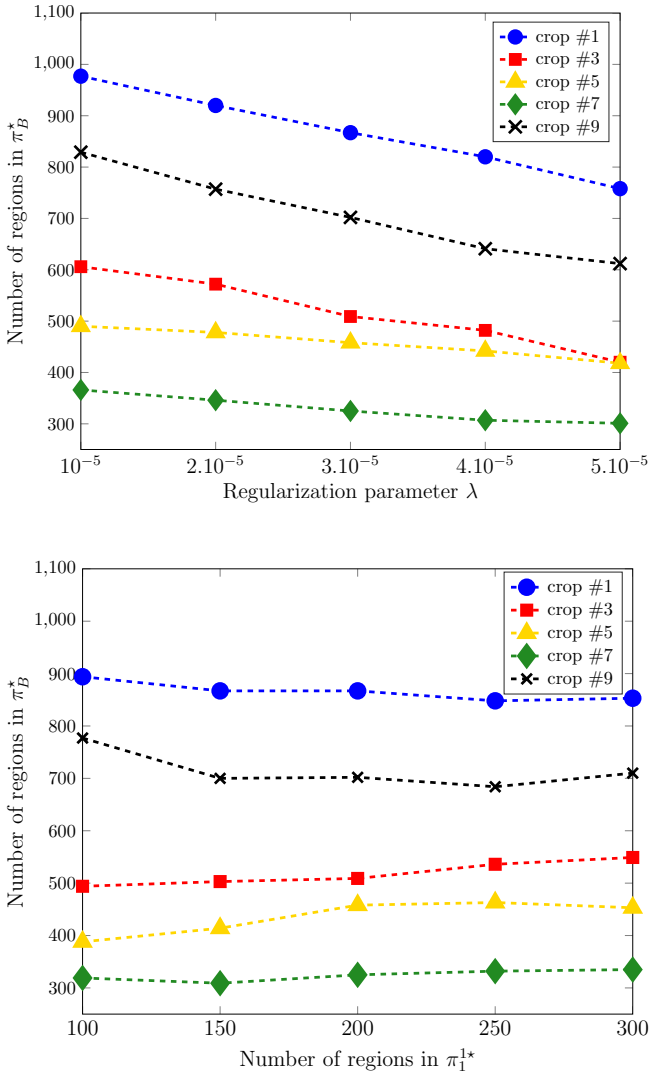

Fig. 8: Influence of (left) regularization parameter λ (for $|\pi_1^{1\star}| = 200$) and (right) number of regions in partition $\pi_1^{1\star}$ for the braid construction procedure (for $\lambda = 3.10^{-5}$) on the size of the braid optimal cut.

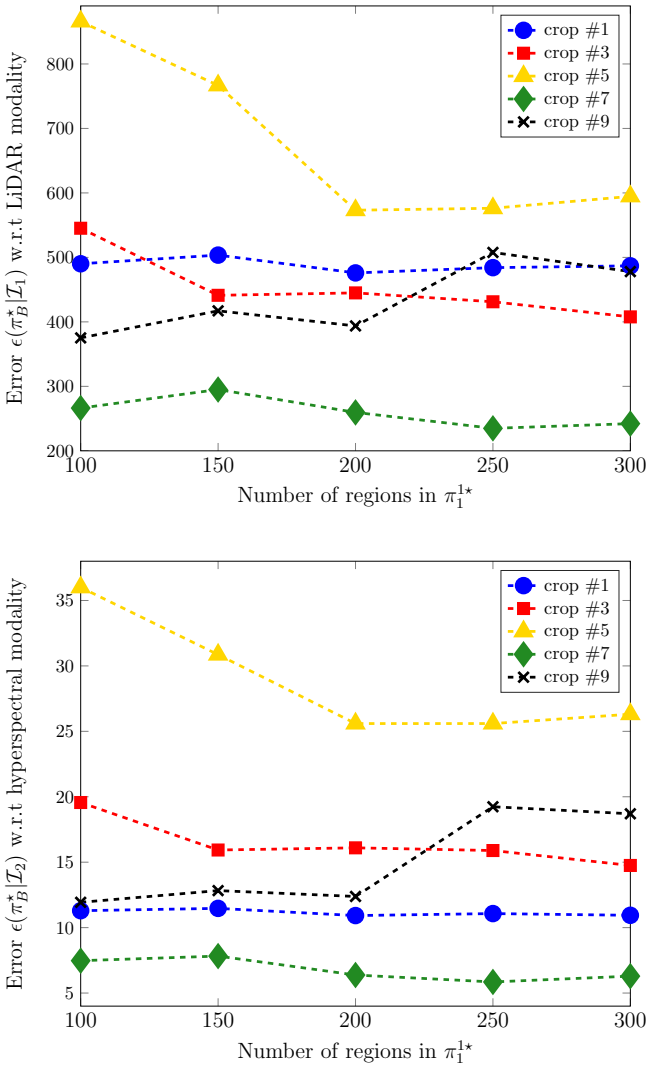

Fig. 9: Influence of $|\pi_1^{1\star}|$ (for $\lambda = 3.10^{-5}$) on the average GOF for the braid optimal cut with respect to both modalities $\mathcal{I}_1$ (LiDAR) and $\mathcal{I}_2$ (hyperspectral).

demonstrated that the braid structure is able to produce a segmentation that not only retains salient regions shared by both modalities, but also regions appearing in only one modality of the multimodal image ; being close to typical marginal segmentation results obtained by considering only one modality at a time and competing with the collaborative BPT strategy.
Up-to-now, the proposed framework is restricted only to multimodal images composed of two co-registered modalities, and the construction of the braid of partition is bound to the extraction of two different cuts per individual hierarchical representation, limiting its use to multimodal data set featuring clear redundancy and complementarity at the same time. In that respect, future work will investigate theoretical aspects related to the construction of the braid of partitions, namely how to extract more cuts coming from various hierarchies and still maintain the braid structure, and how to relax the constraint on the co-registration of the considered modalities. Practical consideration such as a more in-depth quantitative evaluation of the braid structure, as well as the investigation of other applications than segmentation will also be considered.

Acknowledgments

This work was partially funded through the ERC CHESS project, ERC-12-AdG-320684-CHESS, and DECODA, Grant Agreement no. 320594 "DECODA".

References

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

[2] P. Monasse, F. Guichard, Fast computation of a contrast-invariant image representation, Image Processing, IEEE Transactions on 9 (5) (2000) 860–872.

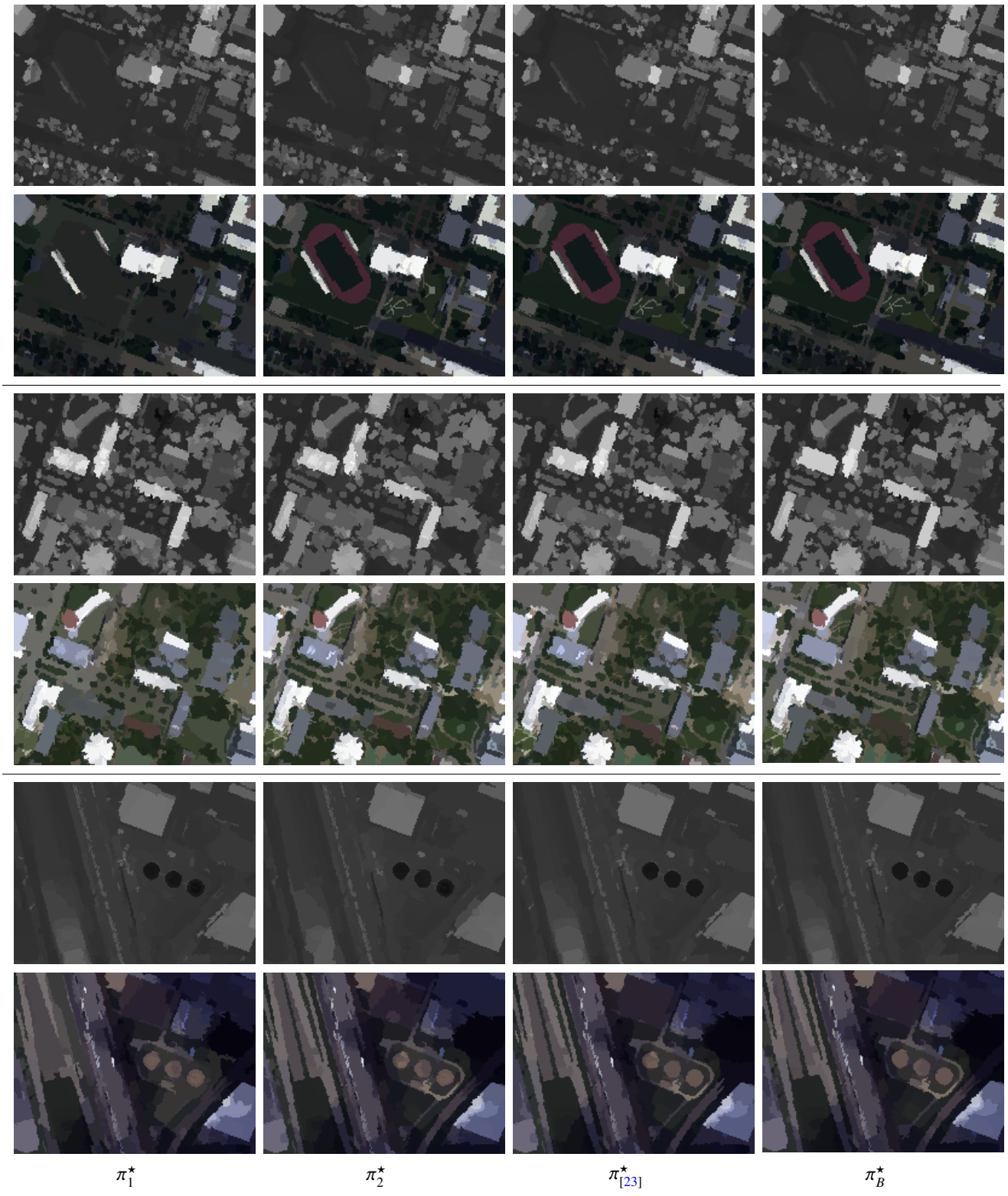

Fig. 10: Optimal partitions $\pi_1^\star$, $\pi_2^\star$, $\pi_{[23]}^\star$ and $\pi_B^\star$ for crops #2 (top two rows), #5 (middle two rows) and #7 (bottom two rows) represented with their mean height with respect to the LiDAR modality I_1 and mean RGB value with respect to the HS modality I_2.

[3] P. Soille, Constrained connectivity for hierarchical image partitioning and simplification, Pattern Analysis and Machine Intelligence, IEEE Transactions on 30 (7) (2008) 1132–1145.

[4] P. Salembier, L. Garrido, Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval, Image Processing, IEEE Transactions on 9 (4) (2000) 561–576.

[5] L. Najman, J. Cousty, A graph-based mathematical morphology reader, Pattern Recognition Letters 47 (2014) 3–17.

[6] P. Bosilj, E. Kijak, S. Lefèvre, Partition and inclusion hierarchies of images: A comprehensive survey, Journal of Imaging 4 (2) (2018) 33.

[7] Y. Xu, T. Géraud, L. Najman, Connected filtering on tree-based shape-spaces, IEEE transactions on pattern analysis and machine intelligence 38 (6) (2016) 1126–1140.

[8] C. Kurtz, B. Naegel, N. Passat, Connected filtering based on multivalued component-trees, IEEE Transactions on Image Processing 23 (12) (2014) 5152–5164.

[9] Y. Xu, T. Géraud, L. Najman, Hierarchical image simplification and segmentation based on Mumford-Shah-salient level line selection, Pattern Recognition Letters.

[10] B. Perret, J. Cousty, S. J. F. Guimaraes, D. S. Maia, Evaluation of hierarchical watersheds, IEEE Transactions on Image Processing 27 (4) (2018) 1676–1688.

[11] B. Perret, J. Cousty, O. Tankyevych, H. Talbot, N. Passat, Directed connected operators: Asymmetric hierarchies for image filtering and segmentation, IEEE transactions on pattern analysis and machine intelligence 37 (6) (2015) 1162–1176.

[12] C. Kurtz, N. Passat, P. Gancarski, A. Puissant, Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology, Pattern Recognition 45 (2) (2012) 685–706.

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

[14] J. Cousty, L. Najman, Y. Kenmochi, S. Guimarães, Hierarchical segmentations with graphs: quasi-flat zones, minimum spanning trees, and saliency maps, Journal of Mathematical Imaging and Vision 60 (4) (2018) 479–502.

[15] D. S. Maia, A. de Albuquerque Araujo, J. Cousty, L. Najman, B. Perret, H. Talbot, Evaluation of combinations of watershed hierarchies, in: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2017, pp. 133–145.

[16] S. Beucher, Watershed, hierarchical segmentation and waterfall algorithm, in: Mathematical morphology and its applications to image processing, Springer, 1994, pp. 69–76.

[17] D. Lahat, T. Adali, C. Jutten, Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects, Proceedings of the IEEE 103 (9) (2015) 1449–1477.

[18] G. Tochon, Hierarchical analysis of multimodal images, Ph.D. thesis, Université Grenoble Alpes (2015).

[19] E. Carlinet, T. Géraud, MToS: A tree of shapes for multivariate images, IEEE Transactions on Image Processing 24 (12) (2015) 5330–5342.

[20] N. Passat, B. Naegel, Component-trees and multivalued images: Structural properties, Journal of Mathematical Imaging and Vision 49 (1) (2014) 37–50.

[21] G. Palou, P. Salembier, Hierarchical video representation with trajectory binary partition tree, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2099–2106.

[22] J. F. Randrianasoa, C. Kurtz, É. Desjardin, N. Passat, Multi-image Segmentation: A Collaborative Approach Based on Binary Partition Trees, in: Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2015, pp. 253–264.

[23] J. F. Randrianasoa, C. Kurtz, E. Desjardin, N. Passat, Binary partition tree construction from multiple features for image segmentation, Pattern Recognition 84 (2018) 237–250.

[24] B. R. Kiran, J. Serra, Braids of partitions, in: Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2015, pp. 217–228.

[25] G. Tochon, M. Dalla Mura, J. Chanussot, Segmentation of Multimodal Images based on Hierarchies of Partitions, in: Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2015, pp. 241–252.

[26] C. Ronse, Partial partitions, partial connections and connective segmentation, Journal of Mathematical Imaging and Vision 32 (2) (2008) 97–125.

[27] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Communications on Pure and Applied Mathematics 42 (5) (1989) 577–685.

[28] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, Pattern Analysis and Machine Intelligence, IEEE Transactions on 23 (11) (2001) 1222–1239.

[29] S. Li, Markov random field modeling in computer vision, Springer-Verlag New York, Inc., 1995.

[30] L. Guigues, J. Cocquerez, H. Le Men, Scale-sets image analysis, International Journal of Computer Vision 68 (3) (2006) 289–317.

[31] J. Serra, Hierarchies and optima, in: Discrete Geometry for Computer Imagery, Springer, 2011, pp. 35–46.

[32] B. Kiran, J. Serra, Global–local optimizations by hierarchical cuts and climbing energies, Pattern Recognition 47 (1) (2014) 12–24.

[33] M. Veganzones, G. Tochon, M. Dalla Mura, A. Plaza, J. Chanussot, Hyperspectral Image Segmentation Using a New Spectral Unmixing-Based Binary Partition Tree Representation, Image Processing, IEEE Transactions on 23 (8) (2014) 3574–3589.

[34] B. R. Kiran, J. Serra, Ground truth energies for hierarchies of segmentations, in: Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2013, pp. 123–134.

[35] J. Angulo, D. Jeulin, Stochastic watershed segmentation, in: Proceedings of the 8th International Symposium on Mathematical Morphology, 2007, pp. 265–276.

[36] C. Ballester, V. Caselles, L. Igual, L. Garrido, Level lines selection with variational models for segmentation and encoding, Journal of Mathematical Imaging and Vision 27 (1) (2007) 5–27.

[37] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (5) (2011) 898–916.

[38] J. F. Randrianasoa, C. Kurtz, P. Gançarski, E. Desjardin, N. Passat, Evaluating the quality of binary partition trees based on uncertain semantic ground-truth for image segmentation, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 3874–3878.

[39] 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, 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) (2014) 2405–2418.

[40] S. Valero, P. Salembier, J. Chanussot, Hyperspectral Image Representation and Processing With Binary Partition Trees, Image Processing, IEEE Transactions on 22 (4) (2013) 1430–1443.

[41] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, Pattern Analysis and Machine Intelligence, IEEE Transactions on 24 (5) (2002) 603–619.

[42] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, P. Westling, High-resolution stereo datasets with subpixel-accurate ground truth, in: German Conference on Pattern Recognition, Springer, 2014, pp. 31–42.

[43] J. Hopcroft, R. Tarjan, Algorithm 447: efficient algorithms for graph manipulation, Communications of the ACM 16 (6) (1973) 372–378.

Supplementary materials

Appendix A. Construction of the braid

Let H_1 and H_2 be two hierarchies of partitions built over the same space E. We prove here that the following procedure (described in section 4.1, and illustrated by figure A.11) allows to create a braid structure:

1. First select arbitrarily some cut $\pi_1^1 \in \Pi_E(H_1)$.
2. Then choose a cut π_2^1 in the constrained set $H_2^{\simeq_h}(\pi_1^1)\backslash\{E\}$, that is, a cut from H_2 which is h-equivalent to π_1^1 and different from the whole space $\{E\}$.
3. Finally, complete by taking a cut in each hierarchy that is a refinement of the cut previously extracted from the other hierarchy, that is $\pi_i^2 \in \Pi_E(H_i), i \in \{1,2\}$ such that $\pi_1^2 \leq \pi_2^1$ and $\pi_2^2 \leq \pi_1^1$.

Proposition 4. *Under this configuration, $B = \{\pi_1^1, \pi_1^2, \pi_2^1, \pi_2^2\}$ has a braid structure.*

Proof. Let $B = \{\pi_1^1, \pi_1^2, \pi_2^1, \pi_2^2\}$ be a family of partitions composed following the previously described procedure, and let $\pi_{i,j}^{k,l} = \pi_i^k \vee \pi_j^l$ denote the pairwise refinement suprema of partitions in B. In particular, the 4 partitions composing B generates $\binom{4}{2} = 6$ different pairwise refinement suprema $\pi_{1,1}^{1,2}, \pi_{1,2}^{1,1}, \pi_{1,2}^{1,2}, \pi_{1,2}^{2,1}, \pi_{1,2}^{2,2}, \pi_{2,2}^{1,2}$. Checking that B is a braid amounts to verify whether the $\pi_{i,j}^{k,l}$ all defines cuts of the same monitor hierarchy H_m, which is equivalent to showing that they are (at least) all h-equivalent to each other. In order to show the braid structure of B, we first demonstrate the following result:

Lemma 1. *Let $\pi_1, \pi_2, \pi_3 \in \Pi_E$ be some partitions of E such that $\pi_1 \simeq_h \pi_3$ and $\pi_2 \leq \pi_3$. Then $\pi_1 \vee \pi_2 \simeq_h \pi_3$.*

Proof. If $\pi_1 \leq \pi_3$, then $\pi_1 \vee \pi_2 \leq \pi_3$ by definition of the refinement supremum, and so $\pi_1 \vee \pi_2 \simeq_h \pi_3$ since any two ordered partitions are also h-equivalent.
On the other hand, if $\pi_1 \geq \pi_3$, then $\pi_1 \geq \pi_2$, hence $\pi_1 \vee \pi_2 = \pi_1$ and so $\pi_1 \vee \pi_2 \simeq_h \pi_3$ for the same reason as above.
In the most general case where π_1 and π_3 are h-equivalent but can nonetheless not be ordered, it means that π_1 is a refinement of π_3 in some parts of E, and is refined by π_3 in the other parts. In the former case, let $\mathcal{R}_3$ be a region of π_3 and $\pi_1(\mathcal{R}_3), \pi_2(\mathcal{R}_3)$ be the refinements (partial partitions) of $\mathcal{R}_3$ in π_1 and π_2. Then, $\pi_1(\mathcal{R}_3) \vee \pi_2(\mathcal{R}_3)$ is also a refinement of $\mathcal{R}_3$, implying that $\pi_1 \vee \pi_2$ refines π_3 in the part of E covered by $\mathcal{R}_3$. In the case where π_3 is locally a refinement of π_1, then given $\mathcal{R}_1 \in \pi_1$, there exists a refinement $\pi_3(\mathcal{R}_1)$ of $\mathcal{R}_1$ in π_3, and therefore a refinement $\pi_2(\mathcal{R}_1)$ of $\mathcal{R}_1$ in π_2 since $\pi_2 \leq \pi_3$. Therefore, $\{\mathcal{R}_1\} \vee \pi_2(\mathcal{R}_1) = \{\mathcal{R}_1\}$ and thus π_3 refines $\pi_1 \vee \pi_2$ in the part of E covered by $\mathcal{R}_1$. Finally, $\pi_1 \vee \pi_2$ either refines or is refined by π_3 in all parts of E, hence $\pi_1 \vee \pi_2 \simeq_h \pi_3$. □

To ease the reading of the proof, we first recall the relations holding between the various partitions composing the braid B:

- $\pi_1^1 \simeq_h \pi_2^1$ by construction.
- $\pi_1^2 \leq \pi_2^1$ and $\pi_2^2 \leq \pi_1^1$ by construction.
- $\pi_1^1 \simeq_h \pi_1^2$ beacuse they are both cuts of the same hierarchy H_1. Similarly, $\pi_2^1 \simeq_h \pi_2^2$.

Following, we prove that all the pairwise refinement suprema of B are at least all h-equivalent to each other. Their relationships are summarized in table A.2.

Table A.2: Summary of the relationships holding between all pairwise refinement suprema of B with their corresponding item in the proof.

	$\pi_{1,1}^{1,2}$	$\pi_{1,2}^{1,1}$	$\pi_{1,2}^{1,2}$	$\pi_{1,2}^{2,1}$	$\pi_{1,2}^{2,2}$	$\pi_{2,2}^{1,2}$
$\pi_{1,1}^{1,2}$	X	1. $\leq$	2. $\leq$	3. $\simeq_h$	4. $\leq$	5. $\simeq_h$
$\pi_{1,2}^{1,1}$		X	6. $\leq$	7. $\leq$	8. $\leq$	9. $\leq$
$\pi_{1,2}^{1,2}$			X	10. $\simeq_h$	11. $\simeq_h$	12. $\simeq_h$
$\pi_{1,2}^{2,1}$				X	13. $\simeq_h$	14. $\leq$
$\pi_{1,2}^{2,2}$					X	15. $\leq$
$\pi_{2,2}^{1,2}$						X

1. $\pi_{1,1}^{1,2} = \pi_1^1 \vee \pi_1^2$. As $\pi_1^2 \leq \pi_2^1$ by construction of B, it follows that $\pi_1^1 \vee \pi_1^2 \leq \pi_1^1 \vee \pi_2^1$, hence $\pi_{1,1}^{1,2} \leq \pi_{1,2}^{1,1}$.
2. $\pi_{1,2}^{1,2} = \pi_1^1 \vee \pi_2^2 = \pi_1^1$ as $\pi_2^2 \leq \pi_1^1$ by construction of B. By property of the refinement supremum, one has $\pi_1^1 \leq \pi_1^1 \vee \pi_1^2 = \pi_{1,1}^{1,2}$, hence $\pi_{1,2}^{1,2} \leq \pi_{1,1}^{1,2}$.
3. By construction of B, one has $\pi_1^1 \simeq_h \pi_2^1$ and $\pi_1^2 \leq \pi_2^1 = \pi_{1,2}^{2,1}$. Using lemma 1, it follows that $\pi_1^1 \vee \pi_1^2 = \pi_{1,1}^{1,2} \simeq_h \pi_{1,2}^{2,1}$.
4. $\pi_2^2 \leq \pi_1^1$ by construction of B, meaning that $\pi_1^2 \vee \pi_2^2 \leq \pi_1^2 \vee \pi_1^1$, hence $\pi_{1,2}^{2,2} \leq \pi_{1,1}^{1,2}$.
5. Using item 3, we first have $\pi_{1,1}^{1,2} \simeq_h \pi_2^1 = \pi_{1,2}^{2,1}$. In addition, $\pi_2^2 \leq \pi_1^1$ by construction of B, implying that $\pi_2^2 \leq \pi_1^1 \vee \pi_1^2 = \pi_{1,1}^{1,2}$. Using lemma 1 finally leads to $\pi_{1,1}^{1,2} \simeq_h \pi_{2,2}^{1,2}$.
6. $\pi_{1,2}^{1,2} = \pi_1^1$ as $\pi_2^2 \leq \pi_1^1$ by construction of B. The basic property of the refinement supremum allows to conclude that $\pi_1^1 \leq \pi_1^1 \vee \pi_2^1$, hence $\pi_{1,2}^{1,2} \leq \pi_{1,2}^{1,1}$.
7. The exact same reasoning as item 6 applied to $\pi_{1,2}^{2,1} = \pi_2^1$ leads to $\pi_{1,2}^{2,1} \leq \pi_{1,2}^{1,1}$.
8. $\pi_1^2 \leq \pi_2^1$ and $\pi_2^2 \leq \pi_1^1$, both by construction of B. It immediately follows that $\pi_1^2 \vee \pi_2^2 \leq \pi_1^1 \vee \pi_2^1$, hence $\pi_{1,2}^{2,2} \leq \pi_{1,2}^{1,1}$.
9. The same reasoning as item 1 applies to $\pi_{2,2}^{1,2} = \pi_2^1 \vee \pi_2^2$, leading to $\pi_{2,2}^{1,2} \leq \pi_{1,2}^{1,1}$.
10. By construction of B, one has $\pi_1^1 = \pi_{1,2}^{1,2} \simeq_h \pi_{1,2}^{2,1} = \pi_2^1$, hence the result.
11. $\pi_{1,2}^{1,2} = \pi_1^1$ as $\pi_2^2 \leq \pi_1^1$ by construction of B. In addition, $\pi_1^1 \simeq_h \pi_1^2$ as they are both cuts of the same hierarchy H_1. Using lemma 1, it follows that $\pi_1^1 = \pi_{1,2}^{1,2} \simeq_h \pi_{1,2}^{2,2} = \pi_1^2 \vee \pi_2^2$.
12. The same reasoning as item 3 applies to $\pi_{2,2}^{1,2}$ and $\pi_{1,2}^{1,2} = \pi_1^1$ and, relying upon lemma 1, leads to $\pi_{1,2}^{1,2} \simeq_h \pi_{2,2}^{1,2}$.
13. The same reasoning as item 11 applies to $\pi_{1,2}^{2,1} = \pi_2^1$ and $\pi_{1,2}^{2,2}$, leading to $\pi_{1,2}^{2,1} \simeq_h \pi_{1,2}^{2,2}$.
14. The same reasoning as item 2 applies to $\pi_{2,2}^{1,2}$ and $\pi_{1,2}^{2,1} = \pi_2^1$, leading to $\pi_{1,2}^{2,1} \leq \pi_{2,2}^{1,2}$.
15. The same reasoning as item 4 applies to $\pi_{1,2}^{2,2}$ and $\pi_{2,2}^{1,2}$, leading to $\pi_{1,2}^{2,2} \leq \pi_{2,2}^{1,2}$.

Finally, all the pairwise refinement supremum $\pi_{i,j}^{k,l} = \pi_i^k \vee \pi_j^l$ that can be formed using the partitions belonging to B are (at least) all h-equivalent to each other. Therefore, there exists some hierarchy H_m such that all $\pi_{i,j}^{k,l} \in \Pi_E(H_m)$, which proves that B has a braid structure when constructed following the proposed procedure. □

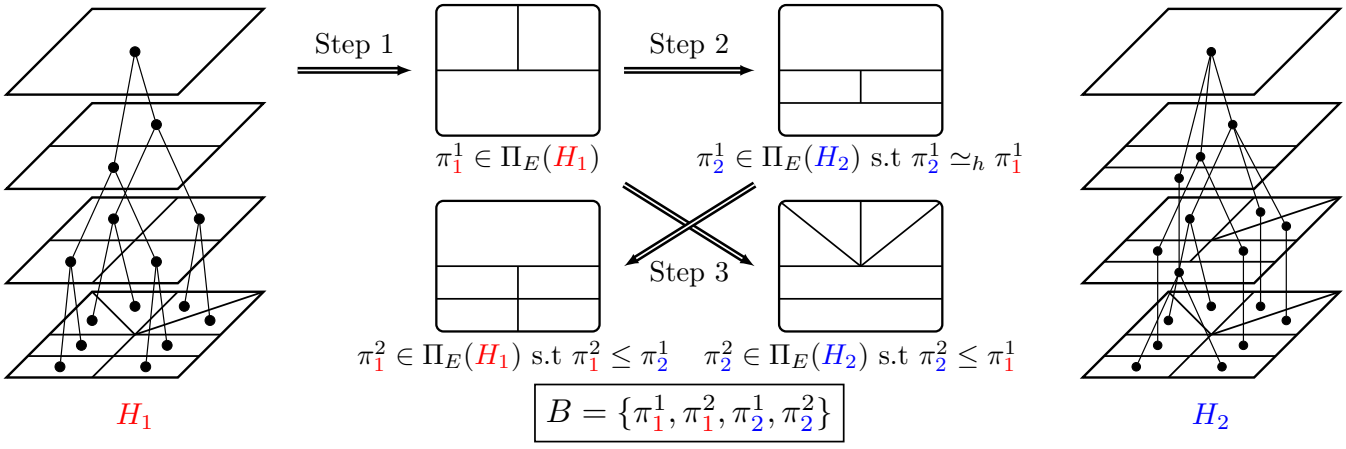

Fig. A.11: Composing a braid B with cuts from two hierarchies H_1 and H_2.

Appendix B. Description of the collaborative BPT approach

We briefly describe here the concept of collaborative binary partition tree [4] (BPT) representation for multimodal images, as was presented in [22; 23]. Contrarily to our proposed approach that performs the fusion process after independent hierarchical representations have been built for each modality, the work presented in [23] operate at the feature level. It builds a single hierarchical representation (namely a BPT), common to all modalities, by somehow pooling together features originating from those various modalities. In the classical setting, the BPT representation relies on some initial partition π_0 of the image and a region merging process parametrized by some region model $\mathcal{M}_\mathcal{R}$ and associated merging criterion $O(\mathcal{R}_i, \mathcal{R}_j)$. More specifically, neighboring regions $\mathcal{R}_i$ and $\mathcal{R}_j$ are iteratively merged based on their similarity (measured by the merging criterion distance $O(\mathcal{R}_i, \mathcal{R}_j)$ evaluated on their respective regions models $\mathcal{M}_{\mathcal{R}_i}$ and $\mathcal{M}_{\mathcal{R}_j}$) until the only remaining region is the whole image support E. The application of this region merging process amounts to the definition of a merging order list $\mathcal{W}$ that stores all pairwise distance between all neighboring regions. This list is updated at each iteration after the fusion of the two regions whose distance is the smallest. In the collaborative framework described in [23], each region $\mathcal{R}$ during the BPT construction has a specific region model $\mathcal{M}_{k,\mathcal{R}}$ and merging criterion $O_k(\mathcal{R}_i, \mathcal{R}_j)$ with respect to each modality $\mathcal{I}_k, k = 1, \ldots, K$. Thus, there is no longer one merging order list $\mathcal{W}$, but as many as the number of modalities $\mathcal{W}_k$. Each merging iteration therefore requires a consensus step to determine which couple of neighboring regions should be fused, according to their position in the various lists $\mathcal{W}_k$. Several consensus strategies were proposed in [23], and we selected the *best median ranking* one, which is implemented as follows: each couple of neighboring regions $(\mathcal{R}_i, \mathcal{R}_j)$ receives K ranks $r_1, \ldots, r_K$, where $r_k = n$ means that $O_k(\mathcal{R}_i, \mathcal{R}_j)$ is the n$^{\text{th}}$ smallest value in the merging order list $\mathcal{W}_k$. The rank attributed to the couple $(\mathcal{R}_i, \mathcal{R}_j)$ after consensus is then defined as the median value of its individual ranks $r_1, \ldots, r_K$, allowing the creation of a consensus merging order list $\mathcal{W}^*$. Finally, the fused couple for the current merging iteration is the one whose rank is the smallest in $\mathcal{W}^*$. Each list $\mathcal{W}_k$ is then updated after the merging, allowing to regenerate the consensus list $\mathcal{W}^*$. The merging process then goes on until the whole image support E has been reached.

Appendix C. Additional experimental validation

In addition to the Hypersectral/LiDAR data set presented in the main article, we consider here a second multimodal data set denoted RGB/Depth in the following. This multimodal scenario originates from the Middlebury Stereo Dataset [42], which features high-resolution left and right views of several indoor scenes as well as their associated disparity maps. While the primary purpose of this database is to provide a rigorous framework for the evaluation of setero algorithms, we take advantage here of the natural complementarity between color and depth information in order to validate the soundness of the proposed braid framework.

Therefore, we picked 9 different scenes out the 2014 database[3]. In each case, we retained the left-view color image and corresponding disparity map, thus composing the 9 RGB/depth multimodal images as displayed in figure C.12. Each image was resized to have width of 400 pixels (leading to a height comprised between 266 pixels and 275 pixels, depending on the case). While the RGB modality is always expected to complement the depth information (since it contains finer details, hence more semantic regions), the selected scene were chosen to also feature salient adjacent regions with different depths but of similar colors.

Appendix C.1. Experimental Set-up

We replicate on the RGB/Depth multimodal data set the experimental set-up that was conducted on the HSI/LiDAR data set. Two BPTs H_1 and H_2 still serve as the base hierarchical representations for each modality $\mathcal{I}_1$ and $\mathcal{I}_2$. The mean color/depth is used as the region model, the merging criterion is defined as the Euclidean distance, and the leaf partition is obtained as the refinement infimum of two mean shift procedures conducted independently on each modality. Those two BPTs H_1 and H_2 are subsequently transformed into their persistent versions $H_1^\star$ and $H_2^\star$, on which is conducted the braid construction procedure (recalled in section Appendix A and figure A.11 of this supplementary material). For each multimodal RGB/Depth scene, we still follow the rule of thumb stated in the main manuscript: the first cut $\pi_1^{1\star}$ steering the whole construction process is extracted from the modality whose salient regions are the coarsest. Thus, from now on, the modality $\mathcal{I}_1$ stands for the Depth while $\mathcal{I}_2$ represents the RGB modality.

Following what we did for the HSI/LiDAR data set, we investigate the

[3] http://vision.middlebury.edu/stereo/data/scenes2014/

Fig. C.12: Selected RGB/Depth multimodal images from the Middleburry stereo data set.

sensibility of the braid optimal cut $\pi_B^\star$ to the two parameters it depends on, namely the number of regions $|\pi_1^{1\star}|$ in the cut $\pi_1^{1\star}$ and the value of the regularization parameter λ. Figure C.13 presents their respective influence (keeping constant the other parameter) on the number of regions $|\pi_B^\star|$ in the braid optimal cut $\pi_B^\star$ for 4 out of the 9 scenes of the RGB/Depth multimodal data set. The range of λ has again been set empirically between 10^{-5} and 5.10^{-5} with steps of 10^{-5} while the one of $|\pi_1^{1\star}|$ is defined between 100 and 300 by steps of 50.

This first observation arising from figure C.13 is that using the same range for λ and $|\pi_1^{1\star}|$ leads to a braid optimal cut $\pi_B^\star$ whose number of regions is in the same order of magnitude than for the HSI/LiDAR data set (a few hundreds in each case). This is particularly interesting for λ, whose setting in any optimization problem is generally achieved by trial-and-error strategies since its optimal numerical value is often bounded to the image nature and content. For the braid optimal cut, the GOF normalization performed in the definition of the multimodal equation $\mathcal{E}_\lambda^B$ (see equation (14) in the main article) seemingly ensures the same correspondance between the ranges of λ and $|\pi_B^\star|$, regardless of the investigated multimodal scenario.

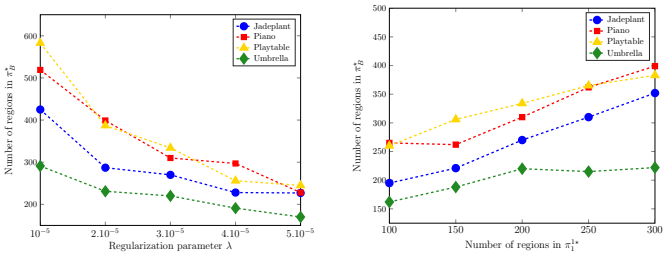

Fig. C.13: Influence of (left) regularization parameter λ (for $|\pi_1^{1\star}| = 200$) and (right) number of regions in partition $\pi_1^{1\star}$ for the braid construction procedure (for $\lambda = 3.10^{-5}$) on the size of the braid optimal cut.

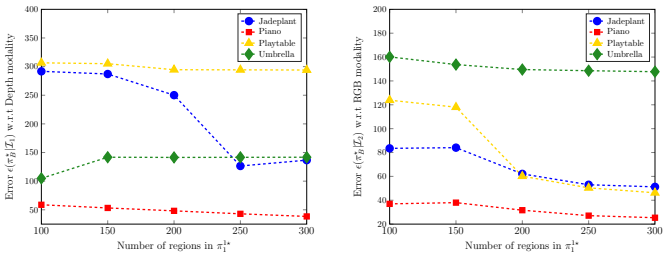

Fig. C.14: Influence of $|\pi_1^{1\star}|$ (for $\lambda = 3.10^{-5}$) on the average GOF for the braid optimal cut with respect to both modalities $\mathcal{I}_1$ (Depth) and $\mathcal{I}_2$ (RGB).

Figure C.13 (left) shows that the number of regions $|\pi_B^\star|$ in the optimal partition $\pi_B^\star$ has again a decreasing behavior with respect to λ. The reason evokated for the HSI/LiDAR data set still holds here: the value of λ controls the trade-off between over- and under-segmentation in the sense that the greater the λ, the fewer the number of regions in the braid optimal cut $\pi_B^\star$. Regarding the influence of $|\pi_1^{1\star}|$ on $|\pi_B^\star|$ (displayed by Figure C.13 (right)), there clearly is an increasing trend for the RGB/Depth (contrarily to the HSI/LiDAR data set where $|\pi_B^\star|$ remained relatively stable with respect to $|\pi_1^{1\star}|$). This implies that, for a fixed value of λ, using a first cut $\pi_1^{1\star}$ with a larger number of regions engenders a braid optimal cut $\pi_B^\star$ with also a larger number of regions. This is plausible since $\pi_1^{1\star}$ acts as some sort of "upper bound" on the two cuts $\pi_2^{1\star}$ and $\pi_2^{2\star}$ that are extracted from $H_2^\star$. Note however that, if $\pi_1^{1\star}$ is indeed an upper bound for $\pi_2^{2\star}$ since $\pi_2^{2\star} \leq \pi_1^{1\star}$, it is strictly speaking not the case for $\pi_2^{1\star}$ since $\pi_2^{1\star} \simeq_h \pi_1^{1\star}$ by construction, hence $\pi_2^{1\star}$ can have less regions than $\pi_1^{1\star}$. We observed in practice that this is however seldom the case, hence the trend reported in figure C.13 (right).

Figure C.14 reports the influence of $|\pi_1^{1\star}|$ on the GOF value $\epsilon(\pi_B^\star|\mathcal{I}_i)$ with respect to the Depth modality $\mathcal{I}_1$ (left) and the RGB modality $\mathcal{I}_2$ (right). The number of regions $\pi_B^\star$ in the braid optimal cut $\pi_B^\star$ varying in the same direction as $\pi_1^{1\star}$, one would expect $\epsilon(\pi_B^\star|\mathcal{I}_i)$ to be decreasing with $\pi_1^{1\star}$ (since a larger number of regions in $\pi_B^\star$ implies a smaller average region size, thus a lower variance within each region). This is indeed the case for $\epsilon(\pi_B^\star|\mathcal{I}_2)$ (figure C.14 (right)), but the decreasing rate seems to greatly depend on the considered image. As a matter of fact, it is divided by a factor 2 for the Playtable scene (decreasing from 120 to roughly 50) while it remains almost unchanged for the Umbrella image (diminishing from 160 to roughly 150). For the depth modality $\mathcal{I}_1$ (figure C.14 (left)) however, the previous two cited scenes have some rather intrigating behaviors: $\epsilon(\pi_B^\star|\mathcal{I}_1)$ remains constant for Playtable, whereas it is even increasing for Umbrella. As for the HSI/LiDAR data set, the proper tuning of $|\pi_1^{1\star}|$ seems to be bounded to the content in the

scene. Thus, λ and $|\pi_1^{1\star}|$ are again respectively set to 3.10^{-5} and 200 in the following.

Appendix C.2. Results

Table C.3 presents the number of regions as well as the average GOF of optimal partitions $\pi_1^\star, \pi_2^\star, \pi_{[23]}^\star$ and $\pi_B^\star$ with respect to both modalities $\mathcal{I}_1$ (the Depth modality) and $\mathcal{I}_2$ (the RGB modality), for all 9 scenes. For each scene, the lowest modality-wise average GOF appears in bold. All listed remarks for the HSI/LiDAR data set are confirmed by the analysis of table C.3. In particular, the marginal segmentation $\pi_i^\star$ almost consistently scores the lowest average GOF with respect to its own modality $\mathcal{I}_i$ again. As a matter of fact, $\pi_1^\star$ is outperformed only once with respect to $\mathcal{I}_1$ (by $\pi_B^\star$ on the Flowers scene). For its part, $\pi_2^\star$ does not rank first only twice with respect to $\mathcal{I}_2$, on the Piano and Vintage images (where $\pi_{[23]}^\star$ and $\pi_B^\star$ achieve the best result overall, respectively). However, the effect noticed for the HSI/LiDAR multimodal scenario occurs again here, being that both marginal approaches perform poorly on the alternate modality. The reason stated for the HSI/LiDAR data set remains valid: each $\pi_i^\star$ is obtained as the minimizer of an energy function acting only on its own modality $\mathcal{I}_i$, thus not accounting for any salient region in the other modality. Conversely, the collaborative BPT and the braid approaches use a multimodal energy. Its minimization forces both the Depth and RGB modalities to collaborate when a region appears salient in one modality but not in the other, thus making full use of the complementary information contained in the multimodal scene. As a consequence, the obtained optimal segmentations $\pi_{[23]}^\star$ and $\pi_B^\star$ achieve some trade-off in terms of fitting accuracy between the Depth and the RGB modalities. This also explains why they tend not to outperform the marginal segmentations on their own modality since both $\pi_{[23]}^\star$ and $\pi_B^\star$ have a total number of regions being the same as (or very close to) the one in $\pi_1^\star$ and $\pi_2^\star$. Note that the bias of $\pi_B^\star$ toward the modality from which is extracted the first cut $\pi_1^{1\star}$ during the braid construction procedure is confirmed here as well, since $\pi_B^\star$ ranks second best 7 times with respect to $\mathcal{I}_1$, but only 3 times with respect to $\mathcal{I}_2$.

Figure C.15 shows the optimal Depth partition $\pi_1^\star$, the optimal RGB partition $\pi_2^\star$, the optimal collaborative partition $\pi_{[23]}^\star$ and the braid optimal partition $\pi_B^\star$ for the Jadeplant, Playroom and Umbrella scenes, represented by their mean Depth and their mean RGB value. It visually supports all the conclusions drawn from the quantitative analysis of Table C.3 as well as those pointed out in the main manuscript for the HSI/LiDAR data set. As a matter of fact, $\pi_1^\star$ accurately captures all salient regions in the depth modality for all three scenes, but mis-segment those with the same depth but not the same RGB color. This can be observed for instance on the Jadeplant image for the blue grid in the foreground which blends with its wooden support and the brown cardboard box in the background which is merged with the green sheet at the same depth. It can also be noticed on the Playroom scene where all drawings in the background wall are completely omitted, or on the Umbrella image where the details on the background wall and the wooden drawers on the right-hand side of the scene suffer from the same issue. The opposite phenomenon inevitably happens for $\pi_2^\star$: all details are well preserved with respect to the RGB modality, at the expense of their depth. This is particularly noticeable on the Jadeplant image, where the branches of the plant appears in a brownish color similar to the one of the cardboard behind, and thus suffer from some sort of "leakage" effect in terms of delineation accuracy. This leakage effect also appears on red shelf of the Playroom scene. The collaborative BPT and the braid structure again produce visually similar results, with nevertheless some slightly more visible differences than int the HSI/LiDAR data set, especially for the Jadeplant and Playroom scenes. In the former case, the braid optimal cut performs notably better to differentiate the branches from the brown background, while the optimal segmentation extracted from the collaborative BPT clearly suffers from the leakage effect. For the latter scene however, the collaborative BPT better retained the details in the drawing hanging on the wall while the braid optimal cut only preserved their overall frame. The major differences between the two methods in the depth images can be spotted in places where a progressive shading of depth occurs. In such case, the gray gradient is better rendered by the segmentation extracted from the braid structure than the one coming from the collaborative BPT, as this latter tends to average the slight variations of gray levels because of the consensus policy adopted during its construction.

Appendix C.3. A note on the computational complexity

We briefly discuss here the computational complexity involved in the definition of the braid structure. We nevertheless recall that obtaining an efficient implementation of the braid structure and conducting a quantitative evaluation of the computational complexity are both beyond the scope of this paper. Given some family of partitions $B = \{\pi_1, \ldots, \pi_n\}$, checking whether B is a braid amounts to verify that all pairwise refinement suprema $\pi_{i,j} = \pi_i \vee \pi_j$ are all h-equivalent to each other (see the proof of proposition 4).

From a practical point of view, obtaining the refinement supremum of two partitions π_i and π_j is strictly equivalent to finding the connected components of the bipartite intersection graph $\mathfrak{G}_{i,j} = (\mathfrak{V}_{i,j}, \mathfrak{E}_{i,j})$. $\mathfrak{G}_{i,j}$ is composed of as many vertices $v_k \in \mathfrak{V}_{i,j}$ as the number of regions both in π_i and π_j and has $e_{k,l} \in \mathfrak{E}_{i,j}$ as an edge linking vertices v_k and v_l if and only if the couple $(\mathcal{R}_k, \mathcal{R}_l) \in \pi_i \times \pi_j$ is such that $\mathcal{R}_i \cap \mathcal{R}_j \neq \emptyset$. The connected components of $\mathfrak{G}_{i,j}$ can be found in a linear time with respect to $\max(|\mathfrak{V}_{i,j}|, |\mathfrak{E}_{i,j}|)$ [43], hence the complexity of computing $\pi_i \vee \pi_j$ is $O(\max(|\pi_i + \pi_j|, |\mathfrak{E}_{i,j}|))$, where $|\mathfrak{E}_{i,j}|$ obviously depends on the structure of the two partitions π_i and π_j. As this must be done for the $N = \binom{n}{2}$ possible pariwise refinement suprema of the family $B = \{\pi_1, \ldots, \pi_n\}$, the overall computational complexity of this first step is $O\left(n^2 \times \max_{1 \leq i < j \leq n} (|\pi_i + \pi_j|, |\mathfrak{E}_{i,j}|)\right)$.

The second step is to check that all pairwise refinement suprema $\pi_{i,j}$ are h-equivalent to each other. In a positive scenario, then B has a braid structure. For now, this verification is done by checking in a brute-force manner that each region in π_{i_0,j_0} is either disjoint or nested with all the other regions in the remaining $N-1$ refinement suprema $\pi_{i \neq i_0, j \neq j_0}$. Thus, denoting by α_N the highest number of regions among the N refinement suprema $\pi_{i,j}$, *i.e.*, $\alpha_N = \max_{1 \leq i < j \leq n} |\pi_{i,j}|$, the complexity of the verification step is $O(N \times \alpha_N)$, which depends both on the initial number of partitions in the family B, as well as the structures of these partitions (impacting the value of α_N).

If the family B has indeed a braid structure, the last step is the computation of the associated monitor hierarchy H_m. This can be efficiently done by first computing the overall infimum $\pi_0^{H_m} = \bigwedge_{1 \leq i < j \leq n} \pi_{i,j}$ of the pairwise refinement suprema $\pi_{i,j}$, that plays the role of the leaf partition of the monitor hierarchy H_m. Then, the inclusion relationships among the various regions of the different refinement suprema $\pi_{i,j}$ can be deduced from the number and the identity of the leaf regions of $\pi_0^{H_m}$ they contain, allowing to create the monitor hierarchy H_m.

In the conducted experiments, we limited the number of partitions in the family B to 4 (and hence, the number of refinement suprema to compute to 6) to guarantee the braid structure following the proposed construction methodology. We noted that the main factor impacting the computational time was related to the number of regions in the partitions composing the braid structure B. All experiments were conducted on a 2.40 GHz Intel Core i7-4700HQ with Matlab R2015a, leading to a computational time for the braid/monitor hierarchy construction and processing being no longer than 70 seconds for the HSI/LiDAR data

Table C.3: Size (top) and GOF with respect to $\mathcal{I}_1$ (bottom left) and $\mathcal{I}_2$ (bottom right) of optimal partitions with respect to the 9 RGB/Depth scenes, and their overall average rank (last column). For each scene, the lowest modality-wise GOF is in bold.

	Flowers	Jadeplant	Motorcycle	Piano	Playroom	Playtable	Recycle	Umbrella	Vintage	avg rank
$\pi_1^\star$	275 15.5 | 182	270 **145** | 120	751 **228** | 131	311 **38.5** | 48.5	319 **52.9** | 112	337 **283** | 47.9	279 **194** | 48.6	219 **102** | 159	365 **46.3** | 41.0	**1.11** | 3.78
$\pi_2^\star$	274 403 | **58.0**	272 654 | **34.8**	752 240 | **48.4**	310 1861 | 26.1	317 2427 | **38.2**	334 356 | **47.6**	275 295 | **29.1**	219 133 | **142**	364 544 | 62.2	3.89 | **1.44**
$\pi_{[23]}^\star$	273 98.8 | 65.2	265 572 | 68.6	752 240 | 61.6	312 56.9 | **16.1**	324 188 | 56.1	334 303 | 71.8	281 272 | 35.1	219 151 | 152	370 106 | 47.1	3.11 | 2.67
$\pi_B^\star$	275 **13.6** | 63.0	270 250 | 62.1	751 231 | 73.7	310 48.2 | 31.7	319 75.9 | 86.9	334 295 | 60.2	277 199 | 37.9	220 141 | 150	364 48.6 | **38.0**	1.89 | 2.44

set (for crop #1 with $|\pi_1^{1\star}| = 300$) and 180 seconds for the RGB/Depth data set (for the Playroom scene, again with $|\pi_1^{1\star}| = 300$). Note that better performances could be obtained following a code optimization pass, but this is beyond the scope of the present paper.

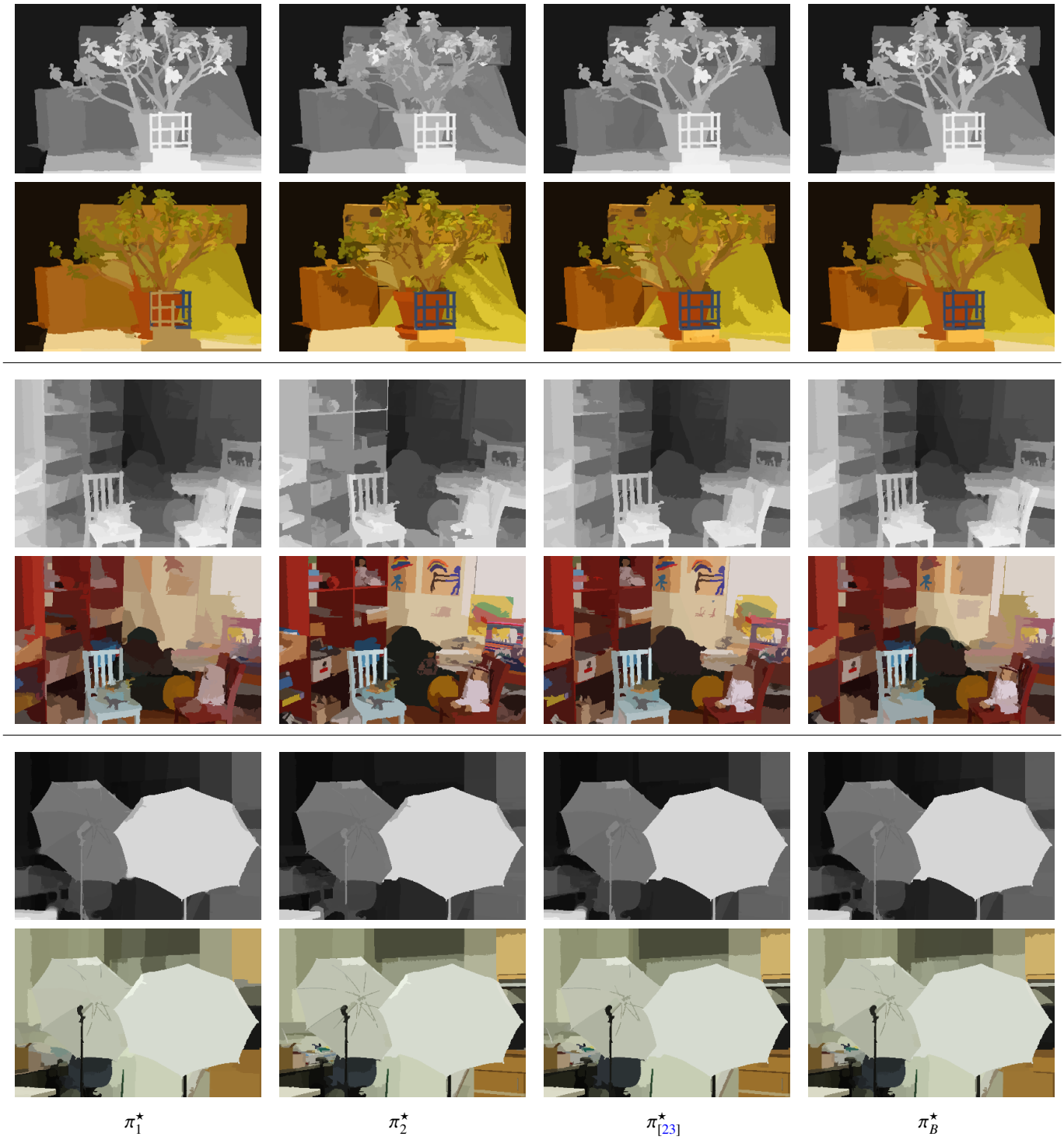

Fig. C.15: Optimal partitions $\pi_1^\star$, $\pi_2^\star$, $\pi_{[23]}^\star$ and $\pi_B^\star$ for images Jadeplant (top two rows), Playroom (middle two rows) and Umbrella (bottom two rows) represented by their mean depth and mean RGB value.