Braids of partitions for the hierarchical analysis of multimodal images

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

[a] *GIPSA-Lab, 11 rue des Mathématiques, 38400 Saint Martin d'Hères, France.*
[b] *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, but it has never been investigated in practical scenarios. In this paper, we propose a novel methodology for the analysis of multimodal images, based on this notion of braids of partitions. In particular, we develop a method to perform the hierarchical segmentation of such multimodal images, relying on an energetic minimization framework. The proposed approach is investigated on various multimodal images scenarios, and the obtained results confirm its ability to efficiently handle the multimodal information to produce more accurate segmentation outputs.
Keywords: Hierarchical representation, multimodal image, braid of partitions, energy minimization, image segmentation

---

Email address: `guillaume.tochon@grenoble-inp.fr` (Guillaume Tochon)

1. Introduction

Multimodality is nowadays increasingly used in signal and image processing. Multimodal signals ensure a more complete and accurate representation of the recorded source, as they jointly consider several single acquisitions (called *modalities*) of this source, each one being acquired with different acquisition configuration (such as different sensor types or different acquisition dates) [1]. In image processing for instance, multimodal images are frequently encountered in medical imaging [2] or remote sensing [3]. However, jointly processing the redundant and complementary information featured by the various multimodalities in a generic manner is an arduous task, as it greatly depends both on the nature of the handled multimodality as well as the underlying application. For this reason, the generic design of multimodal image analysis frameworks remains a real challenge.

Besides, the analysis of a given image is bound to the notion of *scale of analysis*, *i.e.*, the definition of an appropriate level of details to retrieve, within this image, the features of interest with respect to the pursued goal. This concept of scale of analysis obviously depends on the context and the application, as features of various complexities can be extracted from a single image. Thus, hierarchical representations (HRs) were proposed as a possible solution to the intrinsic multiscale nature of images, as they organize in their structures all the potential scales of interest of the image in a nested way. While the construction of a HR should depend only on the specificities of the image, its further analysis is bound to the underlying application. Popular HRs include quad-trees [4], component trees [5], inclusion trees [6], as well as α-trees [7] and binary partition trees [8]. Such structures are now widely used for several image processing and computer vision tasks such as object detection [9] or image segmentation [10]. A review on the use of HRs in the field of mathematical morphology can be found in [11].

While the framework of hierarchical image analysis has been successfully investigated on a wide range of applications, its extension to multimodal images is challenging as the optimal representation of multimodal images by hierarchical structures remains an open question. Recently, the concept of braids of partitions [12] has been introduced as a potential tool to tackle this issue. Here, we define a complete methodology for the hierarchical analysis of multimodal images, based on this concept of braids of partitions. Therefore, this paper brings the following contributions:

- There is up to now no clear guidelines to construct a braid of partitions. We remedy to this point by providing a fully operable way to build the braid structure from two independant HRs.

- We demonstrate the relevance of the obtained braid structure by using it for the hierarchical segmentation of multimodal images, following an energy minimization scheme. This 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.

The remainder of this paper is organized as follows: section 2 reviews the works related to data fusion strategies for multimodal image segmentation, as well as hierachical energy minimization techniques. Section 3 recalls various definitions and properties related to HRs and hierarchical energy minimization procedures. Section 4 presents the concept of braids of partitions proposed by [12] and extends the classical energetic framework on these particular structures. Section 5 details the main contributions of this paper as stated above, while section 6 shows the application of this methodology on two[1] different multimodal datasets and discusses the obtained results. Conclusion and future work are drawn in Section 7.

2. Related work

2.1. Multimodal image segmentation

Image segmentation is a particular application that would surely benefit from the development of multimodal processing tools. As a matter of fact, an optimal use of the complementary and redundant information contained within the multimodal images should lead to a more robust and accurate delineation of the regions composing the segmentation map, in particular when those regions share similar features in one modality but not in the other ones. The use of this information can be integrated at two different stages of the processing chain when performing multimodal image segmentation, namely the *feature* or the *decision* level [14]. In the former case, features are extracted independently from each modality, and further combined in order to produce some unified feature map and a fused image from which the final multimodal segmentation is derived. This fusion strategy is notably investigated in [15] where the various modalities are decomposed following some multiresolution (MR) transformation. Those are all further merged to create a single combined MR, which is in turn inverted to retrieve the fused

[1] Experiments and results on two supplementary additional multimodal data sets are available at https://webfiles.ampere.grenoble-inp.fr/f3lgrj.

image on which classical segmentation algorithms can be applied. Similar ideas are for instance investigated in [16] using independant component analysis coefficients, or in [17] with discrete cosine transform coefficients. Note in addition that co-segmentation [18] (*i.e.*, the extraction of a foreground region from the background using pairs of images) can also be seen as a segmentation procedure with information fusion at the feature level, but the handled pairs of images are not, strictly speaking, multimodal images. In the scenario of a fusion at the decision level on the other hand, each modality is respectively processed to output an individual segmentation map. Those are later on combined in order to produce the final multimodal segmentation map. Several solutions have been proposed to merge several segmentation maps, ranging from geometrical interpolation [19] to homogeneity graph segmentation by random walker [20], flexible couplings [21] or ensemble clustering [22].

Similarly, the integration of the additional information brought by the multimodality can occur either at the feature level or the decision level when dealing with HRs. For the former scenario, which aims at building a single HR that directly encompasses all the specificities of the various modalities, the only existing work is, to the best of our knowledge, the one presented in [23]. In this case, deriving a final multimodal segmentation is eased since it allows to apply classical tools to extract a segmentation from a hierarchy. On the other hand, all the modalities need to cooperate during the construction of the HR, and some features may be "averaged out" by the consensus strategies adopted during the construction. Contrarily, performing the fusion at the decision level implies implies to build one HR per modality, to further combine them all. In that approach, each hierarchy can capture all the specificities of its own modality, but the fusion decision may become complicated due to the increased number of disagreements that could occur between the hierarchies, which is the reason why this strategy remains an open question. Braids of partitions [12] were recently introduced to address this point, and we sketched in [13] how they could be adapted in practice to achieve the hierarchical segmentation of multimodal images.

2.2. Segmentation by hierarchical energy minimization

Image segmentation is an ill-posed problem in itself since a given image can often be properly segmented at various levels of detail, and the precise level to choose depends on the underlying application. Thus, hierarchical representations are well suited for segmentation purposes, as they allow to achieve, with a single structure, For image segmentation purposes, the level of exploration of the hierarchical structure is tuned to extract from the whole set of achievable segmentations the one that best matches the desired goal [24]. This notion of optimality with respect to a task commonly relies on the definition of some objective function (also called cost, or *energy* function) which is minimized over the set of possible outcomes to find the best one. This idea has been for instance investigated in [8] over a BPT representation in a context of rate/distortion optimization, or in [25] to perform image segmentation using a tree of shapes, based on the minimization of the Mumford-Shah functional [26]. Conditions on the energy formulation under which the minimization procedure can be solved were formally studied in [27, 28] for particular energy functions and later generalized in [29] to wider classes of energies. Those conditions are briefly reviewed in the following section 3.2.

3. Hierarchies of partitions

3.1. Hierarchies of partitions

Let $\mathcal{I} : E \to V$, $E \subseteq \mathbb{Z}^2$, $V \subseteq \mathbb{R}^n$, be a generic image, of elements (pixels) $\mathbf{x}_i \in E$ and of pixel values $\mathcal{I}(\mathbf{x}_i)$. E thus denotes the *support* space of image $\mathcal{I}$, while V stands for the space of all its pixel values. Following this definition, a $P-$multimodal image $\mathcal{I}_P$ is characterized by the joint composition of its P modalities $\{\mathcal{I}_1, \ldots, \mathcal{I}_P\}$, with $\mathcal{I}_i : E_i \to V_i$, $i = 1, \ldots P$. 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* of E, denoted π, 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. The *refinement ordering* $\leq$ is a binary order on Π_E defined as follows: for any two partitions $\pi_i, \pi_j \in \Pi_E$, $\pi_i \leq \pi_j$ when each region $\mathcal{R}_i \in \pi_i$ is included in a region $\mathcal{R}_j \in \pi_j$. In such case, π_i is said to refine (or to be a refinement of) π_j. While the refinement ordering is only a partial order, any two partitions π_i and π_j admits a unique *infimum* $\pi_i \wedge \pi_j$ and *supremum* $\pi_i \vee \pi_j$. The former is the largest partition for the refinement ordering that refines both π_i and π_j, and is obtained by taking the intersection of all the regions of π_i and π_j. Conversely, the refinement supremum $\pi_i \vee \pi_j$ is the smallest partition that is refined both by π_i and π_j, and can be viewed as the partition obtained when keeping only the regions with shared boundaries.

A hierarchy of partitions H of a space E is defined as a finite sequence of partitions ordered by refinement:

$$H = \{\pi_i\}_{i=0}^{n} \text{ such that } i \leq j \Rightarrow \pi_i \leq \pi_j \tag{1}$$

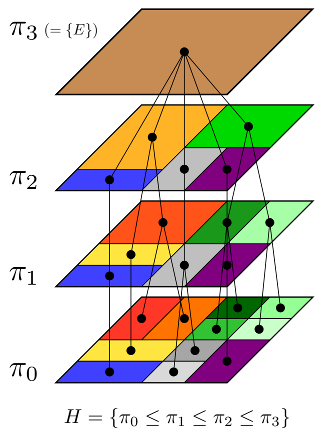
The partitions in the sequence are ranging from the *leaf partition* π_0 to the *root partition* $\pi_n = \{E\}$ of the hierarchy. Equivalently, a hierarchy of partitions can be defined as a collection of regions $H = \{\mathcal{R} \subseteq E\}$ such that $\emptyset \notin H, E \in H$ and $\forall \mathcal{R}_i, \mathcal{R}_j \in H, \mathcal{R}_i \cap \mathcal{R}_j \in \{\emptyset, \mathcal{R}_i, \mathcal{R}_j\}$, meaning that any two regions belonging to a hierarchy are either disjoint or nested. A hierarchy of partitions is often represented as a tree graph, 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.

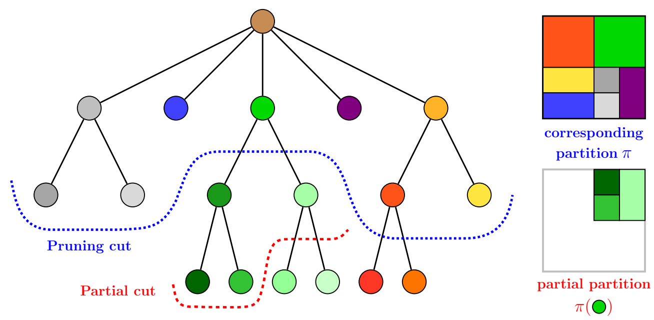
A (pruning) *cut* of H is a partition π of E whose regions belong to H, and $\Pi_E(H)$ denotes the set of all such cuts. $H(\mathcal{R})$ stands for the sub-hierarchy of H rooted at $\mathcal{R}$. Any cut of the sub-hierarchy $H(\mathcal{R})$ is called a *partial partition* of $\mathcal{R}$ following [30], and is denoted $\pi(\mathcal{R})$. All presented notions related to hierarchies of partitions are depicted by figure 1.

3.2. Hierarchical energy minimization

In the following, an energy function will be simply modeled as a mapping $\mathcal{E} : \Pi_E \to \mathbb{R}^+$ that associates to each partition $\pi \in \Pi_E$ (and *a fortiori*, to each cut in $\Pi_E(H)$) a real non-negative 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:

$$\mathcal{E}(\pi) = \underset{\mathcal{R}_i \in \pi}{\mathfrak{D}} \mathcal{E}(\mathcal{R}_i), \tag{2}$$





H	Hierarchy of partitions.
$\mathcal{R} \in H$	Region (subset of E) in H.
$\pi \in \Pi_E(H)$	cut of hierarchy H.
$\pi(\mathcal{R})$	partial partition of $\mathcal{R} \in H$.

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

where $\mathfrak{D}$ is some composition rule to explicit the relationship between the energy of the partition π and those of its regions $\mathcal{R}_i \in \pi$. For instance, the sum (*i.e.* $\mathcal{E}(\pi) = \sum_{\mathcal{R}_i \in \pi} \mathcal{E}(\mathcal{R}_i)$) is generally used in many classical energy functions defined in the literature, such as the the Mumford-Shah functional [26], graph cuts [31] or Markov random fields [32]. 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 $\mathcal{E}$ is defined as:

$$\pi^\star = \operatorname*{argmin}_{\pi \in \Pi_E(H)} \mathcal{E}(\pi). \tag{3}$$

While it is impossible to evaluate the cardinality of $\Pi_E(H)$ since it strongly depends on the structure of H, finding the optimal cut $\pi^\star$ by an exhaustive search is highly unrealistic 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 [27] in the context of separable energies (*i.e.*, $\mathfrak{D} \equiv \sum$) and later on generalized in [29] to wider classes of composition rules $\mathfrak{D}$, 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:

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

$$\pi^\star(\mathcal{R}) = \operatorname{argmin} \left\{ \mathcal{E}(\mathcal{R}), \mathcal{E}\left(\bigsqcup_{r \in \mathsf{S}(\mathcal{R})} \pi^\star(r) \right) \right\} \tag{5}$$

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 [27]. It was shown in particular in [12] that all energies which can be expressed as a Minkowski expression:

$$\mathcal{E}(\pi) = \left(\sum_{\mathcal{R} \in \pi} \mathcal{E}(\mathcal{R})^\alpha \right)^{\frac{1}{\alpha}} \tag{6}$$

are h-increasing for every $\alpha \in [-\infty, +\infty]$, generalizing previously obtained results for energies composed by the sum ($\alpha = 1$) [27, 8], the supremum ($\alpha = +\infty$) [28] and the infimum ($\alpha = -\infty$) [33], notably. Thus, the optimal cut of a hierarchy for any type of Minkowski composed energy function can be easily retrieved following equations (4) and (5).

Energies in the literature often depend in practice on a positive real-valued parameter λ that acts as a trade-off between simplicity (*i.e.*, favoring under-segmentation) and a good data fitting of the segmentation (*i.e.*, leading to

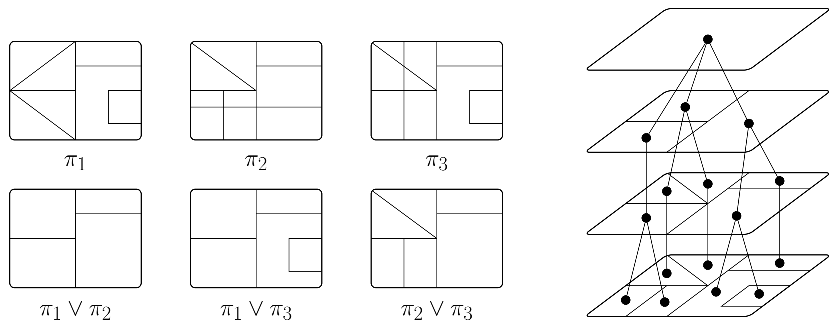


Figure 2: Example of braid of partitions $B = \{\pi_1, \pi_2, \pi_3\}$. On the right is a monitor hierarchy of B since the pairwise refinement suprema $\pi_i \vee \pi_j, i, j \in \{1, 2, 3\}, i \neq j$ define cuts of this hierarchy different from the whole space E.

over-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 [29] 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

$$\lambda_1 \leq \lambda_2 \Rightarrow \pi^\star_{\lambda_1} \leq \pi^\star_{\lambda_2}. \tag{7}$$

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 [27] for more practical implementation details.

4. Braids of partitions

4.1. Definition of a braid

The analysis of a multimodal image by means of a HR 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 [12] as a potential tool to combine multiple hierarchies and thus precisely answer this question [13].

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:*

$$\forall\, \pi_i, \pi_j \in B, \pi_i \vee \pi_{j\neq i} \in \Pi_E(H_m)\backslash\{E\} \tag{8}$$

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 (8). Otherwise, any family of arbitrary partitions would form a braid with $\{E\}$ as a supremum, thus loosing any interesting structure. An example of braid of partitions 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. Therefore, the couple B/H_m can be viewed as a hierarchical representation of the multimodal image that relies both on the complementary and redundant information contained in the data.

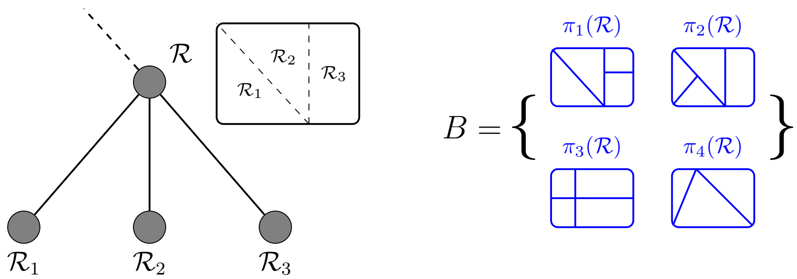
4.2. Minimizing an energy function over a braid

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 [12] that the dynamic program structure holding on hierarchies (equations (4) and (5)) 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:

$$\mathcal{E}^\star(\mathcal{R}) = \min\left\{\mathcal{E}(\mathcal{R}), \mathcal{E}\left(\bigsqcup_{r\in\mathsf{S}(\mathcal{R})}\pi^\star(r)\right), \bigwedge_{\pi_i(\mathcal{R})\in B}\mathcal{E}(\pi_i(\mathcal{R}))\right\} \tag{9}$$

$$\pi^\star(\mathcal{R}) = \begin{cases} \{\mathcal{R}\} & \text{if } \mathcal{E}^\star(\mathcal{R}) = \mathcal{E}(\mathcal{R}) \\ \displaystyle\bigsqcup_{r\in\mathsf{S}(\mathcal{R})}\pi^\star(r) & \text{if } \mathcal{E}^\star(\mathcal{R}) = \mathcal{E}\left(\bigsqcup_{r\in\mathsf{S}(\mathcal{R})}\pi^\star(r)\right) \\ \displaystyle\operatorname*{argmin}_{\pi_i(\mathcal{R})\in B}\mathcal{E}(\pi_i(\mathcal{R})) & \text{otherwise.} \end{cases} \tag{10}$$

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 its 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 $\mathcal{R}$).



$$\mathcal{E}^\star(\mathcal{R}) = \min\left\{\mathcal{E}(\mathcal{R}), \mathcal{E}\Big(\bigsqcup_{i=1,2,3} \pi^\star(\mathcal{R}_i)\Big), \bigwedge_{\pi_i(\mathcal{R})\in B} \mathcal{E}(\pi_i(\mathcal{R}))\right\}$$

Figure 3: A step of the dynamic program (9) applied to a braid structure: one has to choose between $\{\mathcal{R}\}$, $\bigsqcup \pi^\star(\mathcal{R}_i)$ or any other $\pi_i(\mathcal{R}) \in B$. Note however that $\mathcal{R} \neq E$, otherwise B would not be a braid since $\pi_3(\mathcal{R}) \vee \pi_4(\mathcal{R}) = \mathcal{R}$.

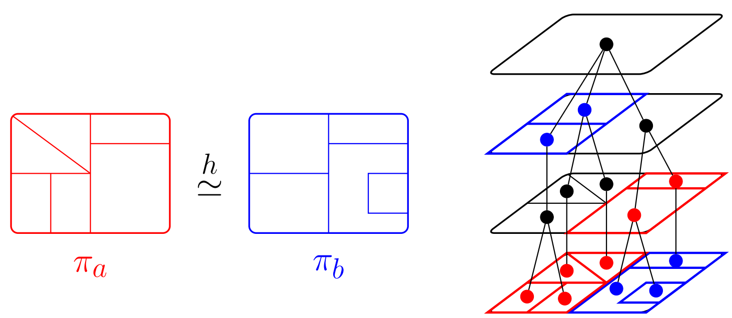


Figure 4: Illustration of the h-equivalence relation: π_a and π_b are h-equivalent (left), they define two different cuts of the same hierarchy (right).

5. Proposed hierarchical analysis of multimodal images with braids

5.1. Generating a braid from multiple hierarchies

As pointed out in [12], 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 (8) is fulfilled), and the generation of general 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. More particularly, all those cuts must be sufficiently related to ensure that all their pairwise refinement suprema are hierarchically organized. 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 $\pi_a \overset{h}{\simeq} \pi_b$ if and only if*

$$\forall \mathcal{R}_a \in \pi_a,\, \forall \mathcal{R}_b \in \pi_b, \mathcal{R}_a \cap \mathcal{R}_b \in \{\emptyset, \mathcal{R}_a, \mathcal{R}_b\}. \tag{11}$$

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. Evidently, if two partitions are globally comparable, they are locally comparable as well. All cuts of a hierarchy H are h-equivalent: $\forall \pi_1, \pi_2 \in \Pi_E(H),\ \pi_1 \overset{h}{\simeq} \pi_2$. Conversely, if two partitions are h-equivalent, they define two cuts of the same hierarchy.

Given some hierarchy H and a partition $\pi_* \in \Pi_E$, we denote by $H \overset{h}{\simeq} \pi_*$ the set of cuts of H that are h-equivalent to π_*: $H \overset{h}{\simeq} \pi_* \subseteq \Pi_E(H)$ 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 equipped with this h-equivalence relation, let $B = \{\pi_i \in \Pi_E\}$ be a braid, and H_m be a monitor hierarchy of it.

Proposition 1. *If there exists $\pi_i, \pi_j \in B$ 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 (8) of a braid, $\pi_i \vee \pi_j \in \Pi_E(H_m)$, so $\pi_j \in \Pi_E(H_m)$. □

Proposition 2. *If there exists $\pi_i, \pi_j, \pi_k, \pi_l \in B$ such that $\pi_i \leq \pi_j$ and $\pi_k \leq \pi_l$, then $\pi_j \overset{h}{\simeq} \pi_l$.*

Proof. Using proposition (1) for both $\pi_j \leq \pi_j$ and $\pi_k \leq \pi_l$, it follows that $\pi_j, \pi_l \in \Pi_E(H_m)$. Thus $\pi_j \overset{h}{\simeq} \pi_l$ using the property of h-equivalence. □

Proposition (2) has an important consequence in practice: if one wants to compose a braid using two ordered cuts $\pi_i^1, \pi_i^2 \in \Pi_E(H_i), \pi_i^1 \geq \pi_i^2$ coming from two different hierarchies $H_i, i \in \{1,2\}$, then for $B = \{\pi_i^j\}, (i,j) \in \{1,2\} \times \{1,2\}$ to be a braid, it is necessary that $\pi_1^1 \overset{h}{\simeq} \pi_2^1$. Following this, we propose to build a braid using 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 $H_2 \overset{h}{\simeq} \pi_1^1 \backslash \{E\}$, 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 5.

Proposition 3. *Under this configuration, $B = \{\pi_i^j\}, (i,j) \in \{1,2\} \times \{1,2\}$ has a braid structure.*

Proof. The proof is provided as a supplementary material. □

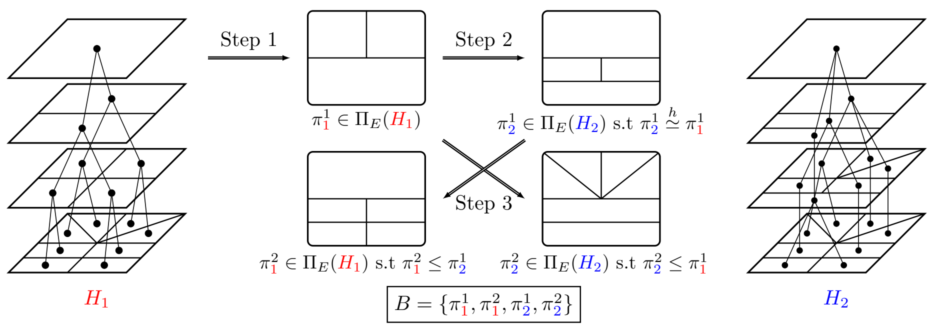


Figure 5: Composing a braid B with cuts from two hierarchies H_1 and H_2.

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 independent 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 independent hierarchies.

5.2. Braid-based multimodal image segmentation

From a conceptual point of view, conducting the energy minimization procedure described in section 4.2 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 a smaller scale, coming from one modality and that fit all the modalities at

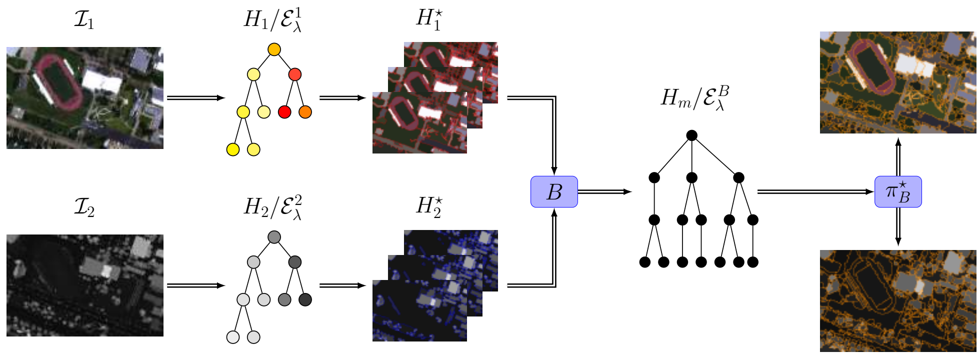


Figure 6: Proposed braid-based multimodal segmentation methodology.

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 $\mathcal{E}^1_\lambda$ and $\mathcal{E}^2_\lambda$ 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 [26]:

$$\mathcal{E}^i_\lambda(\pi) = \sum_{R\in\pi}\left(\Xi_i(\mathcal{R}) + \frac{\lambda}{2}|\partial\mathcal{R}|\right) \tag{12}$$

where

$$\Xi_i(\mathcal{R}) = \int_{\mathcal{R}} \|\mathcal{I}_i(\mathbf{x}) - \boldsymbol{\mu}_i(\mathcal{R})\|_2^2 dx \tag{13}$$

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 [29], is a popular choice when it comes to minimizing some energy function because of its ability to produce consistent segmentations [34].

The braid B is then composed following the procedure previously described: a partition $\pi_1^{1\star}$ is first extracted from $H_1^\star$. Then, two partitions $\pi_2^{1\star}$ and $\pi_2^{2\star}$ are selected in $H_2^\star$ such that $\pi_2^{1\star} \overset{h}{\simeq} \pi_1^{1\star}$ and $\pi_2^{2\star} \leq \pi_1^{1\star}$. The last partition $\pi_1^{2\star}$ is finally chosen in H_1 such that $\pi_1^{2\star} \leq \pi_2^{1\star}$. In practice, the sets $H_2^\star \overset{h}{\simeq} \pi_1^{1\star}$, $H_2^\star \leq \pi_1^{1\star}$ and $H_1^\star \leq \pi_2^{1\star}$ may contain several cuts. We propose to define $\pi_2^{1\star}, \pi_1^{2\star}$ and $\pi_2^{2\star}$ as the largest cut of their respective sets, namely $\pi_2^{1\star} = \bigvee\{H_2^\star \overset{h}{\simeq} \pi_1^{1\star} \backslash \{E\}\}$, $\pi_1^{2\star} = \bigvee\{H_1^\star \leq \pi_2^{1\star}\}$ and $\pi_2^{2\star} = \bigvee\{H_2^\star \leq \pi_1^{1\star}\}$.

Eventually, B is composed of 4 partitions $\{\pi_1^{1\star}, \pi_1^{2\star}, \pi_2^{1\star}, \pi_2^{2\star}\}$ extracted from the two hierarchies $H_1^\star$ and $H_2^\star$, and the braid structure is guaranteed, allowing to construct the monitor hierarchy H_m. A last energy term $\mathcal{E}_\lambda^B$ is defined as a multimodal piece-wise constant Mumford-Shah energy, relying on both modalities of the multimodal image $\mathcal{I}$:

$$\mathcal{E}_\lambda^B(\pi) = \sum_{\mathcal{R}\in\pi} \left(\max\left(\frac{\Xi_1(\mathcal{R})}{\Xi_1(\mathcal{I}_1)}, \frac{\Xi_2(\mathcal{R})}{\Xi_2(\mathcal{I}_2)} \right) + \frac{\lambda}{2}|\partial\mathcal{R}| \right) \tag{14}$$

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. Its minimization over H_m and B following the dynamic program (9) and (10) 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 $\mathcal{I}_2$.

5.3. 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 [35]. 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 by visually comparing the multimodal segmentation result against the marginal segmentation outputs (when each modality is processed individually) [23]. To that extend, we propose here to evaluate the ability of the 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 supplementary materials). This allows a fair visual comparison since all four partitions $\pi^\star_B, \pi^\star_1$, $\pi^\star_2$ and $\pi^\star_{[23]}$ 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 $\pi^\star_B, \pi^\star_1$, $\pi^\star_2$ and $\pi^\star_{[23]}$ with respect to both modalities $\mathcal{I}_1$ and $\mathcal{I}_2$ as follows:

$$\epsilon(\pi|\mathcal{I}_i) = \frac{1}{|E|}\sum_{\mathcal{R}\in\pi} |\mathcal{R}| \times \Xi_i(\mathcal{R}) \tag{15}$$

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, for each modality $\mathcal{I}_i$, with its optimal marginal segmentation $\pi^\star_i$.

6. Experimental validation

In the following, we apply the proposed methodology on two different multimodal data sets, each being composed of two modalities. Two additional multimodal data sets are also presented as a supplementary materials. Let us stress again that the goal of this section is not to conduct an exhaustive validation of the proposed method over several images sharing the same multimodality, but rather to demonstrate is adaptability by investigating different multimodal scenarios with their respective specificities.

6.1. Hyperspectral/LiDAR data set

6.1.1. Description of the data set

The Hyperspectral/LiDAR multimodal data set, described in [36], is composed of a $120 \times 185 \times 144$ hyperspectral (HS) image and a LiDAR-derived

Figure 7: RGB composition of the HS image (left) and corresponding LiDAR image (right).

digital surface model (figure 7), 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 an urban area with several houses and buildings of various heights and roofs made of different materials, an athletics stadium with a running track and two stands some parking lots, walkways, roads and some portions of grass and trees. The HS image depicts the spectral reflectance of the scene, *i.e.*, the way the ground has interacted with the incident light. Since each material has an intrinsic reflectance spectrum, HS images are widely used to identify the different materials composing the scene. The LiDAR image provides a topological map of the scene, therefore giving information about the structure or physical shape of the objects composing it. The HS/LiDAR complementarity lies in the fact that both modalities convey some information of totally different nature. Therefore, two neighboring objects of interest can either be constituted of the same materials, but at different heights, or on the other hand, they can share the same height while not being made of the same materials. Therefore, the integration of the complementary information within the braid framework is expected to resolve those potential errors in the optimal marginal segmentations.

6.1.2. 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 propose to work in practice with the binary partition tree (BPT), which has already proved to be very efficient for hierarchical image representation and segmentation [8, 37, 28]. 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 $\mathcal{O}\left(\mathcal{R}_i, \mathcal{R}_j\right)$. Here, we use the mean spectrum and spectral angle for the region model and merging criterion of the HS modality, and the mean value and Euclidean distance for the LiDAR modality. Those 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 [38] conducted on each modality independently.

Constructing the braid B by following the procedure exposed in figure 5 raises the question of which hierarchy the first cut should be extracted from. 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, the first cut is extracted from the BPT built on the LiDAR modality, since it contains less fine details than the HS modality. This first cut, $\pi_1^{1\star}$ contains 150 regions, which roughly corresponds to the number of expected large salient regions in the LiDAR. It is used to extract $\pi_2^{1\star}$ and $\pi_2^{2\star}$ from $H_2^\star$, which

Table 1: Number of regions $|\pi|$ and average GOF $\epsilon(\pi|\mathcal{I}_i)$ of optimal partitions $\pi_1^\star, \pi_2^\star, \pi_{[23]}^\star, \pi_B^\star$ for the Hyperspectral/LiDAR data set, with respect to both modalities $\mathcal{I}_1$ (LiDAR image) and $\mathcal{I}_2$ (HS image). Lowest values are in bold.

	$\pi_1^\star$	$\pi_2^\star$	$\pi_{[23]}^\star$	$\pi_B^\star$
$\vert\pi\vert$	325	325	325	325
$\epsilon(\pi\vert\mathcal{I}_1)$	1224.8	3884.8	1516.0	**994.9**
$\epsilon(\pi\vert\mathcal{I}_2)$	145.4	52.5	73.2	**48.6**

comprise 406 and 414 regions, respectively. Finally, $\pi_1^{2\star}$ is extracted from $H_1^\star$ using $\pi_2^{1\star}$ and contains 495 regions. The four partitions composing B generate $\binom{4}{2} = 6$ cuts of the monitor hierarchy H_m, which is built by re-organizing those cuts in a hierarchical manner. Finally, the minimization of $\mathcal{E}_\lambda^B$ over H_m, following (9), is conducted with λ being empirically set to 5.10^{-5}, and produces an optimal segmentation $\pi_B^\star$ of the braid composed of 325 regions. 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 (14) to produce the same, or a similar, number of regions than contained in $\pi_B^\star$.

6.1.3. 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). Its demonstrates the effectiveness of the braid structure to make the most of the complementary and redundant information contained within the multimodal data set. As expected, $\pi_1^\star$ and $\pi_2^\star$ achieve a low average GOF value with respect to their corresponding modality, but a greater average GOF with respect to the complementary modality. On the other hand, $\pi_{[23]}^\star$ scores a slightly higher average GOF value with respect to each modality than the marginal approaches. This can be explained by the consensus policy during the construction that favors regions that averagedly fit both modalities at the same time rather than regions that match one modality while misfitting the other. Finally, $\pi_B^\star$ outperforms both $\pi_1^\star$, $\pi_2^\star$ and $\pi_{[23]}^\star$ with respect to $\mathcal{I}_1$ and $\mathcal{I}_2$, while it contains the same number of regions. Thus, $\pi_B^\star$ is able to better fit both modalities at the same time, meaning that the braid structure was able to delineate with a greater precision the salient regions of $\mathcal{I}_1$ and $\mathcal{I}_2$. While this result may seem counter-intuitive, it is an illustration of the principle that *the whole is better than the sum of its parts*: 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, which are both well exploited by the proposed braid-based framework.

Figure 8 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 $\mathcal{I}_1$ (top row) and their mean RGB value with respect to $\mathcal{I}_2$ (bottom row). Close-up views are also provided as supplementary materials. The qualitative analysis of figure 8 leads to similar conclusions. While $\pi_1^\star$ correctly fits $\mathcal{I}_1$ by accurately segmenting all notable regions of the LiDAR modality (the various buildings, the trees or the houses located on the bottom left corner of the image) it fails

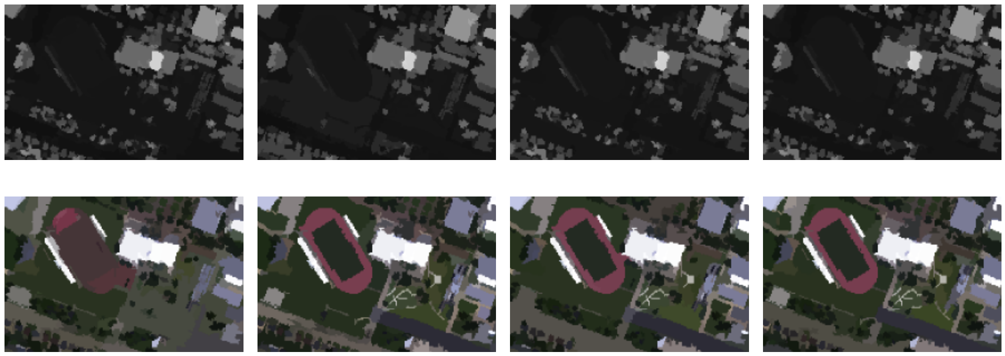


Figure 8: From left to right: optimal partitions $\pi_1^\star$, $\pi_2^\star$, $\pi_{[23]}^\star$ and $\pi_B^\star$ represented with their mean height with respect to the LiDAR modality $\mathcal{I}_1$ (top row) and mean RGB value with respect to the HS modality $\mathcal{I}_2$ (bottom row).

at segmenting regions with similar height but not made of the same materials, such as the running track and the football pitch or the lawns and roads. The reason is straightforward: partition $\pi_1^\star$ is extracted from the hierarchy built on height considerations only and thus cannot account for spectrally different regions, provided that they have the same height. Contrarily, $\pi_2^\star$ conforms $\mathcal{I}_2$ since all spectrally salient regions are well preserved. Regions that have close spectral signatures but not the same height are however generally mis-segmented in $\pi_2^\star$. In particular, several batches of trees are either grouped together, or fused with the neighboring grass (whose spectral response is rather close). Note that for the latter case, despite grass and trees having a close spectral signature, they belong to different semantic classes. The collaborative approach produces an optimal segmentation map $\pi_{[23]}^\star$ that overall well fits both modalities, while several small details have been averaged out, such as small trees or the walkways in the lawns. On the other hand, most erroneous regions of $\pi_1^\star, \pi_2^\star$ or $\pi_{[23]}^\star$ are correctly delineated in the braid

Figure 9: RGB modality (left) and corresponding depth map modality (right).

optimal cut $\pi_B^\star$. That is notably the case for the running track, the lawns and the roads (with respect to $\pi_1^\star$) or the batches of trees (with respect to $\pi_2^\star$).

6.2. RGB/depth data set

6.2.1. Description of the data set

The second considered multimodal data set originates from the Middlebury Stereo Dataset [39] and is composed of an optical image and its associated depth map. Both modalities are composed of 252×370 pixels. The complementarity between the optical and the depth map is expected to benefit the accurate delineation of regions sharing the same properties in one modality but not in the other (*e.g.*, regions appearing with a similar optical color but with different depths).

6.2.2. Experimental Set-up

The procedure followed for the RGB/depth data set is identical to the one described for the Hyperspectral/LiDAR data set. The two BPT representations are built using region models and merging criteria defined as the mean value and Euclidean distance for each modality. The leaf partition π_0, defined as the refinement infimum of two independent mean shift procedure, is composed of 506 regions. The braid B is constructed by first extracting

Table 2: Number of regions $|\pi|$ and average GOF $\epsilon(\pi|\mathcal{I}_i)$ of optimal partitions $\pi_1^\star, \pi_2^\star, \pi_{[23]}^\star, \pi_B^\star$ for the RGB/depth data set with respect to both modalities $\mathcal{I}_1$ (depth map) and $\mathcal{I}_2$ (RGB image). Lowest values are in bold.

	$\pi_1^\star$	$\pi_2^\star$	$\pi_{[23]}^\star$	$\pi_B^\star$
$\|\pi\|$	163	158	162	162
$\epsilon(\pi\|\mathcal{I}_1)$	4.2	30.8	8.0	**3.9**
$\epsilon(\pi\|\mathcal{I}_2)$	51.6	**13.4**	26.5	13.9

a cut from the hierarchy $H_1^\star$ built on the depth map (the modality showing less fine details). This cut, composed of 50 regions, steers the extraction of two cuts from $H_2^\star$, containing 271 and 279 regions, respectively. The final cut is selected from $H_1^\star$ and comprises 417 regions. The construction of the monitor hierarchy H_m is done in the same manner as the previous data set. The multimodal energy $\mathcal{E}_\lambda^B$ is minimized with $\lambda = 2.5.10^{-5}$ and leads to the braid-based optimal segmentation $\pi_B^\star$ composed of 162 optimal regions.

The collaborative BPT construction procedure is equally constructed with region models and merging criteria being defined as those of the marginal approaches, and the consensus policy remaining the best median ranking.

6.2.3. Results

Table 2 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 $\mathcal{I}_1$ and $\mathcal{I}_2$. For the RGB/depth data set, $\mathcal{I}_1$ corresponds to the depth map while $\mathcal{I}_2$ is a frame of the stereo image.

The observations that arise when analyzing table 2 are analogous to those of table 1: each optimal partition $\pi_1^\star$ and $\pi_2^\star$ scores a low average GOF value with

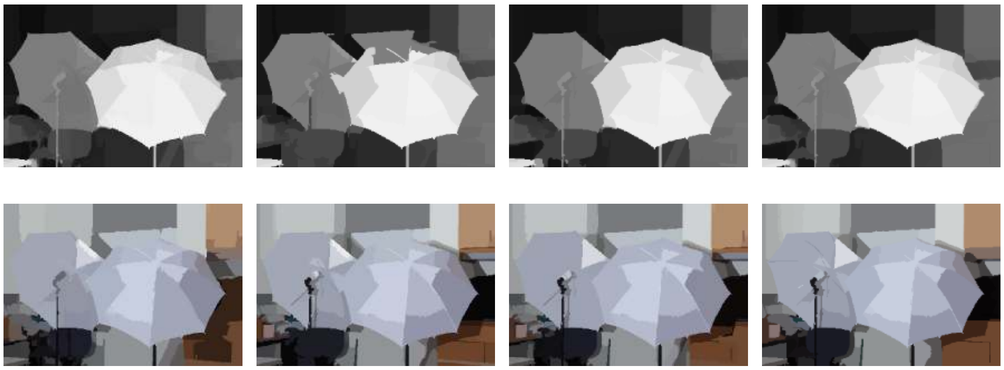


Figure 10: From left to right: optimal partitions $\pi_1^\star$, $\pi_2^\star$, $\pi_{[23]}^\star$ and $\pi_B^\star$ represented with their mean value with respect to the depth modality $\mathcal{I}_1$ (top row) and mean RGB value with respect to the RGB modality $\mathcal{I}_2$ (bottom row).

respect to its own modality and an increased error with respect to the other one. Contrarily, the collaborative optimal cut $\pi_{[23]}^\star$ achieves a higher average GOF values than the marginal approaches with respect to their modality, while scoring a lower value with respect to the alternate modality, illustrating again the averaging phenomenon due to the consensus strategy. On the other hand, the braid optimal cut $\pi_B^\star$ outperforms the depth optimal cut $\pi_1^\star$ with respect to $\mathcal{I}_1$ and achieves a comparable value on $\mathcal{I}_2$ with respect to $\pi_2^\star$, with a similar number of regions. It demonstrates again the consistency of the braid structure for multimodal image representation as its optimal partition is able to better fit both modalities at the same time.

Figure 10 displays the two optimal marginal partitions $\pi_1^\star$ and $\pi_2^\star$, the optimal collaborative partition $\pi_{[23]}^\star$ and the optimal braid partition $\pi_B^\star$, represented by their mean depth with respect to the depth map $\mathcal{I}_1$ (top row) and their mean color with respect to the RGB modality $\mathcal{I}_2$ (bottom row). Its qualitative analysis leads to comparable observations to those of the hyperspectral/LiDAR data set. Both marginal optimal cuts $\pi_1^\star$ and $\pi_2^\star$ contain regions that well describe their respective modality while being inaccurate with respect to the other modality (the various objects on top of the desk located on the bottom left corner of the image, or the half-shaded drawer on the bottom right corner for $\pi_1^\star$, or the most forward umbrella, whose boundaries are either confused with the wall behind or with the second umbrella, due to their similar whitish color for $\pi_2^\star$). The collaborative optimal segmentation $\pi_{[23]}^\star$ preserves all apparent semantic regions at the expense of fine details, such as the small structures within the leftmost umbrella. Finally, all those regions are well segmented in the braid optimal cut $\pi_B^\star$, confirming again that both modalities have collaborated within the braid framework, firstly to derived a consistent hierarchical representation of the multimodal image, and then to design a more accurate partition of this multimodal image.

7. Conclusion

In conclusion, we presented in this article a novel methodology for the hierarchical representation and segmentation of multimodal images. In particular, we took advantage of the newly introduced concept of braids of partitions being a potential solution to the fusion of multiple hierarchical representations issues. 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 representation for such images. The actual main limitation of braids of partitions being the lack of clear guidelines to check the validity of such structure given a family of partitions, we presented here an iterative procedure to extract two cuts from two different and supposedly unrelated hierarchies and guarantee that they

form a braid. Following, we proposed to process the resulting braid structure through 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. The proposed methodology was investigated on several multimodal data sets (two scenarios in this article along with two additional multimodalities presented in supplementary materials) featuring different characteristics. The obtained results demonstrated, quantitatively and qualitatively, the ability of the proposed approach 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, outperforming or equaling typical marginal segmentation results obtained by considering only one modality or by applying some consensus strategy.
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, as well as practical consideration such as investigating other applications than segmentation.

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] 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.

[2] R. W. So, A. C. Chung, A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using Bhattacharyya distances, Pattern Recognition 62 (2017) 161–174.

[3] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, J. A. Benediktsson, Challenges and opportunities of multimodality and data fusion in remote sensing, Proceedings of the IEEE 103 (9) (2015) 1585–1601.

[4] R. A. Finkel, J. L. Bentley, Quad trees a data structure for retrieval on composite keys, Acta Informatica 4 (1) (1974) 1–9.

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

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

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

[8] 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.

[9] 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.

[10] D. Tuia, J. Muñoz-Marí, G. Camps-Valls, Remote sensing image segmentation by active queries, Pattern Recognition 45 (6) (2012) 2180–2192.

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

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

[13] 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.

[14] L. Wald, Some terms of reference in data fusion, IEEE Transactions on geoscience and remote sensing 37 (3) (1999) 1190–1193.

[15] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Information Fusion 4 (4) (2003) 259–280.

[16] N. Cvejic, D. Bull, N. Canagarajah, Region-based multimodal image fusion using ICA bases, Sensors Journal, IEEE 7 (5) (2007) 743–751.

[17] D. Ravì, M. Bober, G. Farinella, M. Guarnera, S. Battiato, Semantic segmentation of images exploiting DCT based features and random forest, Pattern Recognition 52 (2016) 260–273.

[18] C. Rother, T. Minka, A. Blake, V. Kolmogorov, Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs, in: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 1, IEEE, 2006, pp. 993–1000.

[19] T. Rohlfing, C. R. Maurer Jr, Shape-based averaging, Image Processing, IEEE Transactions on 16 (1) (2007) 153–161.

[20] P. Wattuya, K. Rothaus, J.-S. Praßni, X. Jiang, A random walker based approach to combining multiple segmentations, in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE, 2008, pp. 1–4.

[21] Q. Zhou, B. Zheng, W. Zhu, L. J. Latecki, Multi-scale context for scene labeling via flexible segmentation graph, Pattern Recognition(to appear).

[22] L. Franek, D. D. Abdala, S. Vega-Pons, X. Jiang, Image segmentation fusion

using general ensemble clustering methods, in: Computer Vision–ACCV 2010, Springer, 2011, pp. 373–384.

[23] 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.

[24] Y. Tarabalka, J. Tilton, J. Benediktsson, J. Chanussot, A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations, Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 5 (1) (2012) 262–272.

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

[26] 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.

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

[28] 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.

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

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

[31] 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.

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

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

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