



**HAL**  
open science

## Hierarchical segmentation from a non-increasing edge observation attribute

Edward Cayllahua-Cahuina, Jean Cousty, Silvio Jamil F. Guimarães, Yukiko Kenmochi, Guillermo Cámara-Chávez, Arnaldo de Albuquerque Araújo

► **To cite this version:**

Edward Cayllahua-Cahuina, Jean Cousty, Silvio Jamil F. Guimarães, Yukiko Kenmochi, Guillermo Cámara-Chávez, et al.. Hierarchical segmentation from a non-increasing edge observation attribute. Pattern Recognition Letters, 2020, 131, pp.105-112. 10.1016/j.patrec.2019.12.014 . hal-02959334

**HAL Id: hal-02959334**

**<https://hal.science/hal-02959334>**

Submitted on 6 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Hierarchical segmentation from a non-increasing edge observation attribute

Edward Cayllahua-Cahuina<sup>a,b,\*\*</sup>, Jean Cousty<sup>a</sup>, Silvio Jamil F. Guimarães<sup>c</sup>, Yukiko Kenmochi<sup>a</sup>, Guillermo Cámara-Chávez<sup>d</sup>, Arnaldo de Albuquerque Araújo<sup>a,b</sup>

<sup>a</sup>Université Paris-Est, LIGM, CNRS - ENPC - ESIEE Paris - UPEM, 2 Boulevard Blaise Pascal, Noisy-le-Grand 93160, France

<sup>b</sup>Universidade Federal de Minas Gerais, Computer Science Department, Av. Pres. Antônio Carlos 6627, Belo Horizonte 31270-901, Brazil

<sup>c</sup>Pontifical Catholic University of Minas Gerais, Computer Science Department, VIPLAB, Rua Walter Ianni, 255, Belo Horizonte 31980-110, Brazil

<sup>d</sup>Universidade Federal de Ouro Preto, Computer Science Department, Rua P. Paulo M. Gomes 122, Ouro Preto 35400-000, Brazil

### ABSTRACT

Hierarchical image segmentation provides region-oriented scale-spaces: sets of image segmentations at different detail levels in which the segmentations at finer levels are nested with respect to those at coarser levels. Guimarães *et al.* proposed a hierarchical graph-based image segmentation (HGB) method based on the Felzenszwalb-Huttenlocher dissimilarity. It computes, for each edge of a graph, the minimum scale in a hierarchy at which two regions linked by this edge should be merged according to the dissimilarity. We provide an explicit definition of the (edge-) observation attribute and Boolean criterion which are at the basis of this method and show that they are not increasing. Then, we propose an algorithm to compute all the scales for which the criterion holds true. Finally, we propose new methods to regularize the observation attribute and criterion and to set up the observation scale value of each edge of a graph, following the current trend in mathematical morphology to study criteria which are not increasing on a hierarchy. Assessments on Pascal VOC 2010 and 2012 show that these strategies lead to better segmentation results than the ones obtained with the original HGB method.

© 2019 Elsevier Ltd. All rights reserved.

### 1. Introduction

Hierarchical image segmentation gives a multi-scale representation for image analysis. Hierarchical image analysis was initially explored by [1] and has gained a lot of attention, as confirmed by the popularity of [2]. Mathematical morphology has been used in hierarchical image analysis with, *e.g.*, hierarchical watersheds [3, 4, 5, 6], binary partition trees [7, 8, 9], quasi-flat zone hierarchies [10, 11], and tree-based shape spaces [12]. Other methods for hierarchical image analysis consider regular and irregular pyramids [13], scale-set theory [14], multi-scale combinatorial grouping [15], series of optimization problems [16], and generic image classification based on convolutional neural networks [17].

A hierarchical image segmentation is a series of image segmentations at different detail levels such that the segmentation at a given level can be produced by merging regions of the segmentation at the previous level. As a result, in such hierarchical

representation, any region that belongs to a finer detail level is included (or nested) within a region that belongs to a coarser level. The level of a segmentation in a hierarchy is also called an *observation scale*. In [18], Guimarães *et al.* proposed a hierarchical graph-based image segmentation (HGB) method inspired by the Felzenszwalb-Huttenlocher dissimilarity measure [19]. The HGB method computes, for each edge of a graph, the minimum observation scale in a hierarchy at which two regions linked by this edge should merge according to the dissimilarity.

In this article, we give an explicit definition of the edge observation attribute and Boolean criterion which are at the basis of the HGB method and which make possible to select an observation scale for each edge of the graph. We show that this attribute and the related criterion are not increasing with respect to scales. We furthermore notice that the HGB method handles this non-increasing behavior with a simple rule, namely the min-decision rule, known in mathematical morphology since the work of Salembier *et al.* [20]. Recently, more elaborated strategies have been proposed to deal with attributes which are not increasing with respect to scales [12, 21]. These strategies are very successful in applications. Following these trends, we

\*\*Corresponding author: Tel.: +51-54-233272  
e-mail: [edward.cayllahua@esiee.fr](mailto:edward.cayllahua@esiee.fr) (Edward Cayllahua-Cahuina)

propose several new strategies to handle the non-increasing behavior of the observation attribute and criterion behind the HGB method. These strategies allow us to select an observation scale for each edge of the graph. Intuitively, they amount to filtering the set of all the scales for which the Boolean observation criterion holds true, before selecting one scale, with a simple strategy (such as the min-decision rule), among the scales remaining after the filtering. Such filtering is also done with rank and connected filters [22]. Using the segmentation evaluation framework proposed in [6], we show that the newly observation scale selection strategies significantly outperform the original HGB method on Pascal VOC 2010 and 2012 datasets [23, 24] (see an illustration in Fig. 1). Compared to a preliminary version of this work presented in a conference [25], the present article furthermore introduces the observation attribute behind the HGB methods, proposes several new strategies to select edge observation scales, and includes a deeper assessment considering notably a larger image dataset.

Section 2 briefly reminds the main idea of the HGB method. Section 3 introduces the edge observation attribute and criterion behind the HGB method and shows that they are not increasing with respect to scales. Section 4 presents an algorithm to compute all scales for which the observation criterion holds true making possible to consider new observation scale selection strategies as proposed in Section 5. Finally, Section 6 presents the assessment of the newly proposed method.

## 2. Hierarchical graph-based image segmentation

This section aims at explaining the method of hierarchical graph-based image segmentation (HGB) [18]. We first give a series of necessary notions such as quasi-flat zone hierarchies [10], and then describe the HGB method.

### 2.1. Hierarchies

Given a finite set  $V$ , a *partition* of  $V$  is a set  $\mathbf{P}$  of nonempty disjoint subsets of  $V$  whose union is  $V$ . Any element of  $\mathbf{P}$  is called a *region* of  $\mathbf{P}$ . Given two partitions  $\mathbf{P}$  and  $\mathbf{P}'$  of  $V$ ,  $\mathbf{P}'$  is said to be a *refinement* of  $\mathbf{P}$ , denoted by  $\mathbf{P}' \leq \mathbf{P}$ , if any region of  $\mathbf{P}'$  is included in a region of  $\mathbf{P}$ . A hierarchy on  $V$  is a sequence  $\mathcal{H} = (\mathbf{P}_0, \dots, \mathbf{P}_\ell)$  of partitions of  $V$ , such that  $\mathbf{P}_{i-1} \leq \mathbf{P}_i$ , for any  $i \in \{1, \dots, \ell\}$ .

### 2.2. Graph and connected-component partition

A *graph* is a pair  $G = (V, E)$  where  $V$  is a finite set and  $E$  is a subset of  $\{\{x, y\} \subseteq V \mid x \neq y\}$ . Each element of  $V$  is called a *vertex* of  $G$ , and each element of  $E$  is called an *edge* of  $G$ . A subgraph of  $G$  is a graph  $(V', E')$  such that  $V' \subseteq V$  and  $E' \subseteq E$ . If  $X$  is a graph, its vertex and edge sets are denoted by  $V(X)$  and  $E(X)$ , respectively.

If two vertices of a graph  $G$  are joined by an edge, we say that they are adjacent. From the reflexive–transitive closure of this adjacency relation on a finite set  $V(G)$ , we derive the connectivity relation on  $V(X)$ . It is an equivalence relation, whose equivalence classes are called connected components of  $G$ . We denote by  $\mathbf{C}(G)$  the set of all connected components of  $G$ . Note that  $\mathbf{C}(G)$  is a partition of  $V(G)$ , called the *connected-component partition induced by  $G$* .

### 2.3. Quasi-flat zone hierarchies

Given a graph  $G = (V, E)$ , let  $w$  be a map from  $E$  in the set  $\mathbb{R}$  of real numbers. For any edge  $u$  of  $G$ , the value  $w(u)$  is called the weight of  $u$  (for  $w$ ), and the pair  $(G, w)$  is called an *edge-weighted graph*. Any edge-weighted graph can be transformed into a series of connected-component partitions, which constitutes a hierarchy, called the quasi-flat zone hierarchy of  $(G, w)$  (see a formal definition below). This transform is a bijection between the hierarchies and a subset of the edge weighted graphs called the saliency maps [26]. Hence, any edge-weighted graph induces a quasi-flat zone hierarchy and any hierarchy  $\mathcal{H}$  can be represented by an edge-weighted graph whose quasi-flat zone hierarchy is precisely  $\mathcal{H}$  [26]. This bijection allows us to handle any hierarchy with an edge-weighted graph.

Given an edge-weighted graph  $(G, w)$ , let  $X$  be a subgraph of  $G$  and let  $\lambda$  be a value of  $\mathbb{R}$ . The  $\lambda$ -*level edge set* of  $X$  for  $w$  is defined by  $w_\lambda(X) = \{u \in E(X) \mid w(u) < \lambda\}$ , and the  $\lambda$ -*level graph* of  $X$  for  $w$  is defined as the subgraph  $w_\lambda^V(X)$  of  $X$  such that  $w_\lambda^V(X) = (V(X), w_\lambda(X))$ . Then, the connected-component partition  $\mathbf{C}(w_\lambda^V(X))$  induced by  $w_\lambda^V(X)$  is called the  $\lambda$ -*level partition* of  $X$  for  $w$ .

As we consider only finite graphs and hierarchies, the set of considered levels is reduced to a finite subset of  $\mathbb{R}$  denoted by  $\mathbb{E}$  in the sequel of this article. In order to browse values in this set and to round real values to values of  $\mathbb{E}$ , we define, for any  $\lambda \in \mathbb{R}$ ,  $p_{\mathbb{E}}(\lambda) = \max\{\mu \in \mathbb{E} \cup \{-\infty\} \mid \mu < \lambda\}$ ,  $n_{\mathbb{E}}(\lambda) = \min\{\mu \in \mathbb{E} \cup \{\infty\} \mid \mu > \lambda\}$ , and  $\hat{n}_{\mathbb{E}}(\lambda) = \min\{\mu \in \mathbb{E} \cup \{\infty\} \mid \mu \geq \lambda\}$ . In our experiments, the set  $\mathbb{E}$  is a finite integer interval.

Let  $(G, w)$  be an edge-weighted graph and let  $X$  be a subgraph of  $G$ . The sequence of all  $\lambda$ -level partitions of  $X$  for  $w$ , ordered by increasing value of  $\lambda$ , is a hierarchy, defined by  $\mathcal{QFZ}(X, w) = (\mathbf{C}(w_\lambda^V(X)) \mid \lambda \in \mathbb{E} \cup \{\infty\})$ , and called the *quasi-flat zone hierarchy of  $X$  for  $w$* . Let  $\mathcal{H}$  be the quasi-flat zone hierarchy of  $G$  for  $w$ . Given a vertex  $x$  of  $G$  and a value  $\lambda$  in  $\mathbb{E}$ , the region that contains  $x$  in the  $\lambda$ -level partition of the graph  $G$  is denoted by  $\mathcal{H}_x^\lambda$ .

Let us consider a minimum spanning tree  $T$  of  $(G, w)$ . It has been shown in [26] that  $\mathcal{QFZ}(T, w)$  of  $T$  for  $w$  is the same as  $\mathcal{QFZ}(G, w)$  of  $G$  for  $w$ . This indicates that the quasi-flat zone hierarchy for  $G$  can be handled by its minimum spanning tree.

### 2.4. Hierarchical graph-based segmentation method

The input of the HGB method is the edge-weighted graph  $(G, w)$  representing an image, whose pixels correspond to the vertices of  $G$  and whose edges link adjacent pixels. The weight  $w(u)$  of an edge linking two pixels  $x$  and  $y$  represents the dissimilarity between the pixels  $x$  and  $y$ . As the HGB method is grounded in a general graph framework, it can be implemented with any adjacency relation and dissimilarity measure. For the experiments in Section 6, the 4-adjacency relation and the  $L_2$  norm in the RGB color space are used.

Before presenting the HGB method, we describe the observation scale dissimilarity [18] used by the method and which originates from the region merging criterion proposed in [19].

#### 2.4.1. Observation scale dissimilarity

Let  $R_1$  and  $R_2$  be two adjacent regions. The dissimilarity measure between  $R_1$  and  $R_2$  consists of the so-called



Fig. 1: (Left) original image extracted from the Pascal VOC 2010 dataset, and the saliency maps (an equivalent representation of hierarchical segmentation) resulting from the HGB method using the original observation scale (middle) and from one of our proposed observation scale selection strategies (right). A video version of this illustration showing the successive levels of the hierarchy is available at [www.tobeinsertedByPRLwhenAvailable.com](http://www.tobeinsertedByPRLwhenAvailable.com).

---

### Method 1: HGB method

---

**Input** : A minimum spanning tree  $T$  of an edge-weighted graph  $(G, w)$

**Output**: A hierarchy  $\mathcal{H} = \mathcal{QFZ}(T, f)$

```

1 for each  $u \in E(T)$  do  $f(u) := \max\{\lambda \in \mathbb{E}\}$ ;
2 for each  $u \in E(T)$  in non-decreasing order for  $w$  do
3    $\mathcal{H} := \mathcal{QFZ}(T, f)$ ;
4    $f(u) := \mathbb{P}_{\mathbb{E}}(\lambda_{\mathcal{H}}^*(u))$ ;
5 end
6  $\mathcal{H} := \mathcal{QFZ}(T, f)$ ;

```

---

inter-component and within-component differences [19]. The *inter-component difference* between  $R_1$  and  $R_2$  is defined by  $\Delta_{inter}(R_1, R_2) = \min\{w(\{x, y\}) \mid x \in R_1, y \in R_2, \{x, y\} \in E(T)\}$ , while the *within-component difference* of a region  $R$  is defined by  $\Delta_{within}(R) = \max\{w(\{x, y\}) \mid x, y \in R, \{x, y\} \in E(T)\}$ . They lead to the *observation scale of  $R_1$  relative to  $R_2$* , defined by  $S_{R_2}(R_1) = (\Delta_{inter}(R_1, R_2) - \Delta_{within}(R_1)) / |R_1|$ , where  $|R_1|$  is the cardinality of  $R_1$ . Then, a symmetric metric between  $R_1$  and  $R_2$ , called the *observation scale dissimilarity between  $R_1$  and  $R_2$* , is defined by

$$D(R_1, R_2) = \max\{S_{R_2}(R_1), S_{R_1}(R_2)\}. \quad (1)$$

This dissimilarity is used to determine if two regions should be merged or not at a certain observation scale in the following.

#### 2.4.2. HGB Method

The HGB method [18] is presented in Method 1. The input is a graph  $G$  representing an image with its associated weight function  $w$ , where the minimum spanning tree  $T$  of  $G$  is taken. Given  $(T, w)$ , the HGB method computes a new weight function  $f$  which leads to a new hierarchy  $\mathcal{H} = \mathcal{QFZ}(T, f)$ . The resulting hierarchy  $\mathcal{H}$  is considered as the hierarchical image segmentations of the initial image. Thus, the core of the method is the generation of the weight function  $f$  for  $T$ .

After initializing all values of  $f$  to infinity (see Line 1), we compute an observation scale value  $f(u)$  for each edge  $u \in E(T)$  in non-decreasing order with respect to the original weight  $w$  (see Line 2). Hence, at each iteration, the value  $f(u)$  is modified. This means that the weight map  $f$  is modified and we need to update  $\mathcal{H}$  (Line 3) such that we still have  $\mathcal{H} = \mathcal{QFZ}(T, f)$ .

An efficient algorithm for the hierarchy update can be found in [27]. Once  $\mathcal{H}$  is updated, the value  $\lambda_{\mathcal{H}}^*(u)$  of a finite subset  $\mathbb{E}$  of  $\mathbb{R}$ , with  $u = \{x, y\}$ , is obtained by

$$\lambda_{\mathcal{H}}^*(\{x, y\}) = \min\{\lambda \in \mathbb{E} \mid D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda) \leq \lambda\}. \quad (2)$$

We first consider the regions  $\mathcal{H}_x^\lambda$  and  $\mathcal{H}_y^\lambda$  at a level  $\lambda$ . Using the dissimilarity measure  $D$ , we check if  $D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda) \leq \lambda$ . Equation (2) states that the observation scale  $\lambda_{\mathcal{H}}^*(\{x, y\})$  is the minimum value  $\lambda$  for which this assertion holds.

### 3. Non-increasing observation attribute and criterion

In this section, we provide a formal definition of the observation attribute and criterion involved in Equation (2). Then, we discuss their non-increasing behavior opening the doors towards new strategies to select observation scale values based on Felzenszwalb-Huttenlocher dissimilarity measure as used in Method 1.

In the remaining part of this section, we consider that  $\mathcal{H}$  is any hierarchy and that  $u = \{x, y\}$  is any edge of  $T$ .

The *edge observation attribute* (of  $u = \{x, y\}$  in  $\mathcal{H}$ ) is the map  $\mathcal{A}$  from  $\mathbb{E}$  into  $\mathbb{R}$  defined by:

$$\mathcal{A}(\lambda) = \lambda - D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda), \quad (3)$$

for any  $\lambda$  in  $\mathbb{E}$ . The *edge observation criterion* (of  $u$  in  $\mathcal{H}$ ) is the map  $C$  from  $\mathbb{E}$  in the set  $\{true, false\}$  such that  $C(\lambda) = true$  if  $\mathcal{A}(\lambda)$  is greater than or equal to 0; otherwise  $C(\lambda) = false$ . Let  $\lambda$  be any element in  $\mathbb{E}$ , we say that  $\lambda$  is a *positive observation scale* (of  $u$  in  $\mathcal{H}$ ) whenever  $C(\lambda)$  holds true. On the contrary, if  $\lambda$  is not a positive observation scale, then we say that  $\lambda$  is a *negative observation scale* (of  $u$  in  $\mathcal{H}$ ).

Observe that the value  $\lambda_{\mathcal{H}}^*(x, y)$  defined in Equation (2) is simply the lowest element of  $\mathbb{E}$  such that  $C(\lambda)$  is true. In other words,  $\lambda_{\mathcal{H}}^*(x, y)$  is the lowest positive observation scale of  $u$  in  $\mathcal{H}$ . Additionally, we denote by  $\bar{\lambda}_{\mathcal{H}}^*(x, y)$ , the largest negative observation scale of  $u$  in  $\mathcal{H}$ .

Intuitively, a positive observation scale corresponds to a level of the hierarchy  $\mathcal{H}$  for which the two regions linked by  $u$  should be merged according to the observation criterion  $C$ . On the other hand, a negative observation scale corresponds to a level of the hierarchy for which the two associated regions should remain disjoint. Furthermore, for a positive observation scale  $\lambda$ ,

the observation attribute  $\mathcal{A}(\lambda)$  can be intuitively interpreted as a measure of the strength at which the two associated regions should be merged whereas, for a negative observation scale, the observation attribute can be seen as the opposite of the strength at which the two regions should remain disjoint: when the observation attribute is highly positive, the two regions should be merged with a high confidence whereas when the observation attribute is negative and low, the two regions should remain disjoint with a high confidence.

A desirable property would be that the observation criterion  $C$  be increasing with respect to scales, a Boolean criterion  $C$  being increasing whenever, for any scale value  $\lambda \in \mathbb{E}$ ,  $C(\lambda)$  holds true implies that  $C(\lambda')$  holds true for any scale  $\lambda'$  greater than  $\lambda$ . Indeed, in such desirable case, any level in  $\mathbb{E}$  greater than  $\lambda_{\mathcal{H}}^*(x, y)$  would be a positive observation scale, whereas any level not greater than  $\lambda_{\mathcal{H}}^*(x, y)$  would be a negative scale. In other words, we would have  $\lambda_{\mathcal{H}}^*(x, y) = n_{\mathbb{E}}(\bar{\lambda}_{\mathcal{H}}^*(x, y))$ . Hence, it would be easily argued that the observation scale of the edge  $u$  must be set to  $\lambda_{\mathcal{H}}^*(x, y)$ . However, in general, the observation attribute  $\mathcal{A}$  and thus the observation criterion  $C$  are not increasing (see a counterexample in Fig. 2) and we have  $\lambda_{\mathcal{H}}^*(x, y) < n_{\mathbb{E}}(\bar{\lambda}_{\mathcal{H}}^*(x, y))$ . Therefore, it is interesting to investigate strategies which can be adopted to select a significant observation scale between  $\lambda_{\mathcal{H}}^*(x, y)$  and  $\bar{\lambda}_{\mathcal{H}}^*(x, y)$ . A graphical illustration of different situations which may occur for a given edge  $u$  and hierarchy  $\mathcal{H}$  is presented in Fig. 2. In other words, it is interesting to investigate strategies to transform the observation attribute  $\mathcal{A}$  and criterion  $C$  into an increasing attribute  $\mathcal{A}'$  and an increasing criterion  $C'$ .

In the framework of mathematical morphology, non-increasing regional attributes/criteria are known to be useful but difficult to handle. Several rules or strategies to handle non-increasing criteria have been considered in the context of connected filters. Among them, one may cite the min- and max-rules [20], the Vitterbi [20] and shape-space filtering strategies [12], and the maxTree pruning strategies based on graph signal processing [21]. Note that the strategy adopted in Equation (2) corresponds to the min-rule and that the strategy consisting of selecting  $\bar{\lambda}_{\mathcal{H}}^*(x, y)$  corresponds to the max-rule. Our main goal in this article is to investigate other strategies to efficiently

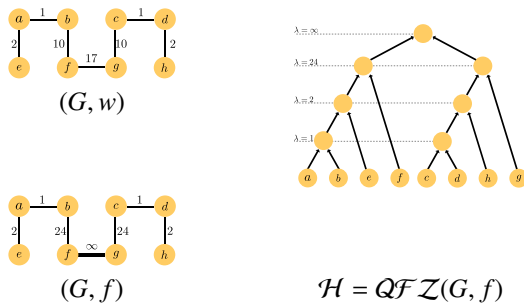


Fig. 2: Counterexample for the increasing property of the observation attribute  $\mathcal{A}$  and criterion  $C$ : for the edge  $u = \{x, y\}$  with  $x = f$ ,  $y = g$ , we have  $D(\mathcal{H}_x^1, \mathcal{H}_y^1) = 17$ ,  $D(\mathcal{H}_x^{23}, \mathcal{H}_y^{23}) = 17$ , and  $D(\mathcal{H}_x^{25}, \mathcal{H}_y^{25}) = 28$ . Hence, we have  $\mathcal{A}(1) = 1 - 17 = -16$ ,  $\mathcal{A}(23) = 23 - 17 = 6$ , and  $\mathcal{A}(25) = 25 - 28 = -3$ , so that  $C(1) = \text{false}$ ,  $C(23) = \text{true}$ , and  $C(25) = \text{false}$ , which shows that  $\mathcal{A}$  and  $C$  are not increasing.

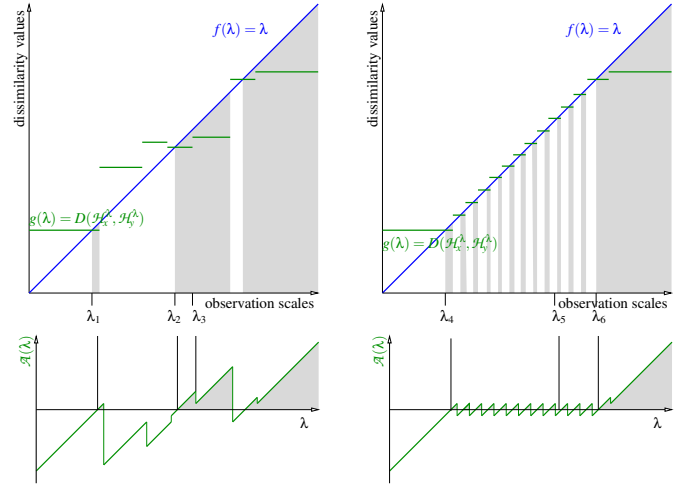


Fig. 3: Illustration of the observation attribute and criterion and of possible observation scale selection strategies. The positive observation intervals (the intervals in which the observation criterion holds true) are represented in gray. On the top, the dissimilarity measure  $D(\mathcal{H}_x^1, \mathcal{H}_y^1)$  is depicted in green as a function of  $\lambda$ , whereas, on the bottom, the observation attribute  $\mathcal{A}(\lambda)$  is depicted in green as a functions of  $\lambda$ . On the left, the min-, the lower-length and the lower  $p$ -rank selection strategies select the scales  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , respectively (for a length threshold which is a little larger than the leftmost gray interval and for  $p = 0.3$ ), whereas, on the right, the max-, the upper-length and the upper  $p$ -rank selection strategies select the scales  $\lambda_6$ ,  $\lambda_4$  and  $\lambda_5$ , respectively. Note that, for the sake of readability, the values of the functions are depicted for all scales taken in  $\mathbb{R}$ , whereas the definitions are given only for a finite subset  $\mathbb{E}$  of  $\mathbb{R}$ .

handle the non-increasing observation criterion  $C$  in the context of hierarchical segmentation based on the Felzenszwalb-Huttenlocher region dissimilarity measure. These strategies are based on the analysis of all positive and negative observation scales and their associated attribute values. Therefore, before presenting these strategies, it is necessary to provide a method to obtain all positive and negative observation scales.

#### 4. Observation intervals: definition and algorithm

As established in the previous section, the observation attribute and observation criterion (obtained by thresholding the observation attribute at value 0) at the basis of the HGB method are not increasing. Hence, in order to find the sets of all scales for which the criterion holds true (*i.e.*, the scales for which the regions linked by an edge  $u$  should be merged), it is not enough to find the minimum of the positive observation scales. In [27], an algorithm to find the minimum observation scales such that the observation criterion holds true is given. Instead, in this section, an algorithm to find all scales at which the observation criterion holds true is provided. The proposed algorithm searches for the bounds of all (maximal) scale intervals for which the criterion holds true. Before presenting the algorithm that computes the bounds of all these intervals, we start with the definition of such intervals, called positive observation intervals.

Let  $\lambda_1$  and  $\lambda_2$  be any two real numbers in  $\mathbb{E} \cup \{-\infty\}$  such that  $\lambda_1 < \lambda_2$ . We denote by  $\llbracket \lambda_1, \lambda_2 \rrbracket_{\mathbb{E}}$  the subset of  $\mathbb{E}$  that contains every element of  $\mathbb{E}$  that is both greater than  $\lambda_1$  and not greater than  $\lambda_2$ :  $\llbracket \lambda_1, \lambda_2 \rrbracket_{\mathbb{E}} = \{\lambda \in \mathbb{E} \mid \lambda_1 < \lambda \leq \lambda_2\}$ . We say that a subset  $I$  of  $\mathbb{E}$  is an *open-closed interval of  $\mathbb{E}$* , or simply an

interval, if there exist two real values  $\lambda_1$  and  $\lambda_2$  in  $\mathbb{E}$  such that  $I$  is equal to  $\llbracket \lambda_1, \lambda_2 \rrbracket_{\mathbb{E}}$ .

**Definition 1 (observation interval).** Let  $\mathcal{H}$  be any hierarchy, let  $u$  be any edge in  $E(T)$ , and let  $I$  be an interval. We say that  $I$  is a positive observation interval (resp. a negative observation interval) for  $(\mathcal{H}, u)$  if the two following statements hold true:

1. any element in  $I$  is a positive (resp. negative) observation scale for  $(\mathcal{H}, u)$ ; and
2.  $I$  is maximal among all intervals for which statement (1) holds true, i.e., any interval which is a proper superset of  $I$  contains a negative (resp. positive) observation scale for  $(\mathcal{H}, u)$ .

The set of all positive (resp. negative) observation intervals is denoted by  $\Lambda_{\mathcal{H}}(u)$  (resp. by  $\bar{\Lambda}_{\mathcal{H}}(u)$ ).

In order to compute  $\Lambda_{\mathcal{H}}(\{x, y\})$ , we follow the strategy presented in [27], which relies on the *component tree* of the hierarchy  $\mathcal{H}$ . The *component tree* of  $\mathcal{H}$  is the pair  $\mathcal{T}_{\mathcal{H}} = (\mathcal{N}, \text{parent})$  such that  $\mathcal{N}$  is the set of all regions of  $\mathcal{H}$  and such that a region  $R_1$  in  $\mathcal{N}$  is a *parent* of a region  $R_2$  in  $\mathcal{N}$  whenever  $R_1$  is a minimal (for inclusion relation) proper superset of  $R_2$ . Note that every region in  $\mathcal{N}$  has exactly one parent, except the region  $V$  which has no parent and is called the *root* of the component tree of  $\mathcal{H}$ . Any region which is not the parent of another one is called a *leaf* of the tree. It can be observed that any singleton of  $V$  is a leaf of  $\mathcal{T}_{\mathcal{H}}$  and that conversely any leaf of  $\mathcal{T}_{\mathcal{H}}$  is a singleton of  $V$ . The *level* of a region  $R$  in  $\mathcal{H}$  is the highest index of a partition that contains  $R$  in  $\mathcal{H}$ . Then, the proposed algorithm, whose precise description is given in Algorithm 1, browses in increasing order the levels of the regions containing  $x$  and  $y$  until finding a value  $\lambda$  such that  $D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda) \leq \lambda$ , or, in other words, such that  $\mathcal{A}(\lambda) > 0$  and hence,  $C = \text{true}$ . This value is then  $\lambda_{\mathcal{H}}^*(x, y)$  defined by Equation (2). This value is also the lower bound of the first positive observation interval. If we keep browsing the levels of the regions containing  $x$  and  $y$  in this tree, as long as  $D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda) \leq \lambda$ , we can identify the upper bound of this first positive observation interval. We can further continue to browse the levels of the regions containing  $x$  and  $y$  in the tree in order to identify all positive observation intervals. Therefore, at the end of the execution, we can return the set  $\Lambda_{\mathcal{H}}(\{x, y\})$  of all positive observation intervals. From the set  $\Lambda_{\mathcal{H}}(\{x, y\})$ , we can obtain by duality the set  $\bar{\Lambda}_{\mathcal{H}}(\{x, y\})$  of all negative observation intervals.

The time complexity of Algorithm 1 depends linearly on the number of regions in the branches of the component tree of  $\mathcal{H}$  containing  $x$  and  $y$  since it consists of browsing all these regions from the leaves to the root. In the worst case, at every level of the hierarchy the region containing  $x$  is merged with a singleton region. Hence, as there are  $|V|$  vertices in  $G$ , in this case, the branch of  $x$  contains  $|V|$  regions. Thus, the worst-case time complexity of Algorithm 1 is  $O(|V|)$ . However, in many practical cases, the component tree of  $\mathcal{H}$  tends to be balanced and each region of  $\mathcal{H}$  results from the merging of two regions of (approximately) the same size. Then, if the tree is balanced, the branch of  $x$  contains  $O(\log_2(|V|))$  nodes and the time-complexity of Algorithm 1 reduces to  $O(\log_2(|V|))$ .

---

**Algorithm 1:** Positive observation intervals
 

---

**Input :** The component tree  $\mathcal{T} = (\mathcal{N}, \text{parent})$  of a hierarchy  $\mathcal{H}$ , an edge  $u = \{x, y\}$  of  $T$ , an array *level* that stores the level of every region of  $\mathcal{H}$   
**Output:** A set  $L$  containing all elements of  $\Lambda_{\mathcal{H}}(\{x, y\})$

```

1  $C_x := \{x\}; C_y := \{y\}; \Lambda_{\mathcal{H}}(\{x, y\}) = \{;$ 
2  $\lambda := \min(\text{level}[C_x], \text{level}[C_y]); \lambda_{prev} := -\infty;$ 
3 do
4   while  $D(C_x, C_y) > \lambda$  do
5      $\lambda_{prev} := \lambda;$ 
6      $\lambda := \min(\text{level}[\text{parent}[C_x]], \text{level}[\text{parent}[C_y]]);$ 
7     if  $\text{level}[\text{parent}[C_x]] = \lambda$  then  $C_x := \text{parent}[C_x];$ 
8     if  $\text{level}[\text{parent}[C_y]] = \lambda$  then  $C_y := \text{parent}[C_y];$ 
9   end
10   $\lambda_{lower} := \max(\text{n}_{\mathbb{E}}(\lambda_{prev}), \hat{\text{n}}_{\mathbb{E}}(D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda)));$ 
11  while  $D(C_x, C_y) \leq \lambda$  and  $(\text{parent}[C_x] \neq \text{root and}$ 
     $\text{parent}[C_y] \neq \text{root})$  do
12     $\lambda_{prev} := \lambda;$ 
13     $\lambda := \min(\text{level}[\text{parent}[C_x]], \text{level}[\text{parent}[C_y]]);$ 
14    if  $\text{level}[\text{parent}[C_x]] = \lambda$  then  $C_x := \text{parent}[C_x];$ 
15    if  $\text{level}[\text{parent}[C_y]] = \lambda$  then  $C_y := \text{parent}[C_y];$ 
16  end
17   $\lambda_{upper} := \max(\text{n}_{\mathbb{E}}(\lambda_{prev}), \hat{\text{n}}_{\mathbb{E}}(\lambda_{prev}));$ 
18   $L.add(\llbracket \lambda_{lower}, \lambda_{upper} \rrbracket);$ 
19 while  $\text{parent}[C_x] \neq \text{root and } \text{parent}[C_y] \neq \text{root};$ 

```

---

## 5. Selecting observation scales

In this section, we investigate selection strategies to choose an observation scale from the sets of positive (resp. negative) observation intervals, i.e., the set of scales at which the observation criterion holds true (resp. false). In Section 5.1, we present the min-, max-, and rank-rules to select observation scales from either the set of positive or negative observation scales. Note that the min- and max-rules correspond to simple strategies known in mathematical morphology to prune the branch of a tree (aka a hierarchy) based on non-increasing regional attributes and criteria whereas the rank rules extend them to take a more robust decision. Then, in Section 5.2, we propose to improve the results obtained with these rules by pre-filtering out positive or negative observation intervals which might be considered as irrelevant with respect to a given attribute, before applying the min-, max- or rank-rules. Whereas, the selection rules can be seen as a way to transform a non-increasing criterion into an increasing one, the filtering strategies can be seen as regularization strategy for the non-increasing attribute on which the (observation) criterion is based. Overall, the proposed strategy allows us to transform a non-increasing attribute-based criterion into an increasing one based on a regularized version of the observation attribute.

### 5.1. Min-, max- and rank-selection rules

In this section, the symbol  $K$  stands for a finite set of observation scales (either negative or positive) and we define several



useful rules to pick one observation scale in such set. Let  $K$  be any subset of  $\mathbb{E}$ . We propose the two following selection rules from  $K$  to set the value of  $f(u)$  in Method 1:

$$\begin{aligned} \text{min-rule: } f(u) &:= \min\{k \in K\}; \text{ and} \\ \text{max-rule: } f(u) &:= \max\{k \in K\}. \end{aligned}$$

Let us now provide a precise definition of the rank rules. The intuitive idea of these rules is to remove from the set  $K$  the scales higher or lower than a given quantile, which are then considered as non-significant, before applying the min- and max-rules.

Let  $n$  be the number of elements in the set  $K$  of possible observation scales, *i.e.*,  $n = |K|$ . Let  $k$  be any positive integer less than  $n$ . We denote by  $\text{rank}_{k/n}(K)$  the element  $e$  of  $K$  such that there are exactly  $k$  distinct elements in  $K$  which are less than  $e$ . Let  $p$  be any real value between 0 and 1, we set  $\text{rank}_p(K) = \text{rank}_{\lfloor p.n \rfloor/n}(K)$ , where  $\lfloor p.n \rfloor$  is the largest integer which is not greater than the real value  $p.n$ .

Let  $p$  be any real value between 0 and 1. We consider the two following additional selection rules from  $K$  to set the value of  $f(u)$  in Method 1:

$$\begin{aligned} \text{lower } p\text{-rank-rule: } f(u) &:= \text{rank}_p(K); \text{ and} \\ \text{upper } p\text{-rank-rule: } f(u) &:= \text{rank}_{1-p}(K). \end{aligned}$$

Using the above rules, we now define several strategies which are tested in Section 6 to set the value  $f(u)$  in Method 1.

When  $K$  is equal to the set of all positive observation scales (resp. the set of all positive observation scales not greater than  $\bar{\lambda}_{\mathcal{H}}^*(x, y)$ ) for given edge  $u$  and hierarchy  $\mathcal{H}$ , the result obtained with the min-rule (resp. lower  $p$ -rank rule) is called the *min strategy* (resp. *lower  $p$ -rank strategy*). With the min-strategy, the value  $f(u)$  in Method 1 is set to the minimum scale  $\lambda$  for which the observation criterion holds true. Therefore, the results obtained with this strategy correspond exactly to the results obtained with the method presented in [18, 27], as described by Equation (2). The lower  $p$ -rank strategy is a new proposal to take more robust decisions, this assertion being practically evaluated in Section 6. Intuitively, it can be seen that the lower  $p$ -rank strategy considers the lower  $p$  percentile of the observation scales between  $\lambda_{\mathcal{H}}^*(x, y)$  and  $\bar{\lambda}_{\mathcal{H}}^*(x, y)$ . Hence, it can be easily seen that, when  $p$  is equal to 0, the lower  $p$ -rank strategy is exactly the min-strategy.

In this article, we also consider the *max-strategy* (resp. *upper  $p$ -rank strategy*), that is the result obtained with the max-rule (resp. upper  $p$ -rank rule) when  $K$  is the set of the negative observation scales (resp. the set of the negative observation scales not less than  $\lambda_{\mathcal{H}}^*(x, y)$ ). It can be observed that when the observation criterion is increasing, the min-, max-, lower- $p$ -rank, and upper- $p$  rank strategies all take the same decision. However, as illustrated in Fig. 3, when the observation criterion is not increasing, these rules provide different alternatives to set up the observation scale  $f(u)$  in Method 1.

## 5.2. Filtering the observation scales with connected operators

In the previous section, we proposed several strategies to select an observation scale from the positive and negative obser-

vation scales. In particular, in order to provide a more robust alternative to the classical min- and max-rules, the  $p$ -rank rule is presented. In this section, following the same goal of robustifying the observation scale selection process, we propose to filter the set of positive and negative observation scales with a family of filters, called connected operators, coming from the field of mathematical morphology. The basic idea of the connected operators is to act at the level of the connected components of a set. Basically, each connected component of the input set is either completely preserved or completely discarded but cannot be partially included in the result of the operator. The choice of the connected components which are kept is based on an attribute designed to measure the importance of the connected components. Such filters are very powerful and able to solve many practical issues in image analysis and signal processing as reviewed in [22]. In this section, we propose to filter the set of positive and negative observation scales with connected operators before selecting an observation scale with the strategies presented in the previous section. We start this section by defining the notion of a component of a set of scales, and then give a formal definition of a connected operator in this framework. We also describe several measures to assess the importance of a component of scales at the end.

Let  $K$  be any subset of  $\mathbb{E}$ . Any interval  $I = ]\lambda_1, \lambda_2]$  that is included in  $K$  and maximal for this property is called a component of  $K$ . The set of all components of  $K$  is denoted by  $CC(K)$ . For instance, if  $K$  is the set of all positive (resp. negative) observation scales for a given pair  $(\mathcal{H}, u)$ , then  $CC(K)$  is precisely the set  $\Lambda_{\mathcal{H}}(u)$  of all positive observation intervals (resp. the set  $\bar{\Lambda}_{\mathcal{H}}(u)$  of negative observation intervals). Any map from the set of all intervals of  $\mathbb{E}$  into the set of real numbers is called an (*interval*) *attribute*. Let  $\mu$  be any interval attribute, let  $t$  be any real number, and let  $K$  be a subset of  $\mathbb{E}$ . We set

$$\gamma_t^\mu(K) = \cup\{I \in CC(K) \mid \mu(I) > t\}.$$

When the attribute  $\mu$  is increasing with respect to its parameter (which can be any interval), the operator  $\gamma_t^\mu$  is called the  *$\mu$ -connected opening of parameter  $t$*  and is indeed an algebraic opening (*i.e.*, an increasing, anti-extensive and idempotent operator). Hence, using an interval measure or attribute  $\mu$ , we are able to analyze and decide if a certain observation interval is relevant or not. If an observation interval is not relevant, then, it is simply discarded. Afterward, an observation scale is selected among the remaining scales using one of the strategies presented in the previous section. To decide if an observation interval  $I$  is relevant, we consider the following attributes:

$$\text{length: } \mu_1(I) = |I|; \quad (4)$$

$$\text{depth: } \mu_2(I) = \max\{\mathcal{A}(\lambda) \mid \lambda \in I\}; \quad (5)$$

$$\text{Ndepth: } \mu_3(I) = \max\{-\mathcal{A}(\lambda) \mid \lambda \in I\}; \quad (6)$$

$$\text{area: } \mu_4(I) = \sum\{\mathcal{A}(\lambda) \mid \lambda \in I\}; \text{ and} \quad (7)$$

$$\text{Narea: } \mu_5(I) = \sum\{-\mathcal{A}(\lambda) \mid \lambda \in I\}. \quad (8)$$

We remind from Equation (3) that  $\mathcal{A}(\lambda) = \lambda - D(\mathcal{H}_x^\lambda, \mathcal{H}_y^\lambda)$ . An illustration of these attributes is provided in Fig. 4. The attributes of every positive and negative observation interval can

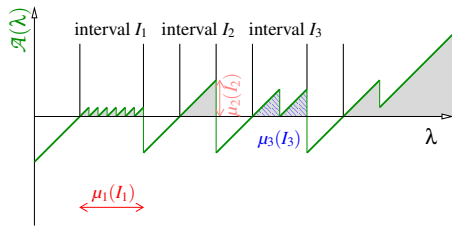


Fig. 4: Illustration of the attributes used to filter the observation intervals. The observation attribute  $\mathcal{A}$  is depicted in green as a function of the scales. The length  $\mu_1(I_1)$  of the interval  $I_1$  is represented in red, the depth  $\mu_2(I_2)$  of  $I_2$  is represented in pink and the area  $\mu_3(I_3)$  of  $I_3$  is shown in blue. Note that, for the sake of readability, the values of the observation attribute  $\mathcal{A}$  are depicted for all scales taken in  $\mathbb{R}$ , whereas the definition is given only for a finite subset  $\mathbb{E}$  of  $\mathbb{R}$ .

be computed on the fly during the execution of Algorithm 1. In our experiments, the positive observation scales are filtered according to the length (Equation (4)), depth (Equation (5)) and area (Equation (7)) measures, whereas the negative observation scales are filtered according to the length (Equation (4)), Ndepth (Equation (6)) and Narea (Equation (8)) measures. Besides the min-, lower- $p$ -rank, max- and upper- $p$ -rank strategies, which are presented in Section 5.1, we also assess in the next section the results of the min-rule on the result of the  $\mu$ -connected openings applied to the positive observation scales when the measure  $\mu$  is either the length, the depth or the area (*i.e.*, either  $\mu_1$ ,  $\mu_2$ , or  $\mu_4$ ). These three strategies are called the *lower-length*, *lower-depth* and *lower-area* selection strategies, respectively. Similarly, we also consider the results of the max-rule on the result the  $\mu$ -connected openings applied to the negative observation scales, when the measure  $\mu$  is either the length, the Ndepth or the Narea (*i.e.*, either  $\mu_1$ ,  $\mu_3$ , or  $\mu_5$ ). These three last strategies are referred to as the *upper-length*, *upper-Ndepth* and *upper-Narea* selection strategies, respectively. Overall, in the next section, we assess 10 different strategies to set up the observation scale  $f(u)$  at Line 4 of Method 1.

## 6. Experiments

This section aims to compare the results of the original HGB method against the results of the new strategies described in this article. Assessing a hierarchical segmentation, per-se, is a difficult task, a hierarchical segmentation being generally not an end-goal but an intermediary representation from which the information that is of practical interest for a given task is extracted. In order to tackle this problem, we use the framework described in [6]. It considers images and ground-truth from Pascal VOC 2010 dataset [23] (2498 images) and Pascal VOC 2012 dataset [24] (3427 objects segmented from the 1449 images). One of the key propositions of [6] is to assess a hierarchy from different view-points corresponding to different use cases of hierarchies. In particular, two aspects are assessed:

- the quality of the “best” cuts or partitions in  $K$  regions appearing at the same scale (FHC measure) or at different scales (FOC measure) of the hierarchy. A cut is assessed with respect to a human-provided ground truth and to a segmentation quality measure (Bidirectional Consistency Error). The measure is averaged for every  $K$  between 2 and twice the number of regions in the ground truth.

- the ease of finding a set of regions in the hierarchy corresponding to a semantic object when different levels of information are given on the position of the object: an object is considered easy to find when it can be well retrieved with few human-like-provided markers. Here, the retrieved object is a union of regions which, in general, does correspond to any cut of the hierarchy.

The assessment of the first aspect on Pascal VOC 2010 dataset is summarized with a score called FOC+FHC, whereas the evaluation of the second aspect on Pascal VOC 2012 dataset is summarized with the ODM score. More details on the computation of these scores are given in [6]. Furthermore, as recommended in [6], to get assessable results, we perform an area-simplification post-processing [28] of the hierarchies that removes regions smaller than 0.04% of the image size. On a typical hierarchy produced during our experiment, this post-processing allows the number of non-singleton regions to be reduced from 180000 to about 800. For the strategies which depend on a parameter, a search was performed to find the optimal parameter. The tested parameters were:

- 3, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 800, 900, 1000, and 1050, for the lower-area and upper-narea strategies;
- 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, and 70, for the lower-depth and upper-Ndepth strategies;
- 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 75, and 80, for the lower-length and upper-length strategies; and
- all values from 0.01 to 0.4 with an increment step of 0.01 for the lower  $p$ -rank and upper  $p$ -rank strategies.

The scores obtained with the 10 strategies described in Section 5 are presented in Tables 1 and 2. The associated source code and the saliency maps of our results on numerous images are given at <https://cayllahe.github.io/hgbcodes>.

A visual inspection of the resulting saliency maps shows that there is significant variation between the considered strategies. From the numerical results, we can observe that filtering the observation intervals leads to a significant improvement against the raw min- and max-strategies, hence over the original HGB method. More specifically, we observe that:

1. for the best cuts assessment:
  - among the strategies based on min-rule, lower  $p$ -rank obtains the highest score (0.9138);
  - among the strategies based on max-rule, upper  $p$ -rank obtains the highest score (0.9357);
  - max-rule based strategies systematically perform better than their min-rule counterpart;
2. for the supervised object retrieval assessment:
  - among the strategies based on min-rule, lower-area and lower-depth obtain the highest score (0.877);
  - among the strategies based on max-rule, upper-Ndepth obtains the highest score (0.870);
  - min-rule based strategies systematically perform better than their max-rule counterpart.



Table 1: Best cuts assessment results (Pascal VOC 2010 dataset).

Strategy	Param.	FOC	FHC	FOC+FHC
Min	-	0.4917	0.4010	0.8927
Lower-length	3	0.4904	0.3993	0.8897
Lower-area	5	0.4901	0.3988	0.8889
Lower-depth	3	0.4897	0.3986	0.8883
Lower $p$ -rank	0.17	0.5020	0.4117	<b>0.9138</b>
Max	-	0.5184	0.4092	0.9277
Upper-length	35	0.5189	0.4099	0.9288
Upper-Narea	15	0.5194	0.4096	0.9290
Upper-Ndepth	20	0.5189	0.4100	0.9289
Upper $p$ -rank	0.36	0.5193	0.4164	<b>0.9357</b>

Table 2: Supervised object retrieval assessment (Pascal VOC 2012 dataset).

Strategy (min-rule)	ODM	Strategy (max-rule)	ODM
Min	0.868	Max-rule	0.865
Lower-length	0.871	Upper-length	0.866
Lower-area	<b>0.877</b>	Upper-Narea	0.867
Lower-depth	<b>0.877</b>	Upper-Ndepth	<b>0.870</b>
Lower $p$ -rank	0.870	Upper $p$ -rank	0.866

In order to go one step further, we also tested some strategies with different parameters for the area-simplification post-processing leading to the following results:

1. for the best cuts assessment, upper-length is the best strategy (FOC+FHC score: 0.9369); and
2. for the supervised object retrieval assessment, lower-length is the best strategy (ODM score: 0.911).

Overall, we see that there is a clear benefit of filtering the observations scales with the proposed strategies. However, the strategy which leads to the best result depends on the targeted assessment and end-use of the hierarchies.

## 7. Conclusions

This article revisits the HGB method to propose new observation scale selection strategies leading to better segmentation results. We introduce the notion of edge observation attribute and criterion, and establish their non-increasing behavior. Then, we propose an algorithm to compute all scales for which the criterion is true. By analyzing these scales, we propose several strategies to select different observation scales among these values. We assess the performance of our strategies on Pascal VOC 2010 and 2012 datasets showing that these strategies outperform the original HGB method. As future work, we will study the impact of the chosen region dissimilarity measure. The inter- and the within-component differences can be computed based on the properties of each region using, for instance, a statistical approach [29]. Another promising direction consists of replacing the region size term of the Felzenszwalb-Huttenlocher dissimilarity measure by a different measure of the importance of a region, such as compactness or visual saliency [30].

## References

- [1] T. Pavlidis, *Structural Pattern Recognition*, Springer, 1977.
- [2] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, *TPAMI* 33 (5) (2011) 898–916.
- [3] S. Beucher, Watershed, hierarchical segmentation and waterfall algorithm, in: *ISMM*, 1994, pp. 69–76.
- [4] L. Najman, M. Schmitt, Geodesic saliency of watershed contours and hierarchical segmentation, *TPAMI* 18 (12) (1996) 1163–1173.
- [5] J. Cousty, L. Najman, Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts, in: *ISMM*, Vol. 6671, Springer, 2011, pp. 272–283.
- [6] B. Perret, J. Cousty, S. J. Guimarães, D. S. Maia, Evaluation of hierarchical watersheds, *TIP* 27 (4) (2018) 1676–1688.
- [7] P. Salembier, L. Garrido, Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval, *TIP* 9 (4) (2000) 561–576.
- [8] S. Valero, P. Salembier, J. Chanussot, Hyperspectral image representation and processing with binary partition trees, *TIP* 22 (4) (2012) 1430–1443.
- [9] J. F. Randrianasoa, C. Kurtz, E. Desjardin, N. Passat, Binary partition tree construction from multiple features for image segmentation, *PR* 84 (2018) 237–250.
- [10] F. Meyer, P. Maragos, Morphological scale-space representation with levelings, in: *Scale-Space Theories in Computer Vision*, Vol. 1682 of LNCS, 1999, pp. 187–198.
- [11] P. Soille, Constrained connectivity for hierarchical image decomposition and simplification, *TPAMI* (2008) 1132–1145.
- [12] Y. Xu, T. Géraud, L. Najman, Connected filtering on tree-based shape-spaces, *TPAMI* 38 (6) (2016) 1126–1140.
- [13] Y. Haxhimusa, A. Ion, W. G. Kropatsch, Irregular pyramid segmentations with stochastic graph decimation strategies, in: *CIARP*, Vol. 4225 of LNCS, Springer, 2006, pp. 277–286.
- [14] L. Guigues, J. P. Cocquerez, H. Le Men, Scale-sets image analysis, *IJCV* 68 (3) (2006) 289–317.
- [15] J. Pont-Tuset, P. Arbeláez, J. T. Barron, F. Marques, J. Malik, Multiscale combinatorial grouping for image segmentation and object proposal generation, *TPAMI* 39 (1) (2017) 128–140.
- [16] J.-H. Syu, S.-J. Wang, L. Wang, Hierarchical image segmentation based on iterative contraction and merging, *TIP* 26 (5) (2017) 2246–2260.
- [17] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, L. Van Gool, Convolutional oriented boundaries: From image segmentation to high-level tasks, *TPAMI* 40 (4) (2017) 819–833.
- [18] S. Guimarães, Y. Kenmochi, J. Cousty, Z. Patrocínio Jr., L. Najman, Hierarchizing graph-based image segmentation algorithms relying on region dissimilarity - the case of the Felzenszwalb-Huttenlocher method, *MMTA* 2 (1) (2017) 55–75.
- [19] P. F. Felzenszwalb, D. P. Huttenlocher, Efficient graph-based image segmentation, *IJCV* 59 (2) (2004) 167–181.
- [20] P. Salembier, A. Oliveras, L. Garrido, Antiextensive connected operators for image and sequence processing, *TIP* 7 (4) (1998) 555–570.
- [21] P. Salembier, S. Liesegang, Ship detection in sar images based on maxtree representation and graph signal processing, *TGRS* (2018) 1–16.
- [22] P. Salembier, M. H. Wilkinson, Connected operators, *SPM* 26 (6) (2009) 136–157.
- [23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge, *IJCV* 88 (2) (2010) 303–338.
- [24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results, <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
- [25] E. Cayllahua-Cahuina, J. Cousty, S. Guimarães, Y. Kenmochi, G. Cámara-Chávez, A. de A. Araújo, A study of observation scales based on felzenszwalb-huttenlocher dissimilarity measure for hierarchical segmentation, in: *DGCI*, Springer, 2019, pp. 167–179.
- [26] J. Cousty, L. Najman, Y. Kenmochi, S. Guimarães, Hierarchical segmentations with graphs: Quasi-flat zones, minimum spanning trees, and saliency maps, *JMIV* 60 (4) (2018) 479–502.
- [27] E. Cayllahua-Cahuina, J. Cousty, Y. Kenmochi, A. de A. Araújo, G. Cámara-Chavez, S. J. F. Guimarães, Efficient algorithms for hierarchical graph-based segmentation relying on the Felzenszwalb-Huttenlocher dissimilarity, *IJPRAI* 33 (11) (2015) 1–28.
- [28] B. Perret, J. Cousty, S. J. Ferzoli Guimarães, Y. Kenmochi, L. Najman, Removing non-significant regions in hierarchical clustering and segmentation, *PRL* 128 (2019) 433–439.
- [29] S. J. F. Guimarães, Z. K. G. do Patrocínio, Y. Kenmochi, J. Cousty, L. Najman, Hierarchical image segmentation relying on a likelihood ratio test, in: V. Murino, E. Puppo (Eds.), *ICIAP*, Springer, 2015, pp. 25–35.
- [30] G. Li, Y. Yu, Visual saliency based on multiscale deep features, in: *CVPR*, 2015, pp. 5455–5463.