



HAL
open science

Evaluation of hierarchical watersheds

Benjamin Perret, Jean Cousty, Silvio J.F. Guimarães, Deise S Maia

► **To cite this version:**

Benjamin Perret, Jean Cousty, Silvio J.F. Guimarães, Deise S Maia. Evaluation of hierarchical watersheds. 2016. hal-01430865v1

HAL Id: hal-01430865

<https://hal.science/hal-01430865v1>

Preprint submitted on 10 Jan 2017 (v1), last revised 9 Jan 2018 (v4)

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.

Evaluation of hierarchical watersheds.

Benjamin Perret, Jean Cousty, Silvio Jamil F. Guimarães, and Deise S. Maya

Abstract—This article aims to understand the practical features of hierarchies of morphological segmentations, namely the quasi-flat zones hierarchy and watershed hierarchies, and to evaluate their potential in the context of natural image analysis. We propose a novel evaluation framework for hierarchies of partitions designed to capture various aspects of those representations: precision of their regions and contours, possibility to extract high quality horizontal cuts and optimal non-horizontal cuts for image segmentation, and ease of finding a set of regions representing a semantic object. This framework is used to assess and to optimize hierarchies with respect to the possible pre- and post-processing steps. We show that, used in conjunction with a state-of-the-art contour detector, watershed hierarchies are competitive with complex state-of-the-art methods for hierarchy construction. In particular, the proposed framework allows us to identify a watershed hierarchy based on a novel extinction value, the number of parent nodes, that outperforms the other hierarchies of morphological segmentations. This coupled with the fact that watershed hierarchies satisfy clear global optimality properties and can be computed efficiently on large data, make them valuable candidates for various computer vision tasks.

Index Terms—hierarchy of partitions, image segmentation, watershed hierarchy, quasi-flat zones hierarchy, image analysis.



1 INTRODUCTION

HIERARCHIES of partitions are multi-scale image representations that were first proposed in [1], [2]. They have since appeared under various names: pyramids, hierarchy of segmentations, partition trees, scale-sets. In a hierarchy (of partitions), an image is represented as a sequence of coarse to fine partitions satisfying the strong causality principle [3], [4]: *i.e.*, any partition is a refinement of the previous one in the sequence. They have various applications in image processing and analysis: image segmentation [5], [6], [7], [8], [9], [10], occlusion boundary detection [11], image simplification [6], [9], [12], object detection [5], objects proposal [10], visual saliency estimation [13]. In particular, they have gained a large popularity with the works of [7] whose hierarchical approach to the general problem of natural image segmentation has outperformed state of the art approaches.

It has long been noted [16], [17], [18] that the classical morphological approach to image segmentation, *i.e.*, the watershed, is compliant with the strong causality principle. This enables to define hierarchies of watersheds (see Figure 1) as a sequence of watershed segmentations of an image whose minima are iteratively removed according to an importance measure, *e.g.*, related to their sizes. This definition has been formalized in the context of minimum spanning forests that already enabled to define watershed cuts as an optimal solution to a combinatorial problem related to minimum spanning tree [19]. It has also been shown that hierarchies of watersheds are linked to the quasi-flat zones hierarchies [20], to the single-linkage clustering problem [21], and to connective segmentation [12], [22].

Hierarchies of watersheds are thus multi-scale representations which satisfies a global optimality property. Moreover, there exist efficient algorithms, with the same time

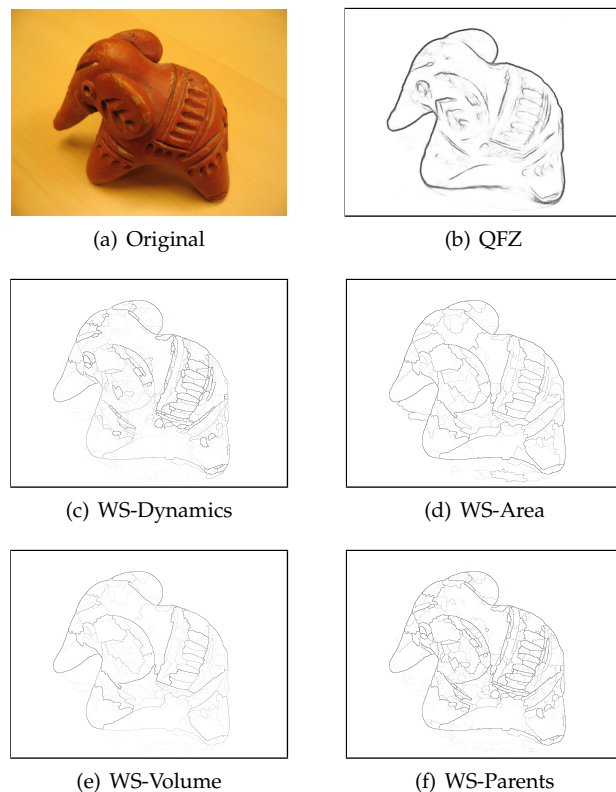


Figure 1. Saliency maps of the morphological hierarchies studied in this article –QFZ, WS-Dynamics, WS-Area, WS-Volume, and WS-Parents– on the elephant image from Grabcut dataset [14] with SED gradient [15].

complexity as minimum spanning tree algorithms, to construct them [23], [24] enabling to process large images in real time. In recent years, they have been used for the computation of morphological operators [25], in the context of stochastic watershed segmentation [26], [27], [28]. However, their practical performances have not yet been studied in the general case of natural image analysis and the aim of

- B. Perret, J. Cousty, and D. S. Maia are with Université Paris-Est, LIGM, ESIEE - UPEMLV - CNRS.
E-mail: {benjamin.perret,jean.cousty,deise.santanamaia}@esiee.fr
- S.J.F. Guimarães is with PUC Minas - ICEI - DCC - VIPLAB.
E-mail: sjamil@pucminas.br

this work is to understand their practical features and to evaluate their potential in this context.

To this end, we propose a novel evaluation framework for hierarchies of partitions specifically designed to capture the various aspects of those representations: 1) quality of regions and contours, 2) quality of produced segmentations with horizontal cuts and optimal cuts, and 3) easiness of finding a set of regions representing a semantic object. These measures are evaluated on two types of natural image datasets: 1) the Berkley segmentation dataset BSDS 500 [7], [29], and 2) Grabcut [14] and Weizmann [30] object detection and segmentation datasets. Compared to the classical approach for hierarchy evaluation that focuses only on the horizontal cuts and the image segmentation problem, we believe that the proposed framework offers a richer assessment that better accounts for the hierarchical nature of the representations and is not limited to a single use case.

This framework is used to evaluate and understand the strengths and weaknesses of the considered hierarchies of morphological segmentations. In particular, it allows us to identify a watershed hierarchy based on a novel extinction value, the number of parent nodes, that outperforms the other hierarchies of morphological segmentations. Then, we study the importance of the gradient measure for all methods and the necessity to perform a filtering of some hierarchies. Finally, the properties of the best found solutions are discussed and compared to a state of the art approach.

The definition of quasi-flat zones and watershed hierarchies are given in Section 2. Existing evaluation methods for hierarchies are discussed in Section 3. Section 4 presents the evaluation framework and the new measures. The experiments and their outcomes are discussed in Section 5. The work is finally concluded in Section 6.

2 PRELIMINARY ON GRAPHS AND HIERARCHIES

In this section, we first review the definitions of graphs and hierarchies of partitions. Then, we recall the definition of the hierarchies of morphological segmentations used in this article, namely the quasi-flat zones hierarchies and the watershed hierarchies.

2.1 Graphs and hierarchies

In the sequel of this paper, the *graph* \mathcal{G} is defined as a pair (V, E) where V is a finite set and E is composed of unordered pairs of distinct elements in V , *i.e.*, E is a subset of $\{\{x, y\} \subseteq V \mid x \neq y\}$. Each element of V is called a *vertex* or a *pixel*, and each element of E is called an *edge*. The graph \mathcal{G} will model the image spatial domain, *e.g.*, V is the regular 2D grid of pixels, and E is the 4 or 8 adjacency.

We denote by W a function from E to \mathbb{R} that weights the edges of \mathcal{G} . Therefore, the pair (\mathcal{G}, W) is an *edge-weighted graph*, and, for any $u \in E$, the value $W(u)$ is the *weight* of u .

A *partition*, also called a *segmentation*, \mathbf{P} of V is a family of subsets of V such that: 1) the intersection of any two distinct elements of \mathbf{P} is empty, and 2) the union of the elements of \mathbf{P} is equal to V . Each element of a partition \mathbf{P} is called a *region* of the partition \mathbf{P} . Given two partitions \mathbf{P}_1 and \mathbf{P}_2 , we say that \mathbf{P}_2 is a *refinement* of \mathbf{P}_1 if every region of \mathbf{P}_2 is included in a region of \mathbf{P}_1 .

A *hierarchy (of partitions)* $\mathcal{H} = (\mathbf{P}_0, \dots, \mathbf{P}_n)$ is a sequence of partitions of V such that \mathbf{P}_0 is the single region partition $\mathbf{P}_0 = \{V\}$, the partition \mathbf{P}_n contains every singletons of V , *i.e.*, $\mathbf{P}_n = \{\{x\} \mid x \in V\}$, and \mathbf{P}_i is a refinement of \mathbf{P}_{i-1} for all i in $\{1, \dots, n\}$ (see Figure 2).

Given a hierarchy $\mathcal{H} = (\mathbf{P}_0, \dots, \mathbf{P}_n)$, the set of *regions of \mathcal{H}* , denoted by $\mathbf{R}_{\mathcal{H}}$ is the union of all partitions of \mathcal{H} . The inclusion relation on $\mathbf{R}_{\mathcal{H}}$ induces a tree structure (or a dendrogram) where: V is the root, the singletons $\{x\}$ with $x \in V$ are the leaves, and the parent of a region $R \neq V$ of $\mathbf{R}_{\mathcal{H}}$, denoted by $Parent(R)$, is the smallest region R' of $\mathbf{R}_{\mathcal{H}}$ that is strictly larger than R (see Figure 2).

Given a hierarchy $\mathcal{H} = (\mathbf{P}_0, \dots, \mathbf{P}_n)$, a partition \mathbf{P} of V made of regions of \mathcal{H} (*i.e.*, $\mathbf{P} \subseteq \mathbf{R}_{\mathcal{H}}$) is called a *cut of \mathcal{H}* (see Figure 2). The set of all cuts of a hierarchy \mathcal{H} is denoted by $\Pi(\mathcal{H})$. A cut \mathbf{P} is said *horizontal* if $\mathbf{P} = \mathbf{P}_i$ for some i in $\{1, \dots, n\}$.

2.2 Quasi-flat zones hierarchy

The quasi-flat zones have been studied since the 70's (see *e.g.*, [12], [31], [32]). They are deeply linked to single-linkage clustering and to the notion of a minimum spanning tree [21]. A *quasi-flat zone* of the weighted graph (\mathcal{G}, W) at level λ is a maximal set of vertices such that there exists a path of maximal weight λ between any two of its vertices. The quasi-flat zones of the weighted graph at a given level λ is a partition of its vertices. The sequence of partitions obtained for all possible values of λ is a hierarchy called the *quasi-flat zones hierarchy* of the weighted graph (see Figure 1), and denoted by QFZ.

2.3 Watershed hierarchies

Watershed hierarchies were first proposed in [16], [17], [18] and have since been formalized in the context of minimum spanning forests [19], [23]. Given a weighted graph and a family of markers (*i.e.*, subsets of the graph vertices identifying the objects of interest), the problem of minimum spanning forest is to find a spanning forest of minimum total weight, defined as the sum of the weights of its edges, such that each connected component of the forest contains (is rooted in) exactly one marker. The connected components of the minimum spanning forest then forms a segmentation with a global optimality property similar to the one of the minimal spanning tree. When the markers are the regional minima of the weight map, the corresponding minimum spanning forest segmentations are indeed the watershed segmentations defined by the drop of water principle [19].

If the markers are ranked, *e.g.*, according to an importance measure, it is possible to obtain a sequence of nested minimum spanning forests such that the k -th minimum spanning forest is rooted in the k -most important markers. Thus, one can obtain a sequence of nested partitions, hence a hierarchy of partitions as defined in this article, where every partition is optimal. A usual choice to define a sequence of markers is to rank the minima of the weight map according to extinction values [33]. Such hierarchies are called hierarchical watersheds; their theoretical properties and some algorithms to construct them are studied in [20], [23], [24].

Extinction values are defined through regional attributes defined on the connected components of the level sets of

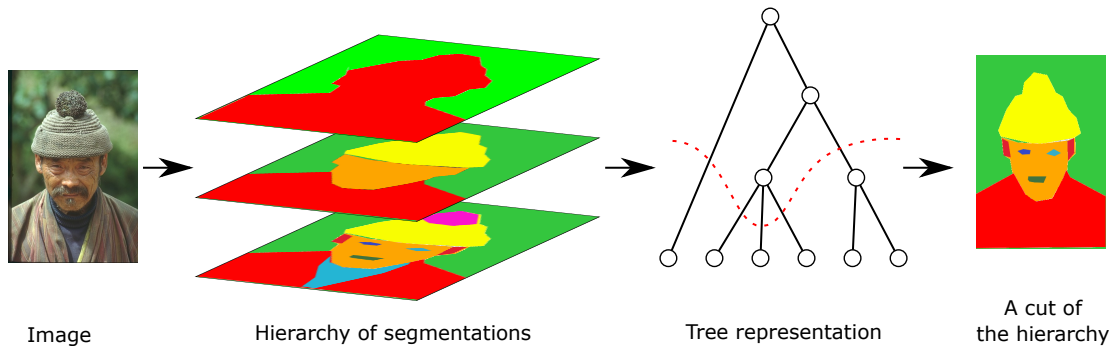


Figure 2. A hierarchy of partitions of an image (source BSDS 500 [7]) is a sequence of coarse to fine partitions. The hierarchy can be represented as a tree of regions. A cut is a partition made of regions of the hierarchy possibly taken at different levels (red dashed curve).

the weight map [33]. Intuitively, the extinction value of a minima m for a given regional attribute is the smallest value λ_m such that the minimum m disappears when all components with an attribute smaller than λ_m are removed. Common regional attributes are related to the size and to the contrast of the components [33], [34]: *e.g.*, dynamics, area, or volume. Other authors have proposed regional measures related to topological properties of the function inside the components: topological height [35], or number of descendants [35]. We propose to use a novel attribute counting the number of parent nodes in the min-tree of the weighted graph, *i.e.*, the number of non leaf nodes among the descendent of a node. All these regional attributes and their associated extinction values can be computed from the quasi-flat zones hierarchy of the weighted graph [24]. Among all the possibilities we have chosen to present the results of (see Figure 1):

- area, volume, and dynamics which are the most widely presented measures in the literature;
- number of parent nodes, which is up to our knowledge a new proposal, that provides the best performances in the following assessments.

Those regional attributes lead to the hierarchies denoted respectively by WS-Area, WS-Volume, WS-Dynamics, and WS-Parents in the following of this manuscript.

3 EVALUATION OF HIERARCHIES

This section reviews the different solutions proposed in the literature for the evaluation of hierarchies. The evaluation of hierarchies is subject to two major difficulties: 1) they are complex structures often leading to large combinatorial problems, 2) as an intermediate tool, they have various usage and one hierarchy may be adapted for some tasks but not for others.

A qualitative assessment of hierarchies can be performed through a visual inspection of the saliency maps [17], *i.e.*, image of contours where the brightness of a region contour is proportional to the scale of the region. Figure 1 shows saliency maps obtained with the quasi-flat zones hierarchy and the considered watershed hierarchies. Such exercise is however not trivial and may even be misleading as the general impression may be largely influenced by the transfer function used to convert the scale measure (that may have a large dynamic range) to a viewable image.

Quantitative assessment of hierarchies focus on the evaluation of the individual segmentations that can be extracted from a hierarchy. This approach enables to reuse the existing image datasets with their ground-truth segmentations and to benefit from the existing works on dissimilarity measures between segmentations.

The most popular approach to evaluate a hierarchy, developed by [7], consists in comparing each partition in the sequence of partitions defining the hierarchy (the horizontal cuts of the hierarchy) to the ground-truth. When the comparison measure produces precision and recall scores, their evaluation along the sequence of partitions produces the so-called precision-recall curves. To evaluate a hierarchy on a whole dataset, two aggregated measures are then defined: the optimal image scale OIS measuring the best achievable score when taking the optimal horizontal cut in each hierarchy, and the optimal data-set scale ODS measuring the best achievable score when taking horizontal cuts at the same level (the optimal scale) in every hierarchy. The difference between the ODS and the OIS measures assesses the consistency of the hierarchy in terms of scale: close OIS and ODS values suggest that regions of equivalent perceptual importance in different images are represented at the same level of their respective hierarchies.

This framework has been applied with three different measures: 1) F-Measure for regions (FR) [36] where image segmentation is viewed as a multi-class clustering problem on the image pixels, 2) F-Measure for boundaries (FB) [36], [37], [38] where image segmentation is viewed as a binary clustering problem on the pixels' boundaries, and 3) F-Measure for objects and parts (FOP) [39] which defines empirical (pseudo) precision, recall based on the heuristic classification of each region of the partitions as an object, a part of an object, or noise. The work of [39] on the evaluation of segmentation assessment measures has shown that FB and FOP are highly discriminant between ground truths of different images on the BSDS 500 image dataset [7]. On the contrary, FR has shown a low discriminant power.

The horizontal cuts considered in that framework represent a subset of all possible partitions that can be constructed from a hierarchy. In order to better evaluate the potential of hierarchies the authors of [40], [41] proposed to look for the optimal cut, generally not horizontal, in a hierarchy according to a given evaluation measure. This leads to combinatorial optimization problems that have been solved

in the two following cases: 1) upper bound on FB solved as a linear fractional combinatorial optimization problem [40], and 2) upper bound on local additive region measures solved with dynamic programming [41] (similar to finding the optimal partition for a given energy function [6], [42]). Those methods does not provide any tool to extract this upper bound cut when the ground truth is unknown: they only measure the full potential of a hierarchy.

Up to our knowledge Maire et al. [43] is the single attempt to provide a hierarchical ground-truths datasets. They defined a hierarchical ontology of semantic objects with 3 levels and asked human subjects to decompose scene according to it: their dataset is thus strongly focused on the category identified in the ontology and does not corresponds to a general segmentation objective.

4 PROPOSED EVALUATION METHODOLOGY

In this section, we present an evaluation framework for hierarchies of partitions. This framework is composed of several supervised assessment measures, each enabling to quantify a different aspect of the hierarchy. First, we explain the choice of the assessment measures and their contributions. Then, we give a detailed description of each measure.

We choose to focus on natural image processing and more precisely on image/object segmentation. Our assessment methodology comprises three parts:

- 1) We propose an evolution of the upper-bound on region measures [41] enabling to quantify the maximal achievable score of a hierarchy for the general segmentation problem. This segmentation ability is studied with respect to the number of desired regions in the target segmentation, allowing us to identify, in the context of hierarchies, some classical properties of image segmentation such as under- and over-segmentation;
- 2) We propose new evaluation measures that aims to quantify the easiness of finding a set of nodes of a hierarchy representing a semantic object in the scene;
- 3) As a standard evaluation measure, we use the F-Measure and precision-recall curves on boundaries FB [36] (see Section 3) that quantify the contours quality in the horizontal cuts of a hierarchy and the consistency of their indexing. This measure is complementary to the other region oriented measures and also provides a reference measure for the comparison with the literature..

4.1 Upper-bound on BCE measure

We propose an evolution of the upper-bound evaluation on regions proposed by Pont-Tuset et al. [41] that consists in two improvements: 1) the use of a dissimilarity measure that enables to penalize both under- and over-segmentation, and 2) the definition of a new type of curve, the fragmentation-upper bound curve that enable to measure the potential of the hierarchy and the potential gain of non-horizontal cuts compared to horizontal cuts.

In [41], the authors focused on the directional Hamming distance [44] which is transparent to over-segmentation, *i.e.*, it does not penalize the subdivision of a region of the ground-truth into multiple regions in the proposal segmentation. In this work, we propose to use the Bidirectional

Consistency Error *BCE* [36] measure. The *BCE* measure is symmetric and it is not transparent to over- or under-segmentation. The evaluation of segmentation measures provided by [39] evaluates BCE as a highly discriminant measure on the image segmentation dataset BSDS 500 [7].

Given an image I , one ground-truth segmentation T_I , and a proposal segmentation S_I , the *BCE measure* of S_I and T_I is defined by [36]:

$$BCE(S_I, T_I) = \frac{1}{N} \sum_{\substack{R \in S_I \\ R' \in T_I}} |R \cap R'| \min \left(\frac{|R \cap R'|}{|R|}, \frac{|R \cap R'|}{|R'|} \right). \quad (1)$$

Given a hierarchy of partitions \mathcal{H}_I on the image I , one ground-truth segmentation T_I and a number k of regions, the *Upper-Bound BCE score* (UB_{BCE}) for \mathcal{H}_I is the highest BCE score for all the cuts of \mathcal{H}_I composed of k regions:

$$UB_{BCE}(\mathcal{H}_I, T_I, k) = \max_{\substack{S \in \Pi(\mathcal{H}_I) \\ |S|=k}} BCE(S, T_I). \quad (2)$$

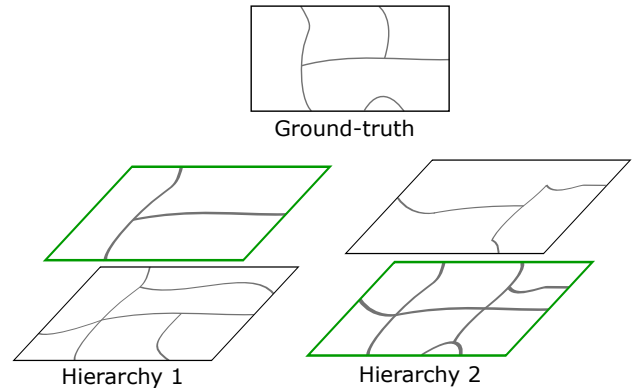


Figure 3. Illustration of under- and over-segmentation for hierarchies. Hierarchies 1 and 2 are both composed of 2 levels. Compared to the ground-truth, the first hierarchy manages to recover long contours in its coarse level but then fails to recover the other contours at a finer level: the optimal horizontal cut is the coarsest one and the hierarchy is said to under-segment the image. With the second hierarchy the inverse situation happens, the coarsest partition recovers all the contours of the ground-truth but also contains extra-contours. However, the finest partition loses the true contours and preserves extra contours: the hierarchy is said to over-segment the image.

In order to better understand the content of the hierarchies and to account for the variations inside the evaluation datasets, we propose the *Fragmentation-Optimal Cut score curve* (FOC) where the mean-average Upper-Bound BCE score (the mean image score over the database, with the image score defined as the average score over the set of ground-truths for the image) is plotted against the *fragmentation* level of the segmentation defined as $k/|T_I|$, the ratio between the number of regions in the segmentation and the number of regions in the ground-truth (see Figure 5). The gain achieved by taking a non horizontal cut in the hierarchy is evaluated with a second curve: the *Fragmentation-Horizontal Cut score curve* (FHC) obtained by taking the successive partitions of the hierarchy (similarly to precision-recall curves). A large difference between, the FOC and FHC curves suggests that the optimization algorithm has selected regions from various levels of the hierarchy to find the

optimal cut: the regions of the ground-truth segmentations are thus spread at different levels in the hierarchy.

The FOC curve starts at the value corresponding to the single region partitions (independent of the evaluated hierarchy). Then, it generally quickly increases at low fragmentation levels as the optimization first selects the largest regions that summarize the ground-truth. Then, the optimal cut starts to include smaller regions that provides only little score gain: this corresponds to the nearly flat part of the curve. At a high level of fragmentation (not visible in the figures), the algorithm cannot add new regions without lowering the score and the curve starts to decrease.

In the ideal case, the maximum of the FOC and FHC curves is achieved for a fragmentation of 1. If the maximum appends at fragmentation level lower than 1, this means that the hierarchy tends to capture the main feature of the ground-truth with a low number of regions but then fails to correctly refine those regions (see *Hierarchy 1* in Figure 3): in this case we say that the hierarchy has a *tendency for under-segmentation*. If the maximum appends at a higher fragmentation level than 1, this means that the hierarchy is able to provide a set of superpixels for the ground-truth but fails to merge them in a correct order (see *Hierarchy 2* in Figure 3): in this case we say that the hierarchy has a *tendency for over-segmentation*.

As an overall performance summary respectively on the FOC and FHC curves, we compute the normalized area under the curve, denoted respectively by AUC-FOC and AUC-FHC. The area under the curve provides an evaluation over a large range of fragmentation levels and thus accounts for the hierarchical nature of the object of study. In order to obtain a measure that is symmetric between under- and over- fragmentation, we choose to calculate the area under the curve on the interval $]0, 2]$. Finally, the area under the curve is normalized with a factor $1/2$ to obtain a score between 0 (worst) and 1 (best).

4.2 Object detection measure

The last measure, introduced in our previous work [45], is based on supervised object detection with markers. It quantifies how well a specific object of a scene can be retrieved with different levels of information given on its position.

In this evaluation we have chosen to use the procedure described in [5] that constructs a two classes segmentation from a hierarchy of partitions and two non-empty markers: one for the background and one for the object of interest. Its principle is to identify the object as the union of the regions of the hierarchy that intersect the object marker but does not touch the background marker. Formally, given an image I , a hierarchy \mathcal{H}_I , an object marker M_o , and a background marker M_b , the extracted object is defined by:

$$O(\mathcal{H}_I, M_o, M_b) = \bigcup \{R \in \mathbf{R}_{\mathcal{H}_I} \mid R \cap M_o \neq \emptyset, R \cap M_b = \emptyset\}. \quad (3)$$

This result can be computed efficiently with Algorithm 1. In the first step of the algorithm, the hierarchy is browsed from the leaves to the root. If the current node is labeled *Background* then its parent node intersects the background marker and is labeled *Background*. If the current node is labeled *Object* and its parent is not currently labeled then it

Algorithm 1: Marker based object detection.

Data: A hierarchy of partition \mathcal{H} on the graph \mathcal{G} .

Data: The labeling ℓ on the leaves of \mathcal{H} s.t. $\forall x \in V$, $\ell(\{x\}) \in \{Object, Background, Undefined\}$

Result: The segmented object O according to Eq. (3).

```

// For all nodes from leaves to root
1 for all regions R in  $\mathbf{R}_{\mathcal{H}}$  in increasing order do
2   if  $R \neq V$  then // R is not the root
3     if  $\ell(R) = Background$  then
4       |  $\ell(Parent(R)) \leftarrow Background$ ;
5     else if  $\ell(R) = Object$  and
6       |  $\ell(Parent(R)) = Undefined$  then
7         |  $\ell(Parent(R)) \leftarrow Object$ ;
// For all nodes from root to leaves
7 for all regions R in  $\mathbf{R}_{\mathcal{H}}$  in decreasing order do
8   if  $R \neq V$  and  $\ell(R) = Undefined$  then
9     |  $\ell(R) \leftarrow \ell(Parent(R))$ ;
10  $O \leftarrow \{x \in V \mid \ell(\{x\}) = Object\}$ 

```

can be labeled *Object*. In the second step, the tree is browsed from the root to the leaves and any non labeled node takes the label of its parent. Finally, the labels of the leaves (the image pixels) give the segmentation result.

In order to perform an objective assessment of the different hierarchies we propose several automatic strategies to generate object and background markers from the ground truths. Our main idea is not to reproduce the interactive segmentation process experienced by a real user but rather to obtain markers representing different difficulty levels or that resembles to human generated markers. The generated markers are the following (see Fig. 4): 1) Erosion (Er): erosion by a ball of radius 45 pixels. If a connected component is completely deleted by the erosion then a single point located in the ultimate erosion of this connected component is added to the marker, 2) Skeleton (Sk): morphological skeleton given by [46], and 3) Frame (Fr): frame of the image minus the object ground truth if the object touches the frame (background only). Using the frame as the background marker is nearly equivalent to having no background marker in the sense that it does not depend of the ground truth or of the image.

In the following, the combination of the background marker MB and the object marker MF is denoted MB-MF (for example, Fr-Sk stands for the combination of a Frame marker for the background and a skeleton marker for the object). Among all the possible combinations of markers, we chose to concentrate on the following ones: 1) Sk-Sk resembles to human generated markers, 2) Er-Er leaves a large space between markers and represent a difficult case. Nevertheless, the combination is symmetric in the sense that the correct segmentation is roughly at equal distance from the object and from the background marker, and 3) Fr-Sk where the object marker resembles to a human generated marker and the background marker conveys nearly no information: this case is thus strongly asymmetric.

The performance of each segmentation result is evaluated with the F-Measure. The median score for the 3

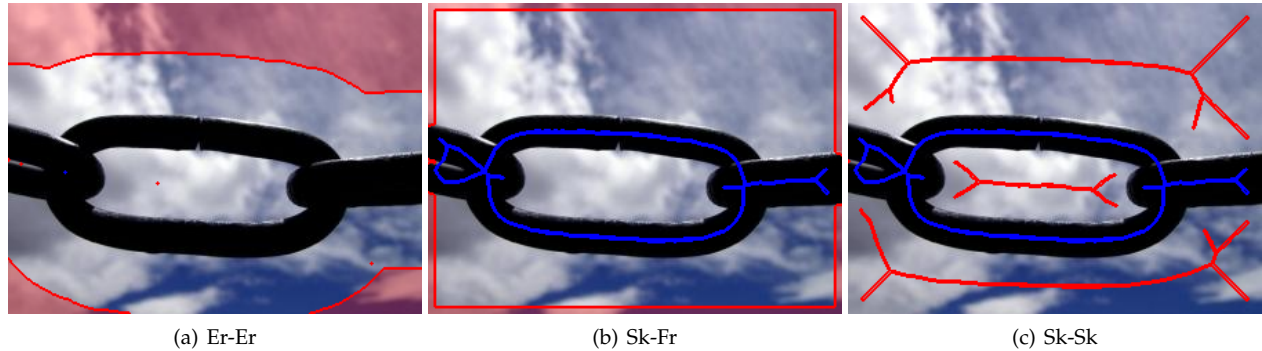


Figure 4. Different combinations of markers. The combination of markers is indicated in the caption of each sub-figure in the form Background Marker-Object Marker. In each figure the background and object markers are respectively depicted in red and blue.

marker combinations, is called *Object Detection Median* and is denoted ODM.

5 EXPERIMENTS

This section presents the results of the experiments and some discussions.

Precision-recall curves for boundaries and upper-bound on BCE measure are evaluated on the BSDS 500 dataset [7], [29] (200 test images). The object detection measure is evaluated on the Grabcut [14] and Weizmann 1 object [30] datasets (respectively 50 and 100 test images). We study the importance of the gradient measure for all methods (Section 5.1) and the necessity to perform a filtering of some hierarchies (Section 5.2). The overall results are discussed and compared to a state of the art approach (Section 5.3).

5.1 Influence of gradient

A classical way to weight the edges of a graph in image analysis in general and for morphological segmentation in particular is to use a gradient measure. The aim of this section is to evaluate the influence of the gradient measure on the quality of the hierarchies.

The most simple gradient measures use only colorimetric information from the two pixels of an edge: in this category, we consider an Euclidean distance in the RGB color space and an Euclidean distance in the Lab color space, the latter being more compliant with human color perception. However, recent advances on contour detection have lead to non local supervised gradient estimators achieving better performance on contour detection benchmarks: in this category, we consider the globalized probability of boundary (gPb) from [7] and the structured edge detector (SED) from [15].

Figure 5 shows the result of WS-Dynamics (top row) and WS-Area (bottom row) with the four considered gradients –RGB, Lab, gPb, and SED–. The results of QFZ (respectively WS-Volume and WS-Parents), not shown here, are similar to the results of WS-Dynamics (respectively WS-Area). A first observation on WS-Dynamics with RGB and Lab gradients is that its PR-Curves on boundaries seem truncated and its FOS-Curves on regions are flat. In the first case, the truncation appears at the level where the partition of the hierarchy contains more than 3000 regions: the evaluation procedure is stopped at this point as it becomes too demanding on

computational power. In the second case, the flat curve is the result of the hierarchy not being able to provide any meaningful partition with at most twice the number of regions in the ground-truth. Those two observations can be a consequence of WS-Dynamics (and similarly QFZ) having its upper levels made only of small salient regions; a solution to this problem is presented in the next section.

While the Lab gradient provides slightly better performance compared to RGB gradient in most cases, we observe a large gain by switching from a local RGB or Lab gradient to a supervised non-local gradient like gPb or SED. The SED gradient improve the results for every measure except the FOC curve with WS-Dynamics compared to gPb gradient. The FOC and FHC curves show that WS-Dynamics requires much more regions to reach its maximal scores with SED gradients which implies that the hierarchy tends to have small irrelevant regions on its top layers. This suggests that despite the regularization effect based on dynamics, which tends to send lowly contrasted regions to the lower levels of the hierarchy, WS-Dynamics remains sensitive to small regions of high contrast that appear more often in SED gradients than in gPb gradients.

In conclusion, we recommend the use of SED gradient to build watershed hierarchies on natural images and the following experiments will be conducted with this gradient. Moreover, SED is about 3 orders of magnitude faster than gPb [15] enabling to reach real time performance without any particular material.

5.2 Small regions removal

As observed in the previous section, QFZ and WS-Dynamics are sensitive to small regions even with a smooth gradient as SED. In this section we evaluate the impact of an area post-filtering on those hierarchies.

The area filter described in [47] removes contours iteratively in the hierarchy: starting from the leaves and moving toward the root, the children of a node are merged if at least one of them contains less than k pixels. In the following, we express the strength of the filter as the ratio $r_k = k/N$, with N the number of pixels in the considered image.

Figure 6 shows the result of the filtering on QFZ (the results on WS-Dynamics are similar) with four different values of r_k : 0 (no filter), 0.4‰ (roughly 50 pixels in a BSDS 500 image [7]), 0.8‰, and 1.6‰. We observe that all

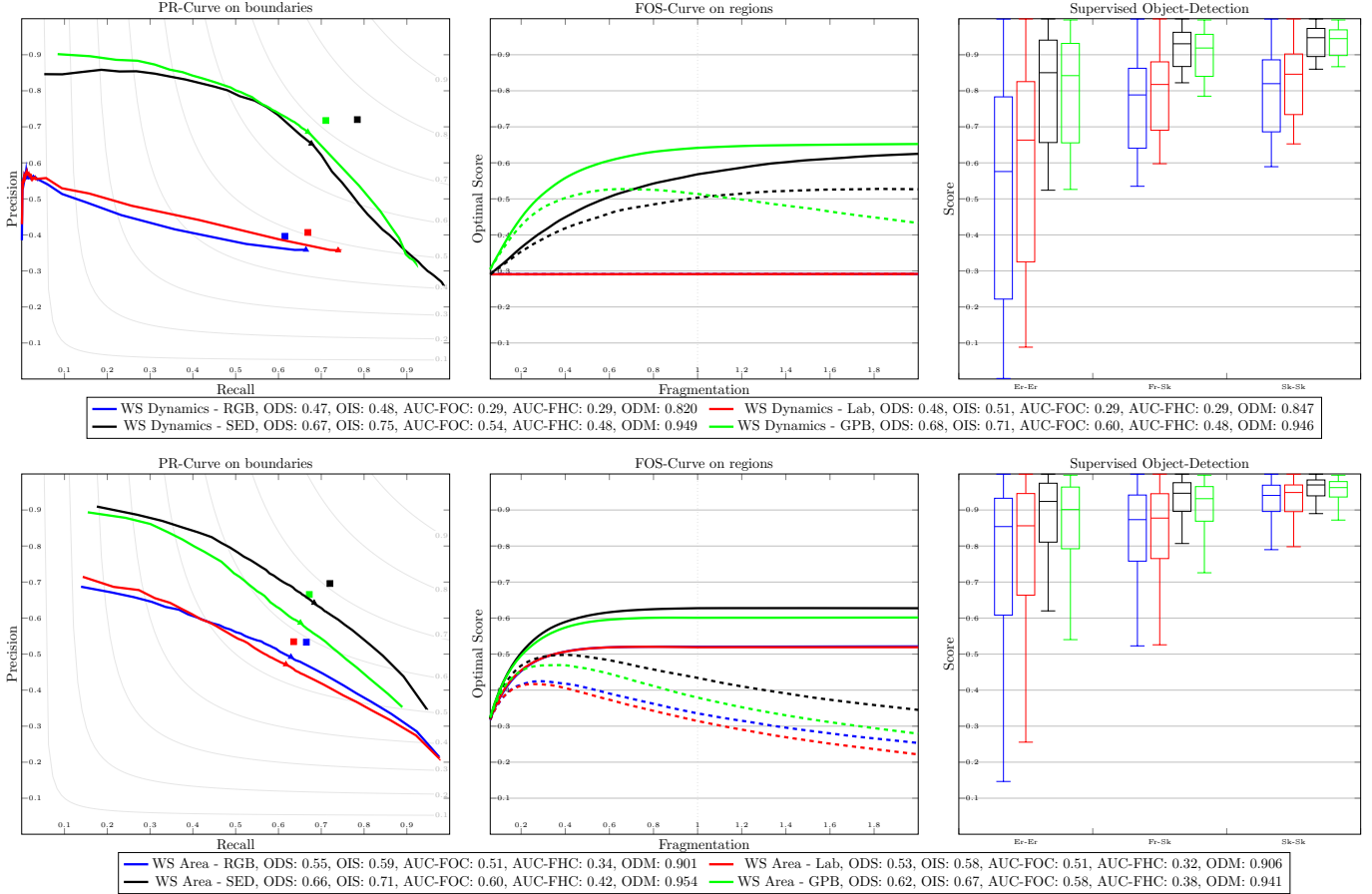


Figure 5. Influence of the gradient on WS-Dynamics (top row) and WS-Area hierarchies (bottom row). **Mean Precision-recall (PR) curves** for boundaries on BSDS 500: each curve represents the variation of precision and recall for the different partitions of the hierarchy. OIS and ODS scores are given in the legend and are respectively represented in the plot by a square and a triangle. **Fragmentation-Optimal Cut score curves (FOC)** for regions on BSDS 500: each plain curve represent the upper-bound score achievable for a given fragmentation value. The corresponding dashed curves represent the score obtained by horizontal cuts. Area under curve for the plain curve (AUC-FOC) and the dashed curve (AUC-FHC) are given in the legend. **Supervised object detection** on Grabcut and Weizmann datasets: for each method and each combination of markers, we see: 1) the median F-measure (central bar), 2) the first and third quartile (extremities of the box), and 3) the lowest datum still within 1.5 inter quartile range (difference between the third and first quartile, and the highest datum still within 1.5 inter quartile range of the upper quartile range (lower and upper extremities). The median score over all markers combinations is given in the legend.

measures increase with r_k , from $r_k = 0$ to $r_k = 0.8\%$. The introduction of the filtering immediately produces a large performance boost. For $r_k = 1.6\%$ compared to $k = 0.8\%$, the situation is mixed with an improvement on FOC measures, stagnation on objection detection measures, but a degradation of OIS and ODS scores: this reflects a tradeoff between the number of regions necessary to describe the scene and the precision of boundaries.

The effect of the filtering on under- and over-segmentation is presented in Figure 7. For each image and each ground-truth of the dataset, we plot the number of regions present in the optimal segmentation found against the number of regions in the ground-truth (we define the optimal segmentation as the segmentation that achieves 99% of the optimal score with the fewest number of regions in the FOC curve). We see that for WS-Dynamics (results are similar for QFZ), larger values of r_k tends to push the optimal segmentation from over-segmentation (position above the diagonal where the optimal segmentation contains more regions than the ground-truth) to under-segmentation (position below the diagonal where the optimal segmentation

contains less regions than the ground-truth). For $r_k = 0.8\%$, the optimal solutions have a mostly symmetrical distribution around the diagonal, suggesting no bias toward under- or over-segmentation.

In conclusion, for QFZ and WS-Dynamics we recommend to perform a post-filtering of the hierarchy by removing regions smaller than 0.8% of the image size. One can notice that the object detection measure is less sensitive to the area filtering than other evaluation measures. This suggests that, for some applications, the processing of the hierarchy is naturally robust to small nodes and this filtering may not be necessary.

5.3 Discussions

This section compares and discuss the best results obtained for each hierarchy (see Figure 8).

As a reference state of the art result we also include the Multiscale Combinatorial Grouping (MCG) hierarchies from [10] in our assessments. MCG also uses SED as the main cue for contour detection, but then merges several

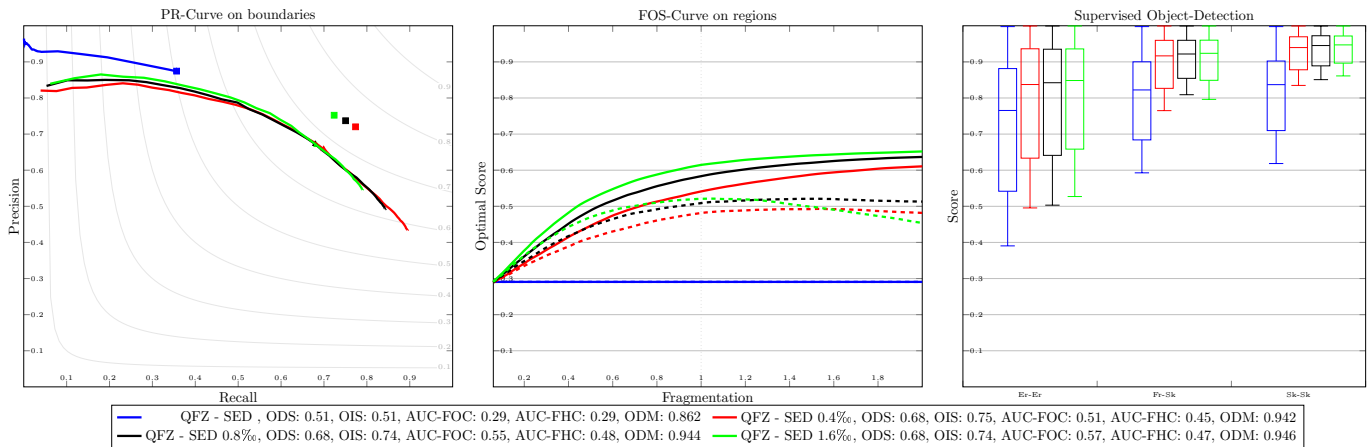


Figure 6. Influence of the area filter on QFZ. (See Figure 5 for explanation).

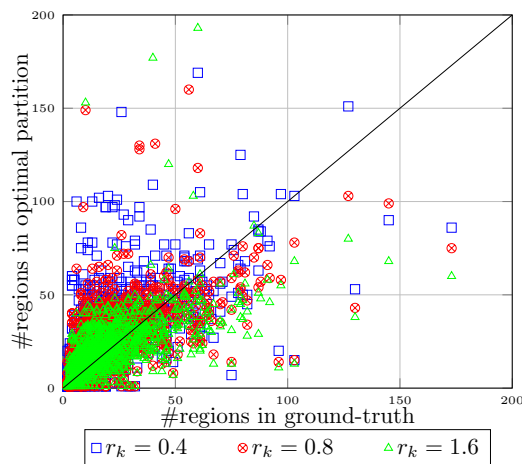


Figure 7. Influence of area filtering on WS-Dynamics.

hierarchies (referred as OWT-UCM in the literature [7]) computed at different scales.

We can observe that QFZ is globally inferior to all the other methods. WS-Dynamics shows good performances in precise contour placement (high ODS and OIS scores) but has a clear tendency for over-segmentation (maximum of FOC and FHC curve occur at large values of fragmentation). We can also notice that WS-Dynamics (and QFZ) performances for object detections in the Sk-Sk case is significantly lower than other methods; this suggests that the hierarchy fails to correctly order regions near the boundaries which is coherent with its tendency to over-segment (more regions are needed to obtain the true contours).

On the contrary, WS-Area and WS-Volume show weaker performance at contour location (average OIS and ODS scores) and have a tendency for under-segmentation (maximum of FOC and FHC curve occur at low values of fragmentation). WS-Area shows a clear advantage over other methods on object detection with Er-Er markers which can be explained by the symmetric nature of the markers in this case: the true contour is located roughly at equal distance from both markers and WS-Area is particularly good at

producing a regular (in size) tiling of the contour image.

WS-Parents offers the best performances on every measure expect OIS compared to other watershed hierarchies. As WS-Area and WS-Volume, it shows a small tendency for under-segmentation. The results of MCG remain higher than WS-Parents except for the object detection assessment: this suggests that MCG, whose various components have been either trained or optimized on BSDS 500 dataset, may over-fit this particular dataset and not be the best method for other applications than general segmentation.

6 CONCLUSION

We have proposed a novel evaluation framework for the evaluation of hierarchies of partitions that enables to capture the quality of different aspects of the hierarchies: regions, contours, horizontal cuts, optimal cuts, nodes grouping, under or over-segmentation. Compared to the classical approach for hierarchy evaluation that concentrates only on the horizontal cuts and the image segmentation problem, we believe that the proposed framework offers a richer assessment that better accounts for the hierarchical nature of the representation and is not limited to a single use case.

This framework was used to assess various hierarchies of morphological segmentations. In particular, we studied the importance of the gradient measure for all methods and the necessity to perform a filtering of some hierarchies. The framework also allowed us to identify a watershed hierarchy based on a novel extinction value, the number of parent nodes, that outperforms the other hierarchies of morphological segmentations. We have shown that, used in conjunction with a state-of-the-art contour detector, most watershed hierarchies are competitive or even sometimes better than the complex state of the art method for hierarchy construction. Moreover, watershed hierarchies are well defined structure satisfying clear global optimality properties and can be computed efficiently on large data: they are thus valuable candidates for various computer vision tasks.

All the programs used to compute the hierarchies and compute the evaluation measures (and their source code) are available online at <http://www.esiee.fr/~perretb/supeval.html>.

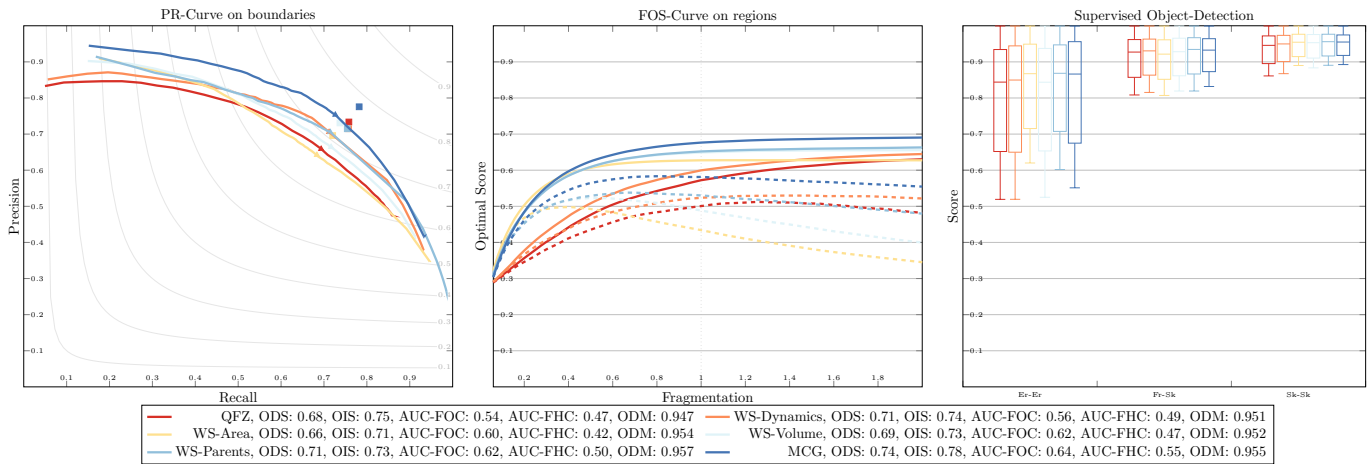


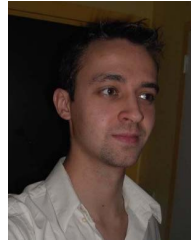
Figure 8. Best achieved results for each hierarchy and a state of the art method. (See Figure 5 for explanation).

In future work, we plan to study the integration of more complex and relevant visual cue to define watershed hierarchies, such as the ongoing works from [48] on iterative stochastic watershed hierarchies generation [28] or [49] on watershed hierarchies combinations. Another challenge will be to take account for richer gradient information as proposed by [50].

REFERENCES

- [1] S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing," *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 104–119, 1975.
- [2] S. L. Horowitz and T. Pavlidis, "Picture segmentation by a tree traversal algorithm," *Journal of the ACM*, vol. 23, no. 2, pp. 368–388, 1976.
- [3] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, no. 5, pp. 363–370, 1984.
- [4] J. M. Morel and S. Solimini, *Variational Methods in Image Segmentation*. Cambridge, MA, USA: Birkhauser Boston Inc., 1995.
- [5] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 561–576, 2000.
- [6] L. Guigues, J. P. Cocquerz, and H. Le Men, "Scale-sets image analysis," *International Journal of Computer Vision*, vol. 68, no. 3, pp. 289–317, 2006.
- [7] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.
- [8] Z. Ren and G. Shakhnarovich, "Image segmentation by cascaded region agglomeration," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 2011–2018.
- [9] B. Perret, J. Cousty, O. Tankyevych, H. Talbot, and N. Passat, "Directed connected operators: asymmetric hierarchies for image filtering and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1162–1176, 2015.
- [10] J. Pont-Tuset, P. Arbelaez, J. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping for image segmentation and object proposal generation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. In press, pp. 1–14, 2016.
- [11] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering occlusion boundaries from an image," *International Journal of Computer Vision*, vol. 91, no. 3, pp. 328–346, 2011.
- [12] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1132–1145, 2008.
- [13] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 1155–1162.
- [14] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive GMMRF model," in *European Conference on Computer Vision (ECCV)*, 2004, pp. 428–441.
- [15] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1558–1570, 2015.
- [16] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Mathematical Morphology and its Applications to Image Processing*, J. Serra and P. Soille, Eds. Kluwer Academic Publishers, 1994, pp. 69–76.
- [17] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1163–1173, 1996.
- [18] F. Meyer, "The dynamics of minima and contours," in *Mathematical Morphology and its Applications to Image and Signal Processing*, P. Maragos, R. Schafer, and M. Butt, Eds. Boston: Kluwer, 1996, pp. 329–336.
- [19] J. Cousty, G. Bertrand, L. Najman, and M. Couprie, "Watershed cuts: Minimum spanning forests and the drop of water principle," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 8, pp. 1362–1374, 2009.
- [20] J. Cousty, L. Najman, and B. Perret, "Constructive links between some morphological hierarchies on edge-weighted graphs," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer, 2013, pp. 86–97.
- [21] J. Cousty, L. Najman, Y. Kenmochi, and S. Guimarães, "Hierarchical segmentations with graphs: quasi-flat zones, minimum spanning trees, and saliency maps," LIGM, Research Report, 2016. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01344727>
- [22] L. Gueguen, S. Velasco-Forero, and P. Soille, "Local mutual information for dissimilarity-based image segmentation," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 3, pp. 625–644, 2014.
- [23] J. Cousty and L. Najman, "Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer, 2011, pp. 272–283.
- [24] L. Najman, J. Cousty, and B. Perret, "Playing with kruskal: algorithms for morphological trees in edge-weighted graphs," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer, 2013, pp. 135–146.
- [25] B. Marcotegui, "Residual approach on a hierarchical segmentation," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 4353–4357.
- [26] F. Meyer and J. Stawiaski, "A stochastic evaluation of the contour strength," in *32nd Annual Symposium of the German Association for Pattern Recognition (DAGM 2010)*, ser. Lecture Notes in Computer Science: Image Processing, Computer Vision, Pattern Recognition, and Graphics, M. Gesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, Eds., vol. 6376. Darmstadt, Germany: Springer Berlin / Heidelberg, 2010, pp. 513–522.

- [27] F. Malmberg, C. L. L. Hendriks, and R. Strand, "Exact evaluation of targeted stochastic watershed cuts," *Discrete Applied Mathematics*, vol. In press, 2016.
- [28] C. A. Pimentel, A. de Albuquerque Araújo, J. Cousty, S. J. F. G. aes, and L. Najman, "Stochastic hierarchical watershed cut based on disturbed topographical surface," in *Technical Papers of the 29th Conference on Graphics, Patterns and Images (SIBGRAPI'16)*, D. G. Aliaga, L. S. Davis, L. A. F. Fernandes, and W. R. Schwartz, Eds., 2016.
- [29] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *International Conference on Computer Vision (ICCV)*, vol. 2, 2001, pp. 416–423.
- [30] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [31] M. Nagao, T. Matsuyama, and Y. Ikeda, "Region extraction and shape analysis in aerial photographs," *Computer Graphics and Image Processing*, vol. 10, no. 3, pp. 195–223, 1979.
- [32] F. Meyer and P. Maragos, "Morphological scale-space representation with levelings," in *Scale-Space Theories in Computer Vision*, ser. Lecture Notes in Computer Science, M. Nielsen, P. Johansen, O. Olsen, and J. Weickert, Eds. Springer, 1999, vol. 1682, pp. 187–198.
- [33] C. Vachier and F. Meyer, "Extinction value: a new measurement of persistence," in *IEEE Workshop on nonlinear signal and image processing*, vol. 1, 1995, pp. 254–257.
- [34] M. Grimaud, "New measure of contrast: the dynamics," in *Proc. SPIE, Image Algebra and Morphological Image Processing III*, vol. 1769, 1992, pp. 292–305.
- [35] A. G. Silva and R. de Alencar Lotufo, "Efficient computation of new extinction values from extended component tree," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 79–90, 2011.
- [36] D. Martin, "An empirical approach to grouping and segmentation," Ph.D. dissertation, EECs Department, University of California, Berkeley, 2003.
- [37] G. Liu and R. M. Haralick, "Assignment problem in edge detection performance evaluation," in *IEEE Conference on Computer Vision and Pattern Recognition.*, vol. 1, 2000, pp. 26–31.
- [38] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, 2004.
- [39] J. Pont-Tuset and F. Marques, "Supervised evaluation of image segmentation and object proposal techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1465–1478, 2016.
- [40] —, "Supervised assessment of segmentation hierarchies," in *European Conference on Computer Vision (ECCV)*, 2012.
- [41] —, "Upper-bound assessment of the spatial accuracy of hierarchical region-based image representations," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012.
- [42] B. R. Kiran and J. Serra, "Global-local optimizations by hierarchical cuts and climbing energies," *Pattern Recognition*, vol. 47, no. 1, pp. 12–24, 2014.
- [43] P. P. Michael Maire, Stella Yu, "Hierarchical scene annotation," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013.
- [44] Q. Huang and B. Dom, "Quantitative methods of evaluating image segmentation," in *International Conference on Image Processing*, vol. 3, 1995, pp. 53–56.
- [45] B. Perret, J. Cousty, J. C. Rivera Ura, and S. J. F. Guimares, "Evaluation of morphological hierarchies for supervised segmentation," in *Mathematical Morphology and Its Applications to Signal and Image Processing*, ser. Lecture Notes in Computer Science, J. Benediktsson, J. Chanussot, L. Najman, and H. Talbot, Eds., vol. 9082. Reykjavik, Iceland: Springer, 2015, pp. 39–50.
- [46] J. Chaussard, M. Couprie, and H. Talbot, "Robust skeletonization using the discrete lambda-medial axis," *Pattern Recognition Letters*, vol. 32, no. 9, pp. 1384–1394, 2011.
- [47] S. Guimarães, Y. Kenmochi, J. Cousty, Z. Patrocinio, and L. Najman, "Hierarchizing graph-based image segmentation algorithms relying on region dissimilarity: the case of the Felzenszwalb-Huttenlocher method," LIGM, Research Report, 2016.
- [48] A. Fehri, S. Velasco-Forero, and F. Meyer, "Automatic selection of stochastic watershed hierarchies," *arXiv preprint arXiv:1609.02715*, 2016.
- [49] D. Satana Maia, J. Cousty, B. Perret, and L. Najman, "Combinations of hierarchical image segmentations," LIGM, Submitted, 2016.
- [50] K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. V. Gool, "Convolutional oriented boundaries," in *European Conference on Computer Vision (ECCV)*, 2016.



Benjamin Perret received his M.Sc. in Computer Science in 2007, and his Ph.D. in Image Processing in 2010 from the Universit de Strasbourg (France). He currently holds a teacher-researcher position at ESIEE Paris, affiliated with the Laboratoire dInformatique Gaspard Monge, Universit Paris-Est. His current research interests include image processing and analysis.



Jean Cousty received his Ingénieurs degree from ESIEE Paris (France) in 2004 and the Ph.D. degree from the Universit de Marne-la-Valle (France) in 2007. After a one-year post-doctoral period in the ASCLEPIOS research team at INRIA (Sophia Antipolis), he is now teaching and doing research with the Informatics and Telecom Department, ESIEE Paris, and with the Laboratoire dInformatique Gaspard Monge, Universit Paris-Est. His current research interests include image analysis and discrete mathematics.



Silvio Jamil F. Guimarães Biography text here.



Deise S. Maia Biography text here.