Evaluation of hierarchical watersheds

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

Evaluation of hierarchical watersheds.

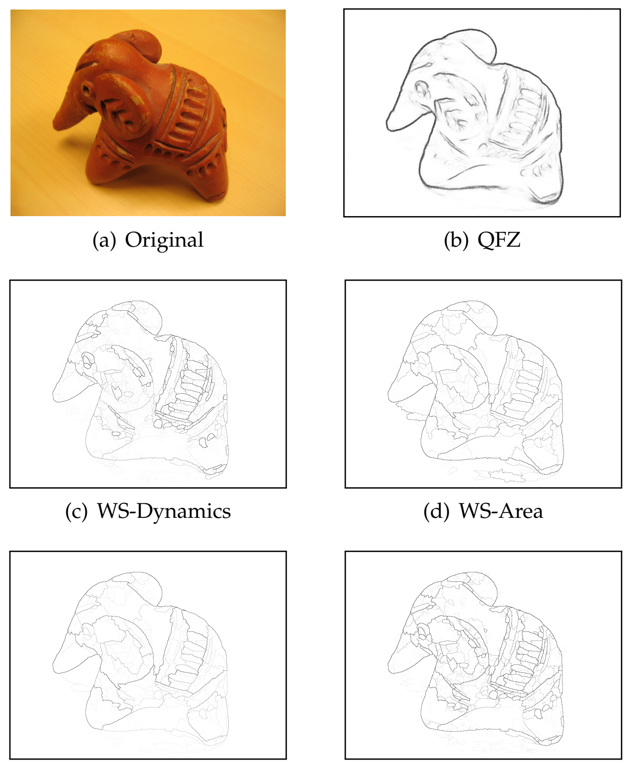
Benjamin Perret, Jean Cousty, Silvio Jamil F. Guimarães, and Deise S. Maia

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 efficiently computed on large data, make them valuable candidates for various computer vision tasks.

Index Terms—mathematical morphology, hierarchy of partitions, watershed segmentation, 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], [11], [12], occlusion boundary detection [13], image simplification [6], [9], [14], object detection [5], object proposal [10], visual saliency estimation [15]. In particular, they have gained a large popularity in [7] whose hierarchical approach to the general problem of natural image segmentation outperformed state-of-the-art approaches.



(a) Original (b) QFZ (c) WS-Dynamics (d) WS-Area (e) WS-Volume (f) WS-Parents

Figure 1. Saliency maps of the morphological hierarchies studied in this article –QFZ, WS-Dynamics, WS-Area, WS-Volume, and WS-Parents– on the elefant image from Grabcut dataset [16] with SED gradient [17].

It has long been noted [18], [19], [20] 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 [21]. It has also been shown that hierarchies of watersheds are linked to the quasi-flat zones hierarchies [22], to the single-linkage clustering problem [23], and to connective segmentation [14], [24].

Hierarchies of watersheds are thus multi-scale representations which satisfy a global optimality property. Moreover, there exist efficient algorithms, with the same time complexity as minimum spanning tree algorithms, to construct them enabling to process large images in real time [25], [26]. In recent years, they have been used for the computation of morphological operators [27], in the context of stochastic watershed segmentation [28], [29], [30]. However, their practical performances have not yet been studied in the general case of natural image analysis and the aim of this work is

- *B. Perret, J. Cousty, and D. S. Maia are with Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge UMR 8049, UPEMLV, ESIEE Paris, ENPC, CNRS, F-93162 Noisy-le-Grand France.*
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

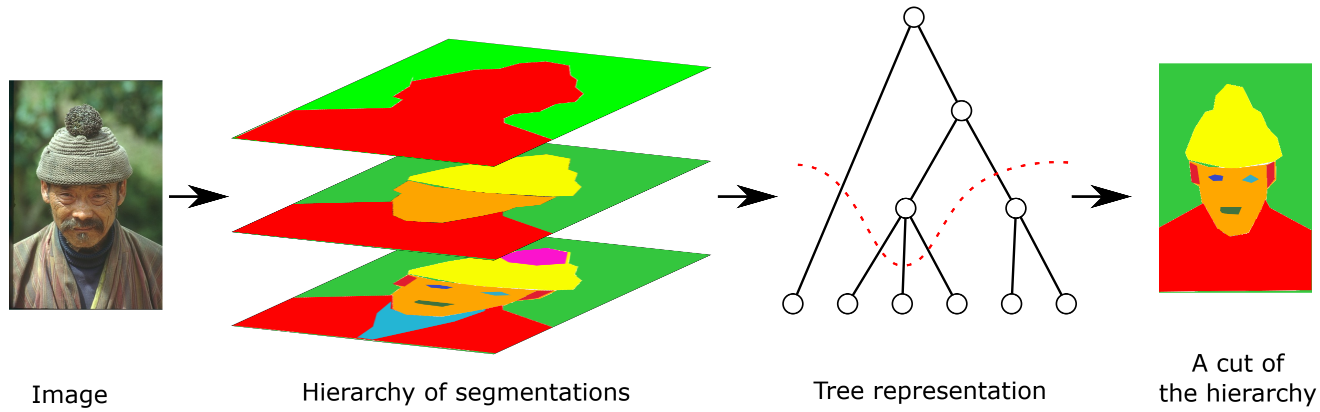


Figure 2. A hierarchy of partitions of an image (source BSDS 500 [7]) is a sequence of coarse to fine partitions. The hierarchy can be represented as a tree of regions. A cut is a partition made of regions of the hierarchy possibly taken at different levels (red dashed curve).

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) Pascal Context segmentation dataset [31], and 2) MS-COCO [32] and Pascal VOC'12 [33] object 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 it is not limited to a single use case.

This framework can be 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 relation.

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

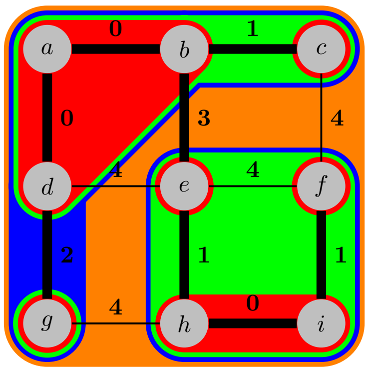
A *hierarchy (of partitions)* is a sequence $\mathcal{H} = (\boldsymbol{P}_0, \ldots, \boldsymbol{P}_n)$ of partitions of V such that $\boldsymbol{P}_0$ contains every singletons of V, *i.e.*, $\boldsymbol{P}_0 = \{\{x\} \mid x \in V\}$, the partition $\boldsymbol{P}_n$ is the single region partition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

Given a hierarchy $\mathcal{H} = (\boldsymbol{P}_0, \ldots, \boldsymbol{P}_n)$, the set of *regions of* $\mathcal{H}$, denoted by $\boldsymbol{R}_{\mathcal{H}}$ is the union of all partitions of $\mathcal{H}$. The inclusion relation on $\boldsymbol{R}_{\mathcal{H}}$ 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 $\boldsymbol{R}_{\mathcal{H}}$, denoted by $Parent(R)$, if the smallest region R' of $\boldsymbol{R}_{\mathcal{H}}$ that is strictly larger than R (see Figure 2).

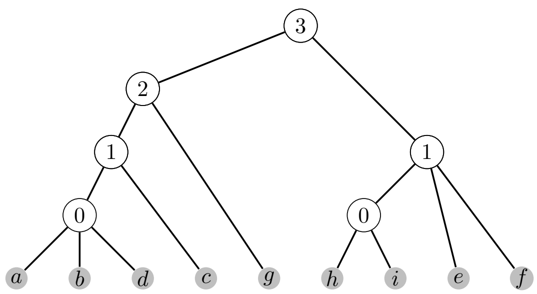
Given a hierarchy $\mathcal{H} = (\boldsymbol{P}_0, \ldots, \boldsymbol{P}_n)$, a partition $\boldsymbol{P}$ of V made of regions of $\mathcal{H}$ (*i.e.*, $\boldsymbol{P} \subseteq \boldsymbol{R}_{\mathcal{H}}$) is called a *cut of* $\mathcal{H}$ (see Figure 2). The set of all cuts of a hierarchy $\mathcal{H}$ is denoted by $\boldsymbol{\Pi}(\mathcal{H})$. A cut $\boldsymbol{P}$ of a hierarchy $\mathcal{H} = (\boldsymbol{P}_0, \ldots, \boldsymbol{P}_n)$ is said *horizontal* if $\boldsymbol{P} = \boldsymbol{P}_i$ for some i in $\{0, \ldots, n\}$.

2.2 Quasi-flat zones hierarchy

The quasi-flat zones have been studied since the 70's (see *e.g.*, [14], [34], [35]). They are deeply linked to single-linkage clustering and to the notion of a minimum spanning tree [23]. 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 set of quasi-flat zones of the weighted graph at a given level λ is a partition of its vertices (see Figure 3.a). 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 (see Figure 3.b).



(a) Weighted graph $(\mathcal{G}, W)$ and its quasi-flat zones



(b) Quasi-flat zones hierarchy of $(\mathcal{G}, W)$

Figure 3. a) A weighted graph $(\mathcal{G}, W)$ and its sets of quasi-flat zones at levels λ equals to 0 (red), 1 (green), 2 (blue), and 3 (orange). The bold edges are in the paths of minimal weight inside the quasi-flat zones. b) The quasi-flat zones hierarchy QFZ of $(\mathcal{G}, W)$: the gray nodes corresponds to the nodes of $\mathcal{G}$. The number inside each node indicates the level λ where the node appears in the hierarchy.

2.3 Watershed hierarchies

Watershed hierarchies were first proposed in [18], [19], [20] and have since been formalized in the context of minimum spanning forests [21], [25]. 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 form a segmentation with a global optimality property similar to the one of the minimum 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 [21].

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 [36]. Such hierarchies are called hierarchical watersheds; their theoretical properties and some algorithms to construct them are in particular studied in [22], [25], [26].

Extinction values are defined through regional attributes defined on the connected components of the level sets of the weight map [36]. Intuitively, the extinction value of a minimum 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 [36], [37]: *e.g.*, dynamics, area, or volume. Other authors have proposed regional measures related to topological properties of the function inside the components: topological height [38], or number of descendants [38]. 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 (see next section 2.4). All these regional attributes and their associated extinction values can be computed from the quasi-flat zones hierarchy of the weighted graph [26]. 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.

2.4 Attribute number of parent nodes

In this section, we present the proposed attribute *Number of Parents*.

Given an edge weighted graph $(\mathcal{G} = (V, E), W)$ and its quasi-flat zones hierarchy QFZ, the min-tree MT of the weighted graph is the set of regions of the QFZ hierarchy minus the singletons, *i.e.*, $\text{MT} = \boldsymbol{R}_{\text{QFZ}} \setminus \{\{x\}, x \in V\}$. The min-tree MT is not a hierarchy of partitions, however equipped with the inclusion relation on its element, it is a tree isomorphic to QFZ in the sense of [22] (see Figure 4). The leaves of MT correspond to the regional minima of the weighting function W and are thus associated to the catchment basins of the watersheds of W.

For any element n of MT, the *number of parent nodes* of n, denoted $NumPar(n)$ is then defined as the number of elements of MT included in n and having at least one child, *i.e.*, $NumPar(n) = |\{x \in \text{MT} \,|\, x \subseteq n, \exists y \in \text{MT}, y \subset x\}|$ (see Figure 4). $Numpar$ measures the number of times the minima of the gradient are modified, either by the addition of new pixels (growth of the associated catchment basin), or by the merging with another minima (fusion of catchment

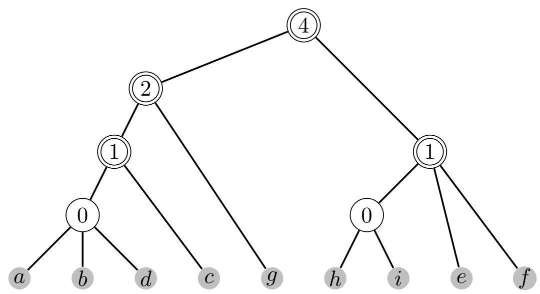


Figure 4. Quasi-flat zones hierarchy and min tree of $(\mathcal{G}, W)$ in Figure 3 a). The gray nodes corresponds to the vertices of $\mathcal{G}$, the other nodes are the elements of the min tree MT of $(\mathcal{G}, W)$. The nodes surrounded by a double circle are the nodes of MT having at least one child. The value of the attribute $NumPar$ is indicated inside the nodes: it counts the number of nodes having at least one child in MT below a given node (included).

basins). Intuitively, it measures the amount of change in a given region where minima and flat components have been contracted as single pixels. Such kind of attributes that measures the *complexity* of a region has also been investigated in the algebraic framework of lattices studied in [39].

As a topological feature, it is invariant to monotone contrast transformations and to geometric transformations (up to discretization effects). It is also increasing (the attribute value of a node is larger than the one of its children) which allows defining an extinction value associated to each minima of the function [36], thus leading to a hierarchical watershed.

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 [19], *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 can 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 similarity 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: 1) the optimal image scale OIS measuring the best achievable score when taking the optimal horizontal cut in each hierarchy, and 2) 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 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) [40] where image segmentation is viewed as a multi-class clustering problem on the image pixels, 2) F-Measure for boundaries (FB) [40], [41], [42] where image segmentation is viewed as a binary clustering problem on the pixels' boundaries, and 3) F-Measure for objects and parts (FOP) [43] 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 [43] 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 [44], [45] 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 [44], and 2) upper bound on local additive region measures solved with dynamic programming [45] (similar to finding the optimal partition for a given energy function [6], [46]). Those methods do 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 [47] 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 motivate the choice of the assessment measures and their contributions. Then, we give a detailed description of each measure.

The popular evaluation strategy for hierarchies of partitions, precision-recall curves for FB and FOP, assesses only the horizontal cuts of hierarchies with similarity measures between segmentations. Indeed, the framework of precision recall curves can be applied to any sequence of partitions without any hierarchical constraint between them. However, hierarchies of partitions are rich representations with strong structural properties that are used by many applications beyond their simple horizontal cuts: energies and algorithms for optimal cut segmentation [6], [11], [12], [46], image filtering with hierarchy pruning [5], [14], [48], objects detection [10], [49], interactive segmentation with multi-scale region selection [5], [9], [50]. Therefore, our goal is to complete the standard precision-recall curves with other measures in order to capture more information relevant to those applications fields.

Upper-bound measures [44], [45], which enable to quantify the maximal achievable score among all possible cuts of a hierarchy, can provide valuable information for all the applications based on non-horizontal cuts: they are thus a candidate of choice to extend precision-recall curves assessment. This segmentation ability can be studied with respect to the number of desired regions in the target segmentation, enabling to identify, in the context of hierarchies, some classical properties of image segmentation such as under- and over-segmentation. The comparison of this upper-bound score with the scores obtained on horizontal cuts also enables to evaluate if the indexing of the partitions of the hierarchy is coherent with the ground truths.

Thus, in order to evaluate the cuts of a hierarchy, both horizontal and non-horizontal, we propose to use both precision-recall curves and upper bound measures. Precision-recall curves and upper-bound measures rely on a similarity measure between segmentations. As shown in [43], the complementarity between region based and boundary based similarity measures is important for the evaluation of segmentations. For precision-recall curves, we will use the combination of similarity measures FB and FOP (see Section 3).

For the upper-bound assessment, we will also use the similarity measure FB to evaluate the quality of boundaries. However, we still lack a computationally tractable algorithmic solution to find the upper-bound partition for FOP measure, which implies that we have to use a less discriminant measure [43] in order to evaluate the upper-bound quality of regions. In [45], the authors focused on the upper-bound of the directional Hamming distance [51]. However, this similarity measure is transparent to under-segmentation, *i.e.*, it does not penalize the merging of several regions of the ground truth into a single region in the proposal segmentation (the single region partition always achieves the highest similarity with any other partition). In this work, we propose instead to use the Bidirectional Consistency Error BCE [40] measure in order to overcome this limitation of the directional Hamming distance. Indeed, BCE is a symmetric measure, *i.e.*, it enables to evaluate both over- and under-segmentation.

Concerning methods focusing on regions of the hierarchy, we propose new evaluation measures that aim to quantify the easiness of finding a set of nodes of a hierarchy representing a semantic object in the scene. This approach considers sets of nodes of the hierarchy which are generally not cuts and the proposed objects generally do not exist in any of the cuts of the hierarchy. Intuitively, an object will be considered easy to find if it can be retrieved with few markers, *i.e.*, if the user of an interactive segmentation procedure can retrieve the object precisely with few annotations.

4.1 Upper-bound measure

We propose an evolution of the upper-bound evaluation on regions proposed by Pont-Tuset et al. [44], [45] that consists in the definition of a new type of curve, the fragmentation-upper bound curve that enables to measure the potential gain of non-horizontal cuts compared to horizontal cuts.

Given a similarity measure s, an image I, one ground-truth segmentation $\boldsymbol{T}_I$, and a proposal segmentation $\boldsymbol{S}_I$, we denote by $s(\boldsymbol{S}_I, \boldsymbol{T}_I)$ the similarity between $\boldsymbol{S}_I$ and $\boldsymbol{T}_I$ for s. Given a hierarchy of partitions $\mathcal{H}_I$ on the image I, one ground-truth segmentation $\boldsymbol{T}_I$ and a number k of regions, the *Upper-Bound score for* s (UB_s) for $\mathcal{H}_I$ is the highest score according to s for all the cuts of $\mathcal{H}_I$ composed of k regions:

$$UB_s(\mathcal{H}_I, \boldsymbol{T}_I, k) = \max_{\substack{\boldsymbol{S} \in \boldsymbol{\Pi}(\mathcal{H}_I) \\ |\boldsymbol{S}| = k}} s(\boldsymbol{S}, \boldsymbol{T}_I). \tag{1}$$

In the context of Upper-Bound measure, when several distinct ground truths are available for a same image, we propose to define the image score as the average score over the set of ground truths. Indeed, a *good* hierarchy should be able to successfully represent the image at different scales. Therefore, if the experts provide different levels of details in different regions of the image, there should exists a *good* partition in the set all cuts of the hierarchy for each ground truth.

This leads to the mean-average Upper-Bound FB score on a database, *i.e.*, the mean image score over the database. Finally, we propose to study the *Fragmentation–Optimal Cut score curve (FOC)*, where the mean-average Upper-Bound score is plotted against the *fragmentation* level of the segmentation defined as $k/|\boldsymbol{T}_I|$, the ratio between the number of regions in the segmentation and the number of regions in the ground truth (see Figure 7).

The gain achieved by taking a non-horizontal cut in the hierarchy is evaluated with a second curve called *Fragmentation-Horizontal Cut score curve (FHC)* where the mean-average FB score of the horizontal cuts of the hierarchies is plotted against the *fragmentation* level of the cuts. 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 provide 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 (*i.e.*, when the proposed partition and the ground truth contain the same

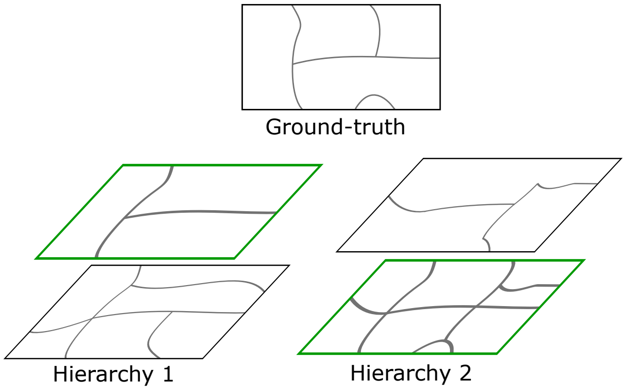


Figure 5. Illustration of under- and over-segmentation for hierarchies. Hierarchies 1 and 2 are both composed of 2 levels. Compared to the ground truth, the first hierarchy manages to recover long contours in its coarse level but then fails to recover the other contours at a finer level: the optimal 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 finest partition recovers all the contours of the ground truth but also contains extra-contours. However, the coarsest partition looses the true contours and preserves extra contours: the hierarchy is said to over-segment the image.

number of regions). If the maximum occurs at fragmentation level lower than 1, this means that the hierarchy tends to capture the main regions of the ground truth with a low number of regions but then fails to correctly refines those regions (see *Hierarchy 1* in Figure 5): in this case we say that the hierarchy has a *tendency for under-segmentation*. If the maximum occurs 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 5): in this case we say that the hierarchy has a *tendency for over-segmentation*.

As a performance summary on the FOC curves, we compute the normalized area under the curve, denoted by AUC-FOC. 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- segmentation, 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 $1/2$.

4.2 Object detection measure

The last measure, introduced in our previous work [50], 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 chose 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:

$$O(\mathcal{H}_I, M_o, M_b) = \bigcup \{R \in \boldsymbol{R}_{\mathcal{H}_I} \mid R \cap M_o \neq \emptyset, R \cap M_b = \emptyset\}\,. \quad (2)$$

The extracted object is thus a union of regions of the hierarchy: *i.e.*, it is generally not a region in any of the partitions of the hierarchy (it is not present in any cut of the hierarchy).

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. $\forall x \in V$, $\ell(\{x\}) \in \{Object, Backgound, Undefined\}$
Result: The segmented object O according to Eq. (2).

// For all nodes from leaves to root
for *all regions* R *in* $\boldsymbol{R}_{\mathcal{H}}$ *in increasing order* **do**
  if $R \neq V$ **then** // R is not the root
    if $\ell(R) = Background$ **then**
      $\ell(Parent(R)) \leftarrow Background$;
    else if $\ell(R) = Object$ ***and*** $\ell(Parent(R)) = Undefined$ **then**
      $\ell(Parent(R)) \leftarrow Object$;

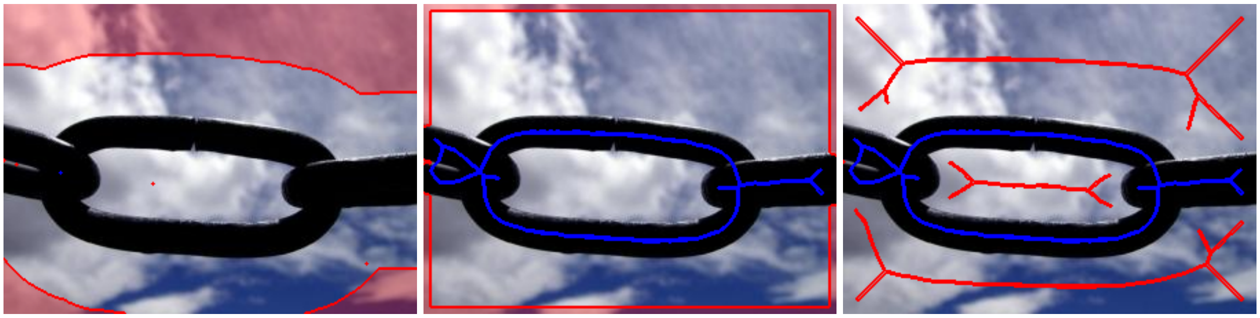
// For all nodes from root to leaves
for *all regions* R *in* $\boldsymbol{R}_{\mathcal{H}}$ *in decreasing order* **do**
  if $R \neq V$ ***and*** $\ell(R) = Undefined$ **then**
    $\ell(R) \leftarrow \ell(Parent(R))$;

$O \leftarrow \{x \in V \mid \ell(\{x\}) = Object\}$

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 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 practice this algorithm enables to process the hierarchy of an image from MS-COCO and Pascal VOC'12 datasets in less than a millisecond[1].

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 Figure 6): 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 [52] of this connected component is added to the marker, 2) Skeleton (Sk): morphological skeleton given by [53], 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

1. Demonstration available online with an interactive segmentation tool at https://perso.esiee.fr/~perretb/ISeg/



(a) Er-Er (b) Sk-Fr (c) Sk-Sk

Figure 6. 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.

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, this combination is symmetric in the sense that the boundaries of the correct segmentation are 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 asymmetric.

The performance of each segmentation result is evaluated with the F-Measure. The median score for the 3 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 on FOP and upper-bound on FB measure are evaluated on the Pascal Context dataset [31] which consists of a pixel-wise segmentation of the 10 103 images on the Pascal VOC'10 dataset [54]. In our experiments, the test set is composed of the last 2 498 images of the Pascal VOC'10 validation set as proposed in [55]. The object detection measure is evaluated on the MS-COCO [32] and Pascal VOC'12 segmentation [33] datasets. Each object of these two datasets is processed independently using the framework described in Section 4.2. This leads to a total of 291 875 objects from the 40 504 images of the MS-COCO 2014 validation set and 3 427 objects from the 1 449 images of the Pascal VOC'12 segmentation test set. 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 state-of-the-art methods (Section 5.3).

5.1 Influence of gradient

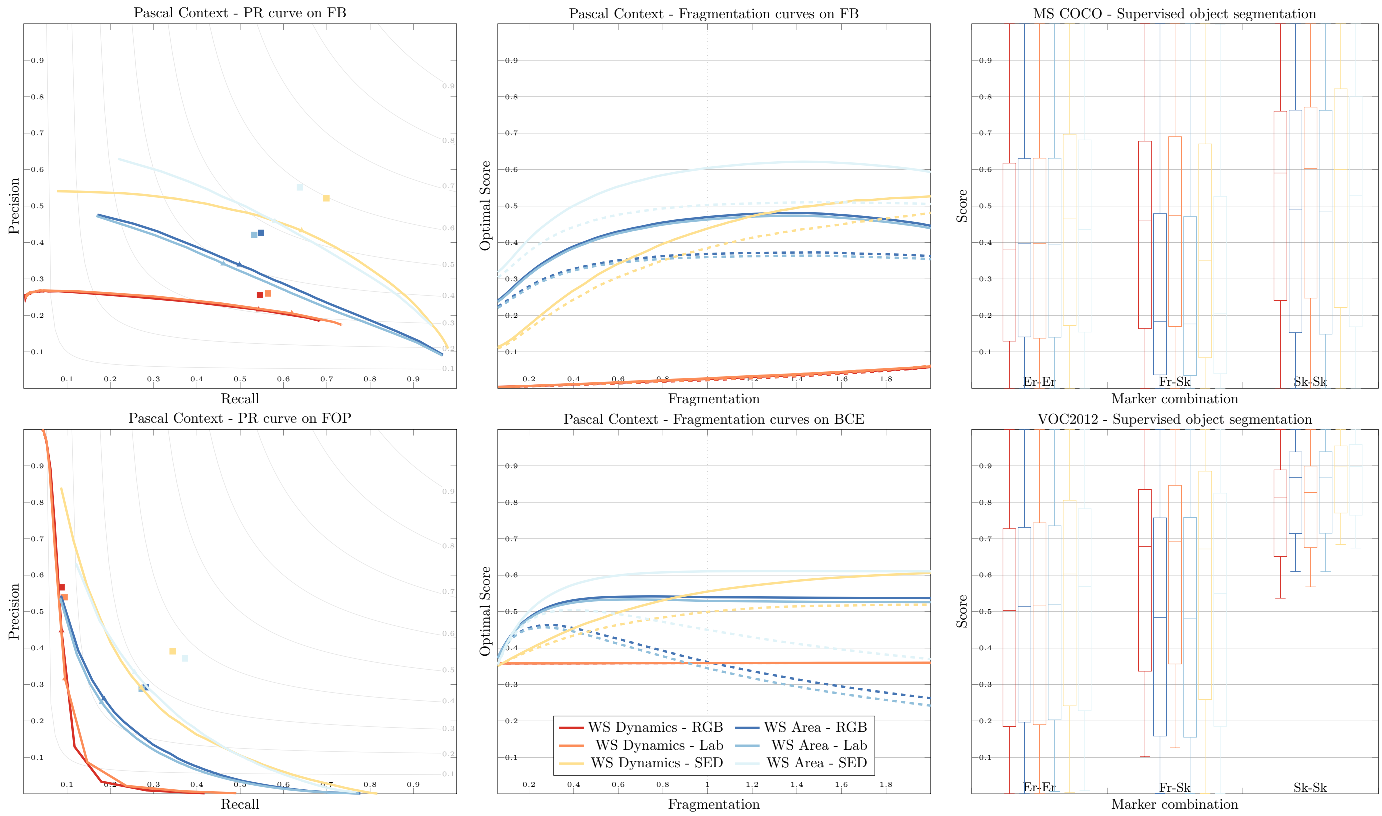
A usual way to weight the edges of a graph in image analysis in general and for morphological segmentation in particular is to use a gradient measure. The aim of this section is to evaluate the influence of the gradient measure on the assessed quality of the hierarchies.

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 [56]. 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 structured edge detector (SED) from [17]. For SED, we used the default parameters given in [17]: single scale with sharpening and 4 decision trees.

Figure 7 shows the result of WS-Dynamics and WS-Area with three considered gradients: 1) RGB, 2) Lab, and 3) 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 FOC and FHC curves on FB and BCE are nearly flat. This 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. It can be a consequence of WS-Dynamics and QFZ having their upper levels made only of small salient regions; a solution to this problem is presented in the next section.

We do not observe a clear improvement with Lab gradient compared to RBG gradient. However, there is almost always a large gain by switching from a local RGB or Lab gradient to the supervised non-local gradient SED. The FOC and FHC curves show that WS-Dynamics requires a large number of regions to reach its maximal scores with SED gradients which implies that the hierarchy tends to have small irrelevant regions on its highest levels. This suggests that despite the regularization effect based on the dynamics, which tends to send lowly contrasted regions to the lowest levels of the hierarchy, WS-Dynamics remains sensitive to small regions of high contrast.

In conclusion, we recommend the use of SED gradient to build watershed hierarchies on natural images and the



	ODS		AUC-FOC		ODM		Mean score
	FB	FOP	FB	BCE	MS-COCO	VOC'12	
WS Dynamics - RGB	0.309	0.146	0.028	0.359	0.609	0.787	0.373
WS Dynamics - Lab	0.309	0.146	0.029	0.359	0.621	0.803	0.378
WS Dynamics - SED	0.517	0.281	0.402	0.532	0.636	0.858	0.538
WS Area - RGB	0.403	0.217	0.436	0.529	0.501	0.770	0.476
WS Area - Lab	0.393	0.209	0.430	0.520	0.496	0.775	0.471
WS Area - SED	0.512	0.290	0.564	0.588	0.552	0.829	0.556

Figure 7. Influence of the gradient on WS-Dynamics and WS-Area hierarchies . **Precision-recall (PR) curves** for FB and FOP on Pascal Context (first column): each curve represents the variation of precision and recall for the different partitions of the hierarchy. OIS and ODS scores are represented in the plot by a square and a triangle. **Fragmentation–Optimal Cut score curves (FOC)** for FB and BCE on Pascal Context (second column): 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. **Supervised object detection** on MS-COCO and Pascal VOC'12 (last column): 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) of the lower quartile, and the highest datum still within 1.5 inter quartile range of the upper quartile range (lower and upper extremities). The principal performance measures are summarized in the table: F-Measure of FB and FOP scores at ODS (precision-recall curves), AUC-FOC for FB and BCE (fragmentation curves), and ODM on MS-COCO and Pascal VOC'12 datasets (supervised object detection). The last column of the table gives the mean score of each method.

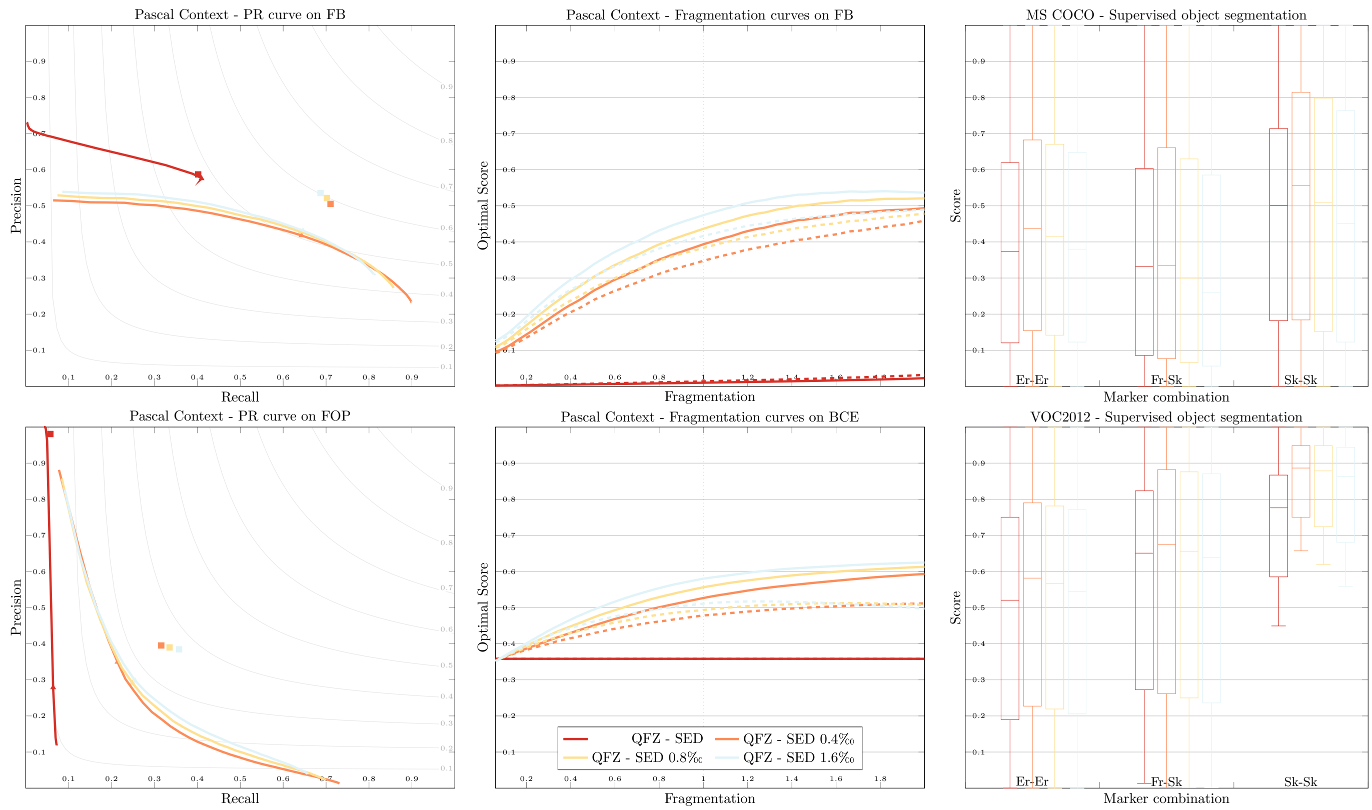
following experiments will be conducted with this gradient.

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 [57] 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 8 shows the results 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 that the introduction of the filtering immediately produces a large performance boost. Between $r_k = 0.4$‰ and $r_k = 0.8$‰, the situation is mixed with an improvement in image segmentation measures (precision-recall curves and fragmentation curves) and a small decrease on objection detection measures. This tendency continues when r_k increases to 1.6‰. This situation can be the results of two factors: 1) the tradeoff between the number of regions necessary to describe the scene and the precision of boundaries, and 2) the difference between Pascal Context dataset, where most image segments are large, and MS-COCO and Pascal VOC'12 datasets which contain a large number of small objects that may be affected by strong filters. The effect of the filtering on over-segmentation can be observed on



	ODS		FOC		ODM		Mean score
	FB	FOP	FB	BCE	MS COCO	VOC 2012	
QFZ - SED	0.479	0.104	0.011	0.358	0.550	0.776	0.380
QFZ - SED 0.4‰	0.514	0.268	0.399	0.533	0.579	0.840	0.522
QFZ - SED 0.8‰	0.517	0.275	0.430	0.550	0.524	0.827	0.521
QFZ - SED 1.6‰	0.519	0.284	0.456	0.567	0.448	0.800	0.512

Figure 8. Influence of the area filter on QFZ. (See Figure 7 for explanation).

fragmentation curves, where the positions of the maxima of the curves tend to move from large fragmentation values toward 1 when the size of the filter r_k increases.

In conclusion, for QFZ and WS-Dynamics we recommend to perform a post-filtering of the hierarchy by removing regions smaller than 0.4‰ or 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 or even be detrimental.

5.3 Discussions

This section compares and discuss the best results obtained for each morphological hierarchy (see Figure 9) and compare them to state-of-the-art methods in terms of evaluation measures (see Figure 10) and computation times (see Table 1).

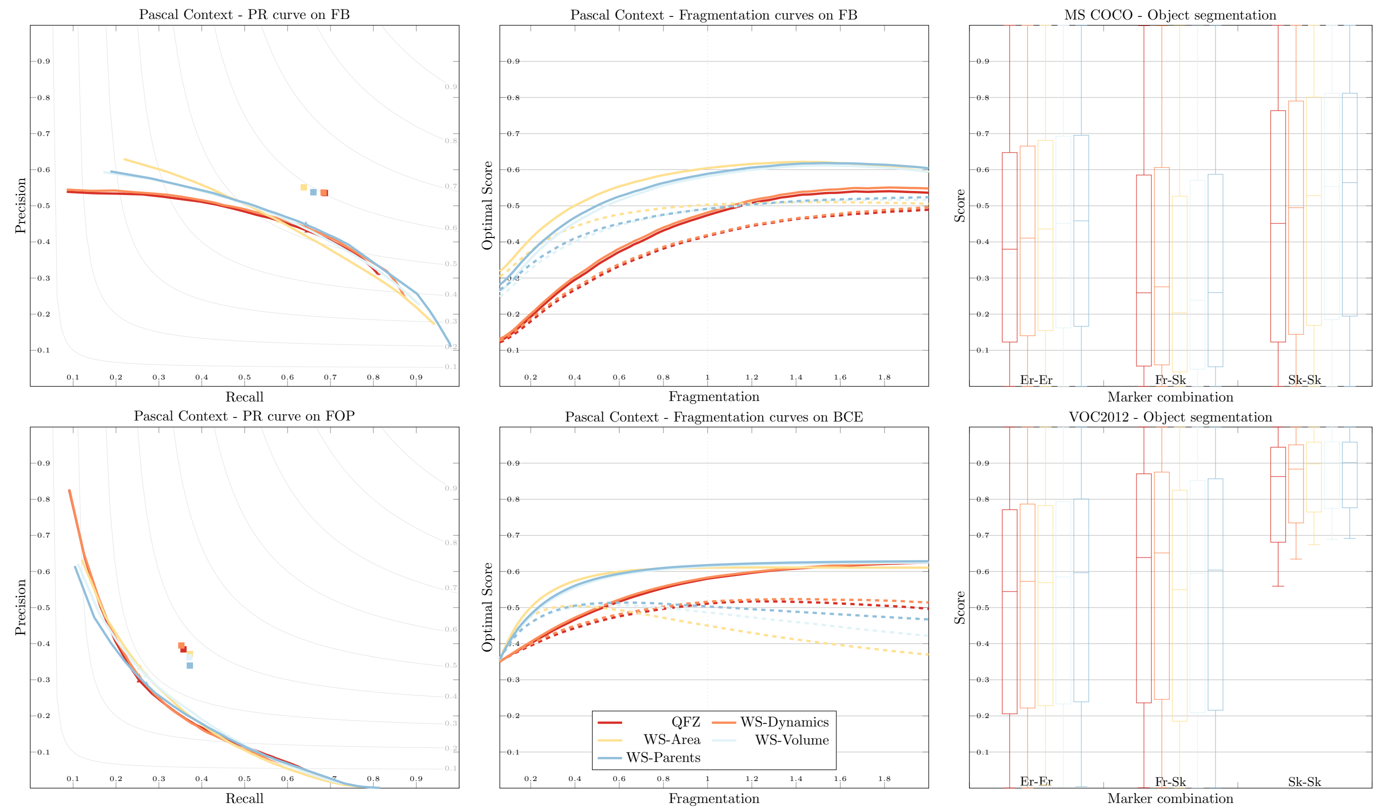
We can observe that the scores of QFZ are lower than the results of the other methods on all the measures.

WS-Dynamics shows good performances on horizontal cuts (precision-recall curves) and medium performance on supervised object detection. However, it is behind in terms of upper-bound measures. It has a tendency for over-segmentation (maximum of FOC and FHC curves occur at large values of fragmentation). We can also notice that WS-Dynamics (and QFZ) performances for object detections in the Er-Er and Sk-Sk cases are significantly lower than other methods; this suggests that those hierarchies fail 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).

WS-Area shows very good performances on quality of regions both on horizontal and non-horizontal cuts. It also has the highest upper-bound FB measure but, a low FB measure on horizontal cuts, suggesting a strong problem in the indexing of its regions. Its scores on supervised object segmentation are quite low compared to other methods: the scores of WS-Area on the Fr-Sk markers are particularly low. This can be explained by the asymmetric nature of the markers in this case (the true contours are located much closer to the object marker than the background frame), WS-Area is then penalized by its tendency to produce a regular (in size) tiling of the image.

WS-Volume behavior is close to WS-Area, its performances are however more balanced with higher scores of FB on horizontal cuts and lower ones on non-horizontal cuts. Its scores on supervised object segmentation are also very good.

WS-Parents offers the best overall performances. While



	ODS		FOC		ODM		Mean score
	FB	FOP	FB	BCE	MS COCO	VOC 2012	
QFZ	0.517	0.275	0.430	0.550	0.524	0.827	0.521
WS-Dynamics	0.523	0.279	0.438	0.554	0.560	0.843	0.533
WS-Area	0.512	0.290	0.564	0.588	0.552	0.829	0.556
WS-Volume	0.521	0.290	0.539	0.588	0.584	0.844	0.561
WS-Parents	0.528	0.279	0.549	0.592	0.594	0.847	0.565

Figure 9. Best achieved results for watershed hierarchy. (See Figure 7 for explanation).

not being first on every measure, it is the most balanced method. As WS-Area and WS-Volume, it shows a small tendency for under-segmentation.

As reference state-of-the-art results, we include Multiscale Combinatorial Grouping (MCG) hierarchies from [10], Convolutional Object Boundaries (COB) hierarchies from [55], [58], and Least Effort Segmentation (LEP) from [59] in our assessments. MCG also uses SED as the main cue for contour detection, but then merges several hierarchies (referred to as OWT-UCM in the literature [7]) computed at different scales. COB relies on hierarchy construction algorithm similar to MCG but uses a CNN based gradient detector. LEP also relies on a modified version of SED for contour detection and utilizes a region merging criterion that combines a classical data fidelity term with a new regularization term that measures the effort needed for a human segmenter to draw the contour of a region. Finally, we also include the results obtained with a *Random Hierarchy* to provide a baseline. The Random Hierarchy of an image is defined as the QFZ hierarchy of a random gradient. This random gradient does not contain any information about the image content: the edges of the graph are weighted by random uniformly distributed values.

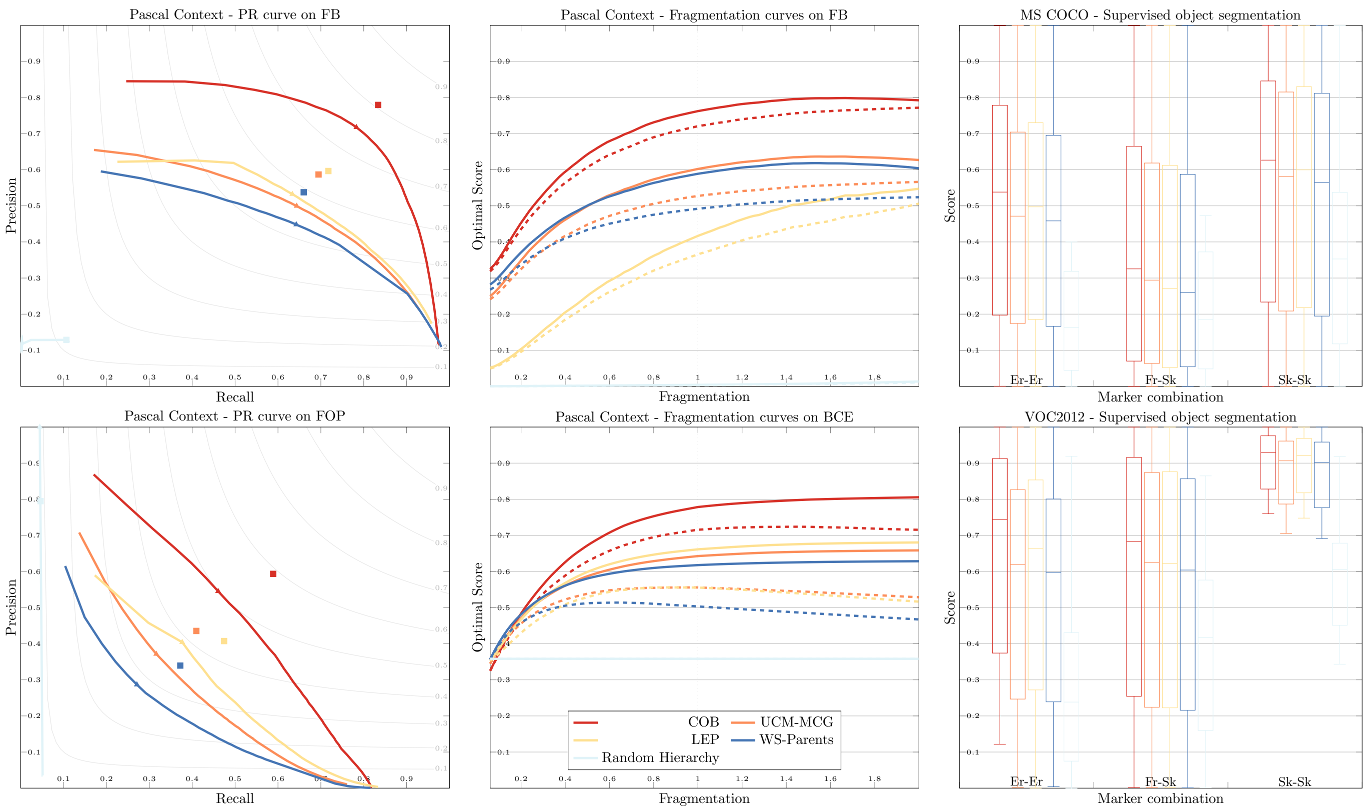
Comparing to state-of-the-art methods, COB, which is the only method to rely on a convolutional neural network trained on Pascal Context train set, clearly dominates. WS-Parents does not perform as well as other methods on precision-recall curves but is competitive in terms of upper-bound measures: this indicates that the indexing of the regions in the WS-Parents is not optimal.

Execution times are reported in Table 1. Watershed hierarchies are at least an order of magnitude faster than other methods on BSDS 500 images [7]. The execution time of the different watershed hierarchies are all similar as the computation of the attribute values represent a negligible part of the total computation. We also performed tests on high resolution RGB images (4160×2340 pixels) giving a mean processing time of 5.5 s (4 s to compute SED gradient and 1.5 s to construct the watershed hierarchy).

	WS	LEP [59]	MCG [10]	COB [55]
gradient	70 ms	1 s	4 s	300 ms*
hierarchy	20 ms	1 s	20 s	500 ms
total	90 ms	2 s	24 s	800 ms

Table 1
Execution times to compute a hierarchy of partition on a BSDS 500 image (481×321 pixels). The asterisk indicates the use of a GPU.

Concerning the evaluation framework, we can observe



	ODS		FOC		ODM		Mean score
	FB	FOP	FB	BCE	MS COCO	VOC 2012	
COB	0.749	0.499	0.705	0.722	0.681	0.919	0.713
UCM-MCG	0.562	0.342	0.556	0.610	0.615	0.864	0.592
LEP	0.579	0.390	0.378	0.626	0.632	0.881	0.581
WS-Parents	0.528	0.279	0.549	0.592	0.594	0.847	0.565
Random Hierarchy	0.117	0.085	0.005	0.358	0.312	0.535	0.235

Figure 10. Comparison with state-of-the-art methods. (See Figure 7 for explanation).

the complementarity of the new metrics – FHC curve, FOC curve, and object detection – with the standard precision-recall curve. The FOC and FHC curves enable to see a difference between the behavior of LEP and MCG that is not noticeable in the precision-recall curves, showing that LEP has a tendency for over-segmentation. They also enable to characterize the indexing problem in the WS-Parents hierarchy which exhibits a large gap between the FHC and FOC curves. The discrepancy between the results of WS-Parents on precision-recall curves and supervised object detection also supports the need for assessment measures that go beyond the use of horizontal cuts. Finally, we can observe that the three metrics are clearly able to discriminate the baseline method (Random Hierarchy).

6 Conclusion

We have proposed a novel evaluation framework for the assessment 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, watershed hierarchies are competitive with complex state-of-the-art methods for hierarchy construction based on the same gradient information. Moreover, watershed hierarchies are well defined structure satisfying global optimality properties and can be efficiently computed on large data: they are thus valuable candidates for various computer vision tasks.

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

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 [30], [60] on iterative stochastic watershed hierarchies generation or [61] on watershed hierarchies combinations. Another challenge will be to take account for richer multi-scale and oriented gradient information provided by deep learning methods that enabled a large performance boost in COB [58].

Acknowledgments

The authors are grateful to FAPEMIG (PPM-00006-16) and CAPES/PVE (88881.064965/2014-01) for the partial financial support to this work.

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. Cocquerez, 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 *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 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] Y. Xu, E. Carlinet, T. Géraud, and L. Najman, "Hierarchical segmentation using tree-based shape space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. in press, 2016.

[12] P. Salembier and S. Foucher, "Optimum graph cuts for pruning binary partition trees of polarimetric sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5493–5502, 2016.

[13] 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.

[14] 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.

[15] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1155–1162.

[16] 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*, 2004, pp. 428–441.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] J. Cousty, L. Najman, Y. Kenmochi, and S. Guimarães, "Hierarchical segmentations with graphs: Quasi-flat zones, minimum spanning trees, and saliency maps," *Journal of Mathematical Imaging and Vision*, Oct 2017.

[24] 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.

[25] 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.

[26] 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.

[27] B. Marcotegui, "Residual approach on a hierarchical segmentation," in *2014 IEEE International Conference on Image Processing*, 2014, pp. 4353–4357.

[28] 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. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, Eds., vol. 6376. Darmstadt, Germany: Springer Berlin / Heidelberg, 2010, pp. 513–522.

[29] F. Malmberg, C. L. L. Hendriks, and R. Strand, "Exact evaluation of targeted stochastic watershed cuts," *Discrete Applied Mathematics*, vol. In press, 2016.

[30] 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*, D. G. Aliaga, L. S. Davis, L. A. F. Fernandes, and W. R. Schwartz, Eds., 2016.

[31] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille, "The role of context for object detection and semantic segmentation in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[32] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, *Microsoft COCO: Common Objects in Context*. Springer International Publishing, 2014, pp. 740–755.

[33] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[34] 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.

[35] 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.

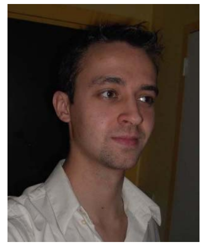
[36] C. Vachier and F. Meyer, "Extinction value: a new measurement of persistence," in *IEEE Workshop on nonlinear signal and image processing*, vol. 1, 1995, pp. 254–257.

[37] M. Grimaud, "New measure of contrast: the dynamics," in *Proc. SPIE, Image Algebra and Morphological Image Processing III*, vol. 1769, 1992, pp. 292–305.

[38] 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.

[39] C. Ronse, "Ordering partial partitions for image segmentation and filtering: Merging, creating and inflating blocks." *Journal of Mathematical Imaging and Vision*, vol. 49, no. 1, pp. 202–233, 2014.

[40] D. Martin, "An empirical approach to grouping and segmentation," Ph.D. dissertation, EECS Department, University of California, Berkeley, 2003.
[41] 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.
[42] 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.
[43] 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.
[44] ——, "Supervised assessment of segmentation hierarchies," in *European Conference on Computer Vision*, 2012.
[45] ——, "Upper-bound assessment of the spatial accuracy of hierarchical region-based image representations," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2012.
[46] 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.
[47] P. P. Michael Maire, Stella Yu, "Hierarchical scene annotation," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013.
[48] E. Aptoula, J. Weber, and S. Lefèvre, "Vectorial Quasi-flat Zones for Color Image Simplification," in *ISMM* , ser. LNCS, vol. 7883, 2013, pp. 231–242.
[49] S. Valero, P. Salembier, and J. Chanussot, "Object recognition in hyperspectral images using binary partition tree representation," *Pattern Recognition Letters*, vol. 56, pp. 45–51, 2015.
[50] 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.
[51] Q. Huang and B. Dom, "Quantitative methods of evaluating image segmentation," in *International Conference on Image Processing*, vol. 3, 1995, pp. 53–56.
[52] J. Serra, *Image analysis and mathematical morphology, v. 1*. Academic press, 1982.
[53] 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.
[54] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results," http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.
[55] K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. V. Gool, "Convolutional oriented boundaries: From image segmentation to high-level tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2017.
[56] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006, ch. 6.
[57] 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," *Mathematical Morphology - Theory and Applications*, vol. 1, pp. 1–22, 2017.
[58] K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. V. Gool, "Convolutional oriented boundaries," in *European Conference on Computer Vision*, 2016.
[59] Q. Zhao, "Segmenting natural images with the least effort as humans," in *BMVC*, 2015, pp. 110.1–110.12.
[60] A. Fehri, S. Velasco-Forero, and F. Meyer, "Automatic Selection of Stochastic Watershed Hierarchies," in *European Conference of Signal Processing (EUSIPCO)*, Budapest, Hungary, 2016.
[61] D. Santana Maia, A. de Albuquerque Araujo, J. Cousty, L. Najman, B. Perret, and H. Talbot, *Evaluation of Combinations of Watershed Hierarchies*. Springer International Publishing, 2017, pp. 133–145.



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 received his Ph.D. degree from the Universidade Federal de Minas Gerais and Université de Marne-la-Valle (France) in 2003. He is now teaching and doing research with the Computer Science Department, at PUC Minas. He is the head of the Audio-Visual Processing Laboratory (VIPLAB) at PUC Minas. His current research interests include image and video analysis and multimedia retrieval.

Deise S. Maia received her bachelor degree in Computer Science from Universidade Estadual do Sudoeste da Bahia (Brazil) in 2014 and her M.Sc. in Computer Science from Universit Paris-Est Marne-le-Valle (France) in 2016. She is currently a PhD student in Image Processing at Universit Paris-Est.