EVALUATING THE QUALITY OF BINARY PARTITION TREES BASED ON UNCERTAIN SEMANTIC GROUND-TRUTH FOR IMAGE SEGMENTATION

Jimmy Francky Randrianasoa[1]*, Camille Kurtz*[2]*, Pierre Gançarski*[3]*, Éric Desjardin*[1]*, Nicolas Passat*[1]

[1] Université de Reims Champagne-Ardenne, CReSTIC, France
[2] Université Paris-Descartes (Sorbonne Paris Cité), LIPADE, France
[3] Université de Strasbourg, CNRS, ICube, France

ABSTRACT

The binary partition tree (BPT) is a hierarchical data-structure that models the content of an image in a multiscale way. In particular, a cut of the BPT of an image provides a segmentation, as a partition of the image support. Actually, building a BPT allows for dramatically reducing the search space for segmentation purposes, based on intrinsic (image signal) and extrinsic (construction metric) information. A large literature has been devoted to the construction on such metrics, and the associated choice of criteria (spectral, spatial, geometric, etc.) for building relevant BPTs, in particular in the challenging context of remote sensing. But, surprisingly, there exists few works dedicated to evaluate the quality of BPTs, i.e. their ability to further provide a satisfactory segmentation. In this paper, we propose a framework for BPT quality evaluation, in a supervised paradigm. Indeed, we assume that ground-truth segments are provided by an expert, possibly with a semantic labelling and a given uncertainty. Then, we describe local evaluation metrics, BPT nodes / ground-truth segments fitting strategies, and global quality score computation considering semantic information, leading to a complete evaluation framework. This framework is illustrated in the context of BPT segmentation of multispectral satellite images.

Index Terms— Binary partition tree, supervised evaluation, uncertainty, semantics, segmentation, mathematical morphology, remote sensing.

1. INTRODUCTION

Image segmentation can be defined under two paradigms: (1) finding one (or many) specific object(s) in an image; and (2) subdividing an image into (spectrally or semantically) homogeneous zones. The second paradigm is particularly relevant for applications that require a global image analysis, for instance in computer vision or remote sensing. However, subdividing an image, i.e. computing a partition of its support, is a challenging task, mainly due to the huge number of possible partitions. In order to tackle this combinatorial issue, a classical solution consists of defining some hierarchies of partitions; the purpose is then to pre-compute a subset of partitions hierarchically organized with respect to the refinement relation.

By assuming that a hierarchy of partitions is "correctly constructed", this strategy then allows for dramatically reducing the search space, and facilitating the segmentation task, without loss of quality of the segmentation result. Indeed, a hierarchy of partitions corresponds to a simple data-structure, i.e. a tree, that can be processed via efficient algorithmic approaches; in particular defining a partition / segmentation is equivalent to defining a cut within the tree.

In this context, several models of hierarchies of partitions have been proposed, many of them in the framework of mathematical morphology [1, Ch. 7]. Most of these hierarchies are *intrinsic* image models, e.g. the component-tree [2] that relies on level-sets, the tree of shapes [3] based on isocontours, or the hierarchical watersheds [4] that rely on saliency (i.e. gradient) measure. By contrast, some *mixed* hierarchies of partitions were also proposed. The most popular is the binary partition tree (BPT) [5] and its variants. BPTs are mixed hierarchies since they rely both on the information embedded in the image content (signal, topological structure) and on extrinsic information, namely a priori knowledge related to the structures of interest, modeled via a metric [6]. Then, both image and metric are involved in the construction of the BPT data-structure.

A large literature has been devoted to the construction of such metrics, and the associated choice of criteria (spectral, spatial, geometric, etc.) for building relevant BPTs, in particular in the challenging context of remote sensing [7, 8, 9, 10]. The design of these metrics (choice and combination of criteria) strongly influences the resulting BPTs, and thus the partitions / segmentations that can be further obtained from them. However, there exists surprisingly very few works devoted to evaluate the quality of BPTs (i.e. their ability to further allow for obtaining a satisfactory segmentation), and then to determine if they were "correctly constructed".

In the case of supervised segmentation evaluation, the results are compared to a ground-truth, composed of reference segments; it is then possible to rely on standard quality indices, e.g. Jaccard index [11], Dice coefficient (a.k.a F-measure) [12, 13], based on spatial overlapping information. From such indices, many frameworks for segmentation quality measures were developed [14, 15, 16, 17, 18, 19], based from example on region or contour-based strategies.

However, in the case of BPTs —and, more generally, hierarchies of partitions— these frameworks can not be easily considered, since several segmentations can be obtained from one BPT. To the best of our knowledge, the only framework for supervised assessment of the quality of a hierarchy of partitions for segmentation purpose, was proposed in [20]. It consists of selecting in the tree a set of segments matching an ideal partition, call upper-bound partition. Such partition is forced to be in the hierarchy and its selection is considered as a linear fractional combinatorial optimization problem.

In this context, we propose a new supervised framework for BPT quality evaluation. Our approach is different from the one proposed in [20], and our contribution is twofold. On the one hand, our purpose is not to match an ideal partition of the image, but to evaluate the ability of a BPT to construct nodes that match at best with a subset of expert-defined segments; in other words, our ground-truth is

This research was partially funded by the French *Agence Nationale de la Recherche* (Grant Agreement ANR-12-MONU-0001).

not required to be global but partial. On the other hand, since the segments are defined by human experts, we assume that they can be imperfect, and we integrate the induced uncertainty in the evaluation process. In addition, these ground-truth segments can be labelled by the expert, then allowing us to embed semantic criteria in the evaluation framework, and to improve the robustness of the global quality score. These different contributions are described in Section 2, where our framework is defined. Section 3 then illustrates its relevance in the context of BPT segmentation of satellite images.

2. BPT QUALITY EVALUATION FRAMEWORK

Our framework for supervised assessment of BPT quality relies on three main components: (i) a local metric that quantitatively evaluates the matching degree between a ground-truth (GT) segment and a node of the BPT (Section 2.2); (ii) the determination of nodes within the BPT that locally maximise matching with the different GT segments (Section 2.3); and (iii) a global quality measure that merges the local scores induced by this optimal cut for each of the GT segments (Section 2.4). The handling of uncertainty of the GT is mainly considered in the local metric of Step (i), while the semantic labelling of the GT segments is mainly used in the definition of the global quality measure of Step (iii).

2.1. Notations and Definitions

An image is defined on a given finite support, i.e. a set of pixels Ω.

A BPT is a hierarchical representation of an image, organized as a binary tree, noted $\mathcal{T}$. Each node N of a BPT is a connected region corresponding to a subset of Ω, i.e. $N \subseteq \Omega$. A node is either a leaf of the tree (i.e. an elementary region, possibly a single pixel) or an internal node (i.e. the union of two nodes modelling two adjacent regions). The root of the BPT is the node corresponding to the entire support Ω of the image. Practically, a BPT is built from its leaves (provided by an initial partition of Ω) to its root, in a bottom-up fashion, by iteratively choosing and merging two adjacent regions which optimize a criterion expressing their likeness. By construction, any cut of the BPT provides a partition of the image support Ω.

A GT segment S provided by the expert is also a subset of the support Ω of the image, i.e. $S \subseteq \Omega$.

2.2. Node / Segment Matching Metrics and Uncertainty

Comparing a node N and a GT segment S consists of quantifying the degree of similarity between them, via the setting of a local score $\Lambda(N, S) \in \mathbb{R}^+$, where Λ is a metric or pseudo-metric that evolves monotonically with respect to this similarity.

Basically, such similarity assessment can be made in a region-based fashion, by considering spatial overlapping information, i.e. by computing the true positives (TP), false positives (FP) and false negatives (FN) between N ans S, given by $TP = |N \cap S|$, $FP = |N \setminus S|$, and $FN = |S \setminus N|$, respectively. The most common quality indices are based on a combination of these three criteria. For instance, the Jaccard index J' [11] is defined as:

$$J'(N, S) = \frac{|N \cap S|}{|N \cup S|} = \frac{TP}{TP + FP + FN} \tag{1}$$

while the Dice coefficient D [12] is expressed as:

$$D(N, S) = \frac{2.|N \cap S|}{|N| + |S|} = \frac{2.TP}{2.TP + FP + FN} \tag{2}$$

The Jaccard and Dice indices J' and D, but also TP, FP and FN (or, possibly, their normalized ratio w.r.t. $|N|$ or $|S|$) are examples of functions Λ. They essentially rely on a region (i.e. combinatorial) paradigm, that consists of "counting" pixels.

Since a GT segment S is assumed to be provided by a human expert, it may be imperfect. This uncertainty mainly comes from its delineation (generally operated by photo-interpretation of the image), i.e. the definition of the contour of the segment. In particular, *the closer a pixel from a segment contour, the less probable its actual correctness*. This statement motivates the computation, for each segment S, of a (signed) distance map [21] to the border of the segment. More precisely, a function $\sigma_S : \Omega \to \mathbb{R}$ is computed, and provides, for each pixel $x \in \Omega$, the distance between x and the border of S. In particular, we set $\sigma_S(x) > 0$ (resp. < 0) outside (resp. inside) S in order to differentiate external and internal pixels.

The degree of uncertainty on S can then be expressed by associating, to each distance value, a probability of correct membership of x to S. Such a membership function is defined as $\mu : \mathbb{R} \to [0, 1]$, with the constraint of being decreasing, and verifying $\lim_{-\infty} \mu = 1$, $\lim_{+\infty} \mu = 0$, and $\mu(0) = 0.5$. The membership function μ considered here is defined as a sigmoid function $\mu_\alpha(d) = 1/(1 + e^{\alpha d})$, where $\alpha > 0$ allows us to control the degree of uncertainty.

Based on this uncertain framework, the notions of true positives, false positives and false negatives can be reformulated[1] as:

$$TP_\alpha(N, S) = \int_N \mu_\alpha(\sigma_S(x)).dx \tag{3}$$

$$FP_\alpha(N, S) = \int_N (1 - \mu_\alpha(\sigma_S(x))).dx \tag{4}$$

$$FN_\alpha(N, S) = \int_{\Omega \setminus N} \mu_\alpha(\sigma_S(x)).dx \tag{5}$$

Practically, the true and false positives can be easily computed as $TP_\alpha(N, S) = \sum_{x \in N} \mu_\alpha(\sigma_S(x))$ and $FP_\alpha(N, S) = |N| - \sum_{x \in N} \mu_\alpha(\sigma_S(x))$. The evaluation of the false negatives requires, in theory, a whole computation over Ω, which is not tractable in practice. By assuming that μ rapidly converges onto 0, it is however sufficient to compute μ in a neighbourhood superset $\Delta S \supseteq S$ of S, and $FN_\alpha(N, S)$ can then be fairly approximated as $\sum_{x \in \Delta S \setminus N} \mu_\alpha(\sigma_S(x))$.

Some uncertain versions of classical scores Λ such as the Jaccard index J' and the Dice index D can be simply obtained by embedding Eqs. (3–5) in Eqs. (1–2). More generally, the use of uncertain notions of true / false positives and false negatives allows us, on the one hand, to take into account the low confidence of the border of the segments and, on the other hand, to introduce some contour-based information into standard region-based metrics, by penalizing mismatching errors far from the GT segment contours.

2.3. Finding Matching Nodes in the Binary Partition Tree

Once defined a similarity metric for nodes / GT segments, the algorithmic question of determining, for each GT segment, what is the most similar node within the BPT, is raised. Given k GT segments S_i ($i \in [\![1, k]\!]$), and a BPT $\mathcal{T}$ composed of n nodes N_j ($j \in [\![1, n]\!]$), we then have to solve k times the following optimization problem

$$N^i_\star = \arg \max_{j \in [\![1,n]\!]} \Lambda(N_j, S_i) \tag{6}$$

A brute-force approach would be to browse, for each GT segment S_i, the whole BPT. This would lead to compute $\Lambda(N_j, S_i)$ for

[1] For $\alpha \to +\infty$, we retrieve the standard notions of TP, FP and FN.

all $(i,j) \in [\![1,k]\!] \times [\![1,n]\!]$ with, at least, a time cost $\mathcal{O}(k.n)$ (the cost of Λ computation also has to be considered). For instance, with a 2D image of size 1000×1000, with an initial set of leaves where each leaf represents a single pixel, the number of nodes n would be $2.1000^2 - 1 \simeq 2.10^6$, leading to a total cost of $2k.10^6$ of Λ computations. It is then crucial to reduce this cost, by avoiding useless node and GT segment comparisons.

To this end, let us recall that our objective is to optimize the matching between a GT segment S and a node N: On the one hand, we want to maximize the intersection between S and N, i.e. $|S \cap N|$, and then get a high value of true positives TP. On the other hand, we want to minimize the mismatchings between S and N, i.e. $|S \setminus N|$ and $|N \setminus S|$, and then get low values of false positives FP and false negatives FN. These two goals are often antagonistic, and classical indices, e.g. Eqs. (1–2), handle the trade-off between them. Nevertheless, this provides us useful information for setting spatial constraints and quantitative heuristics.

Spatial constraints: vertical selection – The intersection between S and N is mandatory to guarantee a non-null value of TP, itself required to avoid null values of standard quality indices as J' or D. A node N that does not intersect S is built, in the BPT, from the union of leaves that present the same property. By contra-position, a node that intersects S is composed by at least one leaf that also intersects S. In particular, for finding an optimal node $N_\star$ it is then sufficient, for a given GT segment S, to restrict our study to the BPT branches whose ending leaves also intersect S. This provides a way to "vertically" restrict the search space within the BPT, by only considering the branches with such leaves (see vertical arrows in Fig. 1).

Quantitative heuristics: horizontal selection – To correctly match a GT segment S, a node N has to be of comparable size. Indeed, if we have $|N| \ll |S|$, the TP value will be low and the FN value high w.r.t. the size of S. On the contrary, if we have $|N| \gg |S|$, the FP value will be high. Consequently, having $|N| \simeq |S|$ is a necessary (yet non-sufficient) condition for obtaining satisfactory matching scores between nodes and segments. Practically, it is then relevant to restrict the actual computation of the Λ values to nodes N with size within a confidence interval $[\kappa^-, \kappa^+] \subseteq \mathbb{N}$, with $\kappa^- < |S| < \kappa^+$. This is indeed a heuristic reduction of the search space, that consists of selecting an "horizontal" set of nodes within the BPT (see horizontal arrows in Fig. 1). In particular, the choice of the κ values has to be wisely made, to handle the trade-off between time cost reduction and near-optimal matching node determination.

These two strategies for selecting candidate nodes to solve Eq. (6) allow for reducing the computational cost. In particular, we consider a bottom-up approach, initialized with the set of leaves intersecting S, and directly climbing the branches up to the first nodes of size κ^-. At this stage, the scores Λ are explicitly computed up to the nodes of size κ^+; then the climbing stops and the score $\Lambda(N_\star, S)$ of the best node is kept as the quality score $\lambda(S)$ of $\mathcal{T}$ for the GT segment S. In certain cases, the properties of the chosen metric Λ could lead to algorithmic optimizations; for instance, separable [22, 23] and, more generally, hierarchically increasing metrics [24] avoid an exhaustive computation of the Λ scores over the whole set of nodes.

2.4. Global Quality Score

By performing this optimization process for each GT segment S_i, we obtain a set of k nodes $N^i_\star$, associated with a (near-)optimal quality score $\lambda(S_i) = \Lambda(N^i_\star, S_i)$. The issue is then to gather the information provided by these local quality scores to finally define a global quality metric Γ that will express the quality of the BPT, i.e. its ability to fit at best the set of GT segments.

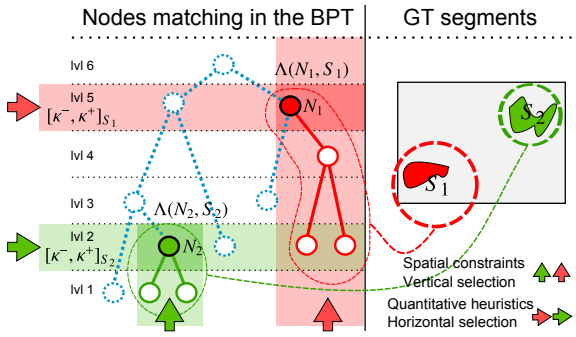


Fig. 1. Illustration of the spatial constraints (vertical selection) and quantitative heuristics (horizontal selection) employed to optimize the search of matching nodes in the BPT.

The very first idea for defining Γ is indeed to compute an average value of all the local quality scores, i.e. $\Gamma = 1/k. \sum_{i=1}^{k} \lambda(S_i)$. However, since we assume that each GT segment is endowed with a given label, it is important to consider this semantic information in order to avoid potential bias effects. For instance, if 25% of the GT segments have a label A, while the other 75% have a label B, a simple mean value Γ for a BPT with an average quality value of 0.1 for the class A and 0.9 for the class B will lead to a global quality score of $\frac{1}{4} \times 0.1 + \frac{3}{4} \times 0.9 = 0.7$, which is non-relevant if both classes have comparable importance, that should lead to a 0.5 global score. It is then crucial to model in the global measure the relative importance of each label / class of segment. In addition, it may be also relevant to model the relative importance of each segment within each class.

To this end, we define a weighted formulation of the global quality score Γ as

$$\Gamma = \sum_{\ell \in L} w_\ell \sum_{S_i \in \mathcal{C}_\ell} w_i.\lambda(S_i) \qquad (7)$$

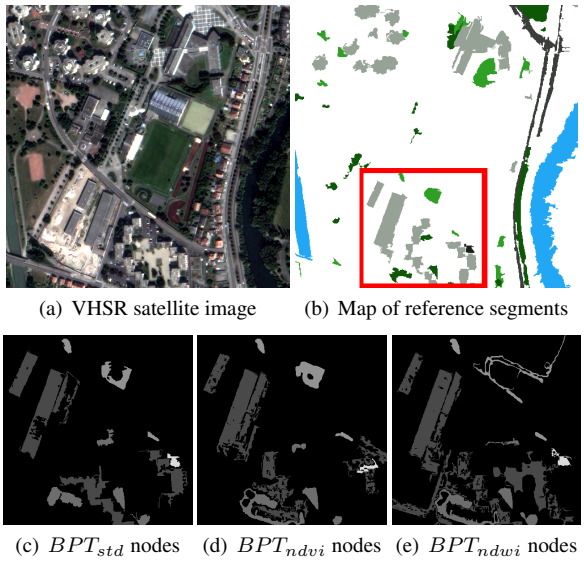
with $\sum_{\ell \in L} w_\ell = 1$, $\sum_{S_i \in \mathcal{C}_\ell} w_i = 1$, and $w_\star \geq 0$, where L is the label set and $\mathcal{C}_\ell$ are the different semantic classes of GT segments.

The weights w_ℓ can be used to assess the relative importance of each semantic class $\mathcal{C}_\ell$. In particular, $w_\ell = 1/|L|$ if each class has the same importance. The weights w_i can be used for normalizing the local quality metric (e.g. in the case of extensive metrics as TP), and / or to discriminate the importance of each GT segment, i.e. , the necessity to correctly segment it. They can also be used for quantifying the relevance of a GT segment; for instance, if these segments are obtained from a crowdsourcing campaign, the w_i weight can be proportional to the confidence assigned to S_i.

It is worth mentioning that the use of semantic information for designing the global quality metric also argues in favour of the possible design of non-linear definitions of Γ. Two specific formulations can be proposed: $\Gamma_{\min} = \min_{\ell \in L} \sum_{S_i \in \mathcal{C}_\ell} w_i.\lambda(S_i)$ and $\Gamma_{\max} = \max_{\ell \in L} \sum_{S_i \in \mathcal{C}_\ell} w_i.\lambda(S_i)$. The first allows to characterize BPTs that are able to efficiently characterize *all* the classes of objects provided as ground-truth; the second provides a way to discriminate BPTs that detect (at least) one among a set of given classes.

3. EXPERIMENTAL STUDIES

To illustrate our framework, we involved it in the domain of remote sensing image analysis, and in particular for the segmentation of very-high spatial resolution (VHSR) satellite images, where the use



(a) VHSR satellite image (b) Map of reference segments

(c) BPT_{std} nodes (d) BPT_{ndvi} nodes (e) BPT_{ndwi} nodes

Fig. 2. (a) VHSR image 1 (1000 × 1000 pixels) at a spatial resolution of 60 cm. (b) GT map with reference segments belonging to 6 (coloured) semantic classes. (c) Matched nodes from the standard BPT. (d) Matched nodes from the BPT built using the NDVI metric. (e) Matched nodes from the BPT built using the NDWI metric.

of BPTs has been widely considered [7, 8, 9, 10]. Our purpose is mainly to give the intuition of potential uses of this framework for evaluating the quality of BPTs in complex imaging domains.

3.1. Data

Our dataset (courtesy LIVE, UMR CNRS 7263) contains two VHSR images (1000 × 1000 pixels), sensed over the town of Strasbourg (France) by the PLÉIADES satellite. These pansharpened multispectral images have a spatial resolution of 60 cm with four spectral bands (R, G, B, NIR). They represent high-density urban areas composed of typical geographical objects. The first image of the dataset is presented in Fig. 2 (a). A ground-truth map of different urban objects represented in the scene is also available; this map has been derived from a public crowd-sourcing campaign in the context of the COCLICO research project. From this campaign, we have only retained the reference segments that led to the highest consensus between the crowders. The segments are labelled with $L = 6$ different semantic classes (*built area*, *forest area*, *herbaceous area*, *roads*, *shadow*, *water*). Fig. 2 (b) presents the GT map of the first image.

3.2. Method and results

From the two images, different BPTs were built by choosing three simple construction metrics, to avoid any bias related to this choice. The first is a standard "colour" metric, defined as the increase of the ranges of the pixel intensity values for each radiometric band, potentially induced by the fusion of incident regions. The second relies on the difference of the NDVI (Normalized Difference Vegetation Index) between two adjacent regions, emphasizing vegetated areas, while the third relies on the NDWI (Normalized Difference Water Index), emphasizing water surfaces.

We used our framework to evaluate the quality of these three BPTs relatively to the reference segments of the GT maps. Our framework was parametrized as follows. To compare nodes and GT segments (Section 2.2), we used the uncertain versions of the Jaccard index J' and the Dice coefficient D. The membership function μ was computed by using $\alpha = 1.2$. As quantitative heuristics to reduce the search space of matching nodes (horizontal selection, Section 2.3), we defined κ as a function depending on the size of the GT segment S to be matched, by only evaluating BPT nodes whose size is included in an interval $[0.5 \times |S|, 1.5 \times |S|]$. Finally, the weights involved in the computation of the global quality scores (Section 2.4) were fixed as $w_\ell = 1/|L|$ and $\omega_i = |S_i| / \sum_{S_j \in \mathcal{C}_\ell} |S_j|$ for a GT segment S_i belonging to the semantic class C_ℓ.

Image	Index	Std	NDVI	NDWI	N/S	Time (s)
1	D	**0.632**	0.511	0.566	46/48	387
	J'	**0.480**	0.369	0.430	46/48	378
2	D	**0.670**	0.523	0.516	51/51	450
	J'	**0.531**	0.389	0.399	51/51	484

Table 1. Global quality scores of BPT_{std}, BPT_{ndvi} and BPT_{ndwi} from two VHSR images (1000×1000 pixels). N/S: number of BPT nodes retrieved according to the number of reference segments.

Table 1 presents the global scores obtained, assessing the quality of the different BPTs. We notice that the best scores were obtained from the BPT_{std} built upon a standard "colour" metric. Indeed, despite the presence of some vegetation and water segments in the images, the scene is mainly composed of artificial objects such as buildings, leading to lower results for the BPT_{ndvi} and the BPT_{ndwi}. This result seems to be confirmed by Fig. 2 (c, d, e) showing that the matched nodes for the BPT_{std} are more similar to the GT segments (see red crop, Fig. 2 (b)) than those from the other BPTs. From Table 1, we also remark that two GT segments (out of 48) were not matched for the first image. This result is due to the quantitative heuristics applied during the research that ignored the nodes having their size not included in the restricted interval. If no restrictions were made in the optimization strategy, we could have found them with very low matching rates (nodes too large or too small) that may impair the global score. Finally, to illustrate the interest of reducing the search space of matching nodes (Section 2.3), we provide the average computation times required to evaluate each BPT. By coupling the two proposed optimization strategies to speed-up the selection of candidate nodes, the required computation times is 387 *s* while 28312 *s* are needed by a brute force search in the BPT.

4. CONCLUSION

In this paper, we proposed a framework for evaluating the quality of the BPTs in a supervised way. A global quality score of a BPT is computed taking into consideration the uncertainty / semantic information of a given GT. The experiments highlighted the usefulness of our framework in the context of remote sensing image analysis. The optimization strategy proposed to reduce the search space of matching BPT nodes is computationally helpful but may slightly bias the results and should be used carefully. In addition, the choice of the GT segments should lead to a fair evaluation. Finally, this work can be generalized for other hierarchical structures to help their comparison. In other works, we recently proposed a new method for building BPTs relying on a multi-feature / multi-image paradigm [10, 25]. As perspective, we plan to use this evaluation framework to help the parameter choices related to the construction of these new BPTs, leading to improvement of image segmentation results.

5. REFERENCES

[1] L. Najman and H. Talbot, Eds., *Mathematical Morphology: From Theory to Applications*, ISTE / J. Wiley & Sons, 2010.

[2] P. Salembier, A. Oliveras, and L. Garrido, "Anti-extensive connected operators for image and sequence processing," *IEEE Transactions on Image Processing*, vol. 7, pp. 555–570, 1998.

[3] P. Monasse and F. Guichard, "Scale-space from a level lines tree," *Journal of Visual Communication and Image Representation*, vol. 11, pp. 224–236, 2000.

[4] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1163–1173, 1996.

[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, pp. 561–576, 2000.

[6] P. Soille, "Constrained connectivity for hierarchical image partitioning and simplification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1132–1145, 2008.

[7] S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image representation and processing with binary partition trees," *IEEE Transactions on Image Processing*, vol. 22, pp. 1430–1443, 2013.

[8] 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.

[9] C. Kurtz, N. Passat, P. Gançarski, and A. Puissant, "Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology," *Pattern Recognition*, vol. 45, pp. 685–706, 2012.

[10] J. F. Randrianasoa, C. Kurtz, É. Desjardin, and N. Passat, "Multi-image segmentation: A collaborative approach based on binary partition trees," in *International Symposium on Mathematical Morphology (ISMM)*, 2015, vol. 9082 of *Lecture Notes in Computer Science*, pp. 253–264.

[11] P. Jaccard, "The distribution of the flora in the Alpine zone," *New Phytologist*, vol. 11, pp. 37–50, 1912.

[12] L.R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297–302, 1945.

[13] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons," *Biologiske Skrifter*, vol. 5, pp. 1–34, 1948.

[14] M. Polak, H. Zhang, and M. Pi, "An evaluation metric for image segmentation of multiple objects," *Image and Vision Computing*, vol. 27, pp. 1223–1227, 2009.

[15] H. Vojodi and A.M.E. Moghadam, "A supervised evaluation method based on region shape descriptor for image segmentation algorithm," in *International Symposium on Artificial Intelligence and Signal Processing (AISP)*, 2012, pp. 18–22.

[16] J. Pont-Tuset and F. Marques, "Measures and meta-measures for the supervised evaluation of image segmentation," in *Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2131–2138.

[17] H. Li, J. Cai, T.N.A. Nguyen, and J. Zheng, "A benchmark for semantic image segmentation," in *International Conference on Multimedia and Expo (ICME)*, 2013, pp. 1–6.

[18] 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. 39, pp. 128–140, 2017.

[19] 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, pp. 1465–1478, 2016.

[20] J. Pont-Tuset and F. Marques, "Supervised assessment of segmentation hierarchies," in *European Conference on Computer Vision (ECCV)*, 2012, vol. 7575 of *Lecture Notes in Computer Science*, pp. 814–827.

[21] R. Fabbri, L.D.F. Costa, J.C. Torelli, and O.M. Bruno, "2D Euclidean distance transform algorithms: A comparative survey," *ACM Computing Surveys*, vol. 40, pp. 1–44, 2008.

[22] L. Guigues, J.-P. Cocquerez, and H. Le Men, "Scale-sets image analysis," *International Journal of Computer Vision*, vol. 68, pp. 289–317, 2006.

[23] N. Passat, B. Naegel, F. Rousseau, M. Koob, and J.-L. Dietemann, "Interactive segmentation based on component-trees," *Pattern Recognition*, vol. 44, pp. 2539–2554, 2011.

[24] J. Serra, "Hierarchies and optima," in *Discrete Geometry for Computer Imagery (DGCI)*, 2011, vol. 6607 of *Lecture Notes in Computer Science*, pp. 35–46.

[25] J. F. Randrianasoa, C. Kurtz, É. Desjardin, and N. Passat, "Binary partition tree construction from multiple features for image segmentation," Tech. Rep. hal-01248042, 2016.