

Improved Partition Trees for Multi-Class Segmentation of Remote Sensing Images

Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat

▶ **To cite this version:**

Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat. Improved Partition Trees for Multi-Class Segmentation of Remote Sensing Images. 2015 IEEE International Geoscience and Remote Sensing Symposium - IGARSS 2015, IEEE, Jul 2015, Milan, Italy. hal-01182772

HAL Id: hal-01182772
https://inria.hal.science/hal-01182772

Submitted on 3 Aug 2015

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

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

IMPROVED PARTITION TREES FOR MULTI-CLASS SEGMENTATION OF REMOTE SENSING IMAGES

Emmanuel Maggiori[1]*, Yuliya Tarabalka*[1] *and Guillaume Charpiat*[2]

[1] Inria Sophia Antipolis Méditerranée, TITANE team; [2] Inria Saclay, TAO team, France
Email: emmanuel.maggiori@inria.fr

ABSTRACT

We propose a new binary partition tree (BPT)-based framework for multi-class segmentation of remote sensing images. In the literature, BPTs are typically computed in a bottom-up manner based on spectral similarities, then analyzed to extract image objects. When image objects exhibit a considerable internal spectral variability, it often happens that such objects are composed of several disjoint regions in the BPT, yielding errors in object extraction. We pose the multi-class segmentation problem as an energy minimization task and solve it by using BPTs. Our main contribution consists in introducing a new dissimilarity function for the tree construction, which combines both spectral discrepancies and supervised class-specific information to take into account the within-class spectral variability. The experimental validation proved that the proposed method constitutes a competitive alternative for object-based image classification.

Index Terms— Binary partition tree, segmentation, object-based classification, energy minimization.

1. INTRODUCTION

Recent advances in remote sensing image classification show a clear tendency toward the inclusion of spatial information in the algorithms [1]. This has been proven to enhance the traditional pixel-based approach, in which the image is viewed as an unordered array of spectra and no information about the interactions among the pixels is exploited. The spatial information typically includes adjacency relations and properties of entire regions at different scales.

Binary partition trees (BPTs) [2] constitute an efficient alternative to represent images in a spectral-spatial manner and have been intensively used to process color, multispectral and hyperspectral imagery [1, 3]. A BPT is a hierarchical partition of an image: the root node represents the entire image, the following level partitions the image into two non-overlapping regions, and so on. The following strengths of the BPTs make them very suitable for image representation and analysis:

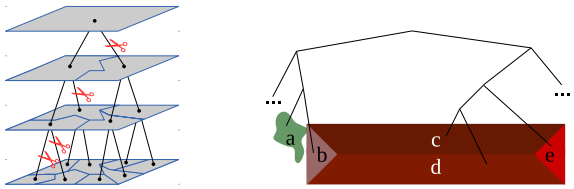
- The nodes of the tree can contain rich information about the corresponding regions at different scales, including spectral histograms and shape information.
- The computational complexity required to construct the BPT is low.
- Most processing later done on them (including the extraction of objects and segmentations) is also efficient.

Recent works have proposed to use BPTs for classification [1] and object detection [4]. Even though these methods enhance the pixel-wise classification accuracy, an object-based analysis of the output has not been carried out. It is known that the BPT construction approach might lead to trees that do not represent objects in an accurate way [2].

In this work, we address the problem of multi-class segmentation, which consists in an exhaustive partitioning of the pixels into a non-overlapping set of regions $\mathcal{R} = (R_i)$, with associated class labels $\mathcal{L} = (L_i)$. We aim to detect regions that correspond to actual objects in the images, as well as to classify them accurately. For this purpose, we pose multi-class segmentation as an energy minimization problem and explore the use of BPTs to solve it (Sec. 2-3). After an evaluation of the shortcomings with the traditional approach, we propose a new BPT construction function that encourages BPTs to better represent entire objects in single nodes (Sec. 4), hence improving object-based classification results.

2. BUILDING A BPT

BPTs are constructed in a bottom-up manner, by recording the history of merges when applying a region merging algorithm. Starting from a fine partition (usually one region per pixel), the most similar pair of regions is merged in every iteration, until there is only one region left. In every step, a new node is inserted into the tree, with two child pointers to the regions that were merged. The BPT construction requires to define how a region is represented, the so-called region *model*, and how the regions are to be compared, the so-called merging *order*. The latter is a dissimilarity function between two models, denoted by $O(R_1, R_2)$. The time required to run the full region merging algorithm is $\mathcal{O}(n \log(n) k)$ [5], n being the size of the input image and k the maximum number of neighbors of a region during the execution. Provided that $k \ll n$, the BPT construction time can be considered quasilinear in practice. The required storage is linear in the size of the input image.



(a) Segmentation by cutting branches on a BPT (b) Faulty BPT: an object not represented in a single node

Fig. 1: Binary partition trees.

Many research efforts have been focused on the definition of the region model and order function. Notably, different alternatives in the context of grayscale image segmentation [3] and remotely sensed hyperspectral data [1] have been studied. The main observations suggest that:

- A non-parametric region model (*e.g.,* per-band histograms) is a better approach than its parametric counterpart (*e.g.,* average spectrum), given that the internal variability of every region is represented.
- A cross-bin dissimilarity function to compare the histograms is a better alternative than a bin-per-bin distance (*e.g.,* Euclidean), in order to cope with possibly variable lighting conditions or histogram "shifts".

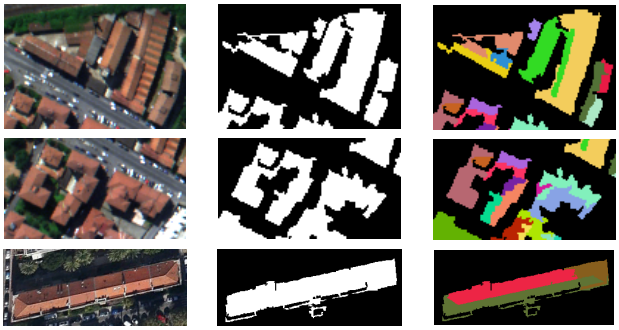
3. MULTI-CLASS SEGMENTATION WITH BPTS

We formulate multi-class segmentation as an energy minimization problem and describe the algorithm to extract the optimal partition from a BPT with respect to that energy. Let us consider a B-band image as a set of n pixel vectors $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^B, i = 1, 2, ..., n\}$. Our task is to find the labeled partitioning $(\mathcal{R}, \mathcal{L})$ that minimizes:

$$E(\mathcal{R}, \mathcal{L}) = \lambda||\mathcal{R}|| - \sum_{R_j \in \mathcal{R}} \sum_{\mathbf{x}_i \in R_j} \log P(L_j|\mathbf{x}_i), \qquad (1)$$

where $P(L_j|\mathbf{x}_i)$ is a posterior probability data term, which can be derived from a probabilistic support vector machine (SVM) [6]. λ is a regularization parameter to select the coarseness of the partition, $||\mathcal{R}||$ being the number of regions.

On a given BPT, the partitioning, or *cut* (see Fig. 1a), that optimizes (1) can be found in one bottom-up traversal of the tree, by using a dynamic programming approach [7]. Let us denote $\mathcal{E}(R)$ the lowest possible energy of a given region R (by selecting its best label), and $C(R)$ the energy of the cut under R that optimizes (1). At the pixel level, the optimal partition is the pixel itself and the label is the one with the lowest energy. Let us call R_{left} and R_{right} the children of R. At every step of the traversal, the following property is assessed: $\mathcal{E}(R) < C(R_{\text{left}}) + C(R_{\text{right}})$. If it does not stand, we chose to keep the best partitions under the children separated, and set $C(R) = C(R_{\text{left}}) + C(R_{\text{right}})$. Otherwise, we replace the partitions under R_{left} and R_{right} by R, and set $C(R) = \mathcal{E}(R)$. The traversal finishes when we compute $C(\text{root})$, *i.e.* the optimal partition of the whole image.



(a) Original image (b) Areas labeled *tile* (c) Extracted objects

Fig. 2: Detection of tile buildings on a typical BPT. The extracted objects hardly correspond to actual buildings.

4. NEW DISSIMILARITY FUNCTION

As a consequence of inaccuracies in the BPT construction, it is known that nodes in the tree might end up not representing actual objects in the image [2]. Our experiments supported this observation since we found that faulty node/object correspondences are very prevalent. We here illustrate some examples, in which the following dissimilarity function was used:

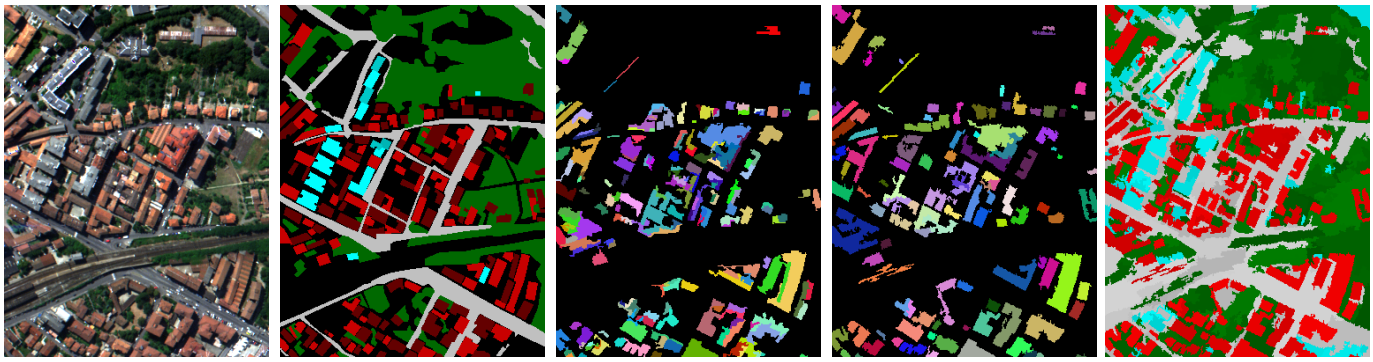
$$O(R_1, R_2) = \min(|R_1|, |R_2|)^{\frac{1}{2}} D(R_1, R_2), \qquad (2)$$

where the first factor is a typical area-weighting term to encourage the balancing of the tree [3], and D is the average of Earth Mover's Distances [8] among band histograms. This is an efficient cross-bin histogram difference measure, which has already been used in the context of BPTs [9].

Once the BPT was constructed, we applied the proposed energy minimization approach to extract a multi-class segmentation. From this segmentation, we have isolated the objects assigned to the class *tiles* to aid the interpretation. Fig. 2 shows the results. The first two rows correspond to a fragment of a hyperspectral image over Pavia, and the last one to a fragment of a color pansharpened image over Nice. Note that if we observe the area covered by the detected objects, the results are quite accurate from a pixel-wise perspective (Fig. 2b). In previous remote sensing works (*e.g.,* [4]), the performance has been analyzed using similar pixel-based approaches. However, from an object-based perspective (Fig. 2c), the extracted regions tend to be oversplit and hardly correspond to the actual buildings.

To better understand this behavior, we proceeded analyzing the resulting BPTs. We observed that often a part of an object merges first to other regions than to the rest of the object itself, which is amplified when a class features a considerable internal spectral variability. We illustrate this behavior in Fig. 1b. This example contains a non-uniform roof. During the tree construction, a part of the roof (b) is merged first to something else (a) rather than to the rest of the object (c-d-e).

Let us now remember that the purpose of using non-parametric models to represent regions is to cope with the



(a) Color composition (b) Reference data: tiles, bitumen, roads , veget. (c) Tile building detection (BPT with $\alpha = 0$) (d) Tile building detection (BPT with $\alpha = 0.5$) (e) Classification map (BPT with $\alpha = 0.5$)

Fig. 3: *Center of Pavia* image.

internal variability. However, we must point out that a typical order function such as (2) always penalizes the merging of dissimilar regions. Therefore, variability can be represented but is discouraged.

In this work, we propose to augment this function by including a force that clusters regions belonging to the same class, despite being spectrally dissimilar. The new function we propose is as follows:

$$O(R_1, R_2) = \min(|R_1|, |R_2|)^{\frac{1}{2}} \Big[(1-\alpha) D(R_1, R_2) - \alpha \log P(L_{R_1} = L_{R_2}) \Big]. \quad (3)$$

As in the typical order function (2), we include an area-weighting factor and an unsupervised term $D(R_1, R_2)$, which is computed by comparing spectral histograms of regions without any preliminary training. In our formulation, we add a *supervised* term $P(L_{R_1} = L_{R_2})$, the probability of assigning the same label to both regions. This way, while the unsupervised term penalizes spectral dissimilarity, the supervised term will encourage to merge regions that are likely to belong to the same class. The trade-off between both terms is controled by a parameter α.

An SVM is first run to compute per-pixel class probabilities $P(L_j|\mathbf{x}_i), j = 1...K$, where K is the number of classes. The supervised term $P(L_{R_1} = L_{R_2})$ is then estimated as follows:

1. Estimate class probabilities for an entire region:
$$P(L_j|R_k) = \frac{1}{|R_k|} \sum_{\mathbf{x}_i \in R_k} P(L_j|\mathbf{x}_i), k = \{1, 2\},$$
2. $P(L_{R_1} = L_{R_2}) = \sum_{j=1}^{K} P(L_j|R_1) P(L_j|R_2)$.

SVM with kernels have the capability of dealing with internal class variability. Now that our new function (3) is capable of clustering parts of objects that are correctly identified by the SVM, we expect to enhance the performance of the BPTs at representing entire objects in single nodes.

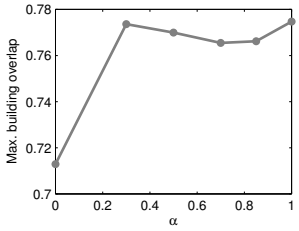


Fig. 4: The new order function enhances BPT construction.

5. EXPERIMENTAL RESULTS

The proposed method was applied to the *Center of Pavia* image acquired by the ROSIS sensor over the city of Pavia, Italy. It is of size 400×300 pixels, with a spatial resolution of 1.3 m/pixel and 102 spectral bands. A color composition of the image is shown in Fig. 3a. A detailed reference image was constructed by combining visual inspection and cadastral records (Fig. 3b). It contains four classes: tile buildings, bitumen buildings, roads and vegetation.

Overall accuracy (the proportion of correctly classified points) is used as a measure of classification quality. Dice's coefficient is used to measure the overlap between a region in the tree/map and a reference object: $\frac{2|R_1 \cap R_2|}{|R_1| + |R_2|}$.

In the first experiment, we assess the accuracy of the BPT construction. For every *tile building* object in the reference data, we search for the most overlapping region in the whole BPT. The average overlap among all reference buildings (117 elements) is used as a measure of the maximum possible object detection performance that can be achieved by cutting the tree. We carried out the experiment for different values of α, in order to evaluate the relevance of using the class probabilities in the BPT construction dissimilarity function. The results are depicted in Fig. 4. We can observe that the inclusion of the new term in (3), *i.e.* when $\alpha > 0$, improves the BPT quality. In addition, larger values of α (even when ignoring the histograms, *i.e.* $\alpha = 1$) enhance the results.

Criterion / Method	SVM	GC	BPT $\alpha = 0$	BPT $\alpha = 0.5$
Building overlap	0.51	0.51	0.54	**0.56**
Overall accuracy	0.88	**0.94**	0.91	**0.94**

Table 1: Results for the *Center of Pavia* image.

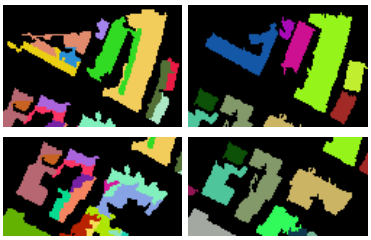


Fig. 5: Building detection. *Left:* $\alpha = 0$, *Right:* $\alpha = 0.5$.

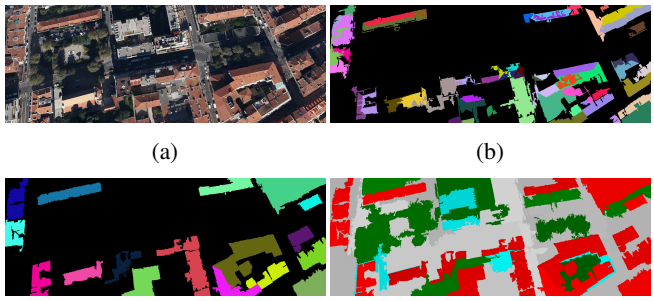
In the second experiment, we produce multi-class segmentations. We compared the following methods: SVM with Gaussian kernel ($c = 128, \gamma = 2^{-5}$, adjusted by 5-fold cross validation), graph-cut with α-expansion [10] (which has proven to outperform the recent multi-label classification approaches [11]), and the proposed BPT cut (λ was empirically set to 20) with $\alpha = 0$ and $\alpha = 0.5$ (to compare Eqs. 2 and 3). Numerical results are shown in Table 1. It can be observed that the proposed method yields the most competitive results. Even though a similar overall accuracy is attained by an advanced technique such as graph-cut, our BPT-based technique outperforms it in terms of tile building overlap, thus proving a better per-object performance. This is supported visually by comparing the building regions in Figs. 3c-d. Some fragments have been amplified in Fig. 5. An overall classification map is depicted in Fig. 3e.

Fig. 6 illustrates our technique on a pansharpened color satellite image over Nice, France, extracted from Google Maps. Our dissimilarity function extracts more accurate tile building objects.

6. CONCLUDING REMARKS

Binary partition trees have been studied for classification in the context of remote sensing imagery. However, it is known that the errors in the tree construction induce errors in the BPT-based object extraction. We have posed the problem of multi-class segmentation as an energy minimization task, which we have solved by using BPTs. We have proposed a new dissimilarity function for tree construction, which enables to output regions that better correspond to real objects, as verified by our experiments.

Now that the nodes in the trees tend to include complete objects, we are exploring the use of shape features and the BPT optimization, in order to further enhance the object-based classification results.



(a) (b) (c) (d)

Fig. 6: Image of *Nice*. (a) Color. (b) Tiles with BPT ($\alpha = 0$). (c) Tiles with BPT ($\alpha = 0.5$). (d) Classification map.

7. REFERENCES

[1] Silvia Valero, Philippe Salembier, and Jocelyn Chanussot, "Hyperspectral image representation and processing with binary partition trees," *IEEE TIP*, vol. 22, no. 4, pp. 1430–1443, 2013.

[2] Veronica Vilaplana, Ferran Marques, and Philippe Salembier, "Binary partition trees for object detection," *IEEE TIP*, vol. 17, no. 11, pp. 2201–2216, 2008.

[3] Felipe Calderero and Ferran Marques, "Region merging techniques using information theory statistical measures," *IEEE TIP*, vol. 19, no. 6, pp. 1567–1586, 2010.

[4] Silvia Valero, Philippe Salembier, and Jocelyn Chanussot, "Object recognition in urban hyperspectral images using binary partition tree representation," in *IGARSS*, 2013.

[5] T Kurita, "An efficient agglomerative clustering algorithm for region growing," in *IAPR Workshop on Machine Vision Applications*, 1994, pp. 210–213.

[6] Yuliya Tarabalka, Jocelyn Chanussot, and Jón Atli Benediktsson, "Segmentation and classification of hyperspectral images using MSF grown from automatically selected markers," *IEEE SMC*, vol. 40, no. 5, 2010.

[7] Philippe Salembier, Samuel Foucher, and Carlos López-Martínez, "Low-level processing of PolSAR images with binary partition trees," in *IGARSS*, 2014.

[8] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas, "A metric for distributions with applications to image databases," in *ICCV*, 1998, pp. 59–66.

[9] Guillem Palou and Philippe Salembier, "Occlusion-based depth ordering on monocular images with binary partition tree," in *ICASSP 2011*, pp. 1093–1096.

[10] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *IEEE TPAMI*, vol. 23, no. 11, pp. 1222–1239, 2001.

[11] Yuliya Tarabalka and Aakanksha Rana, "Graph-cut-based model for spectral-spatial classification of hyperspectral images," in *IGARSS*, 2014.