

# CONTEXT-BASED ENERGY ESTIMATOR: APPLICATION TO OBJECT SEGMENTATION ON THE TREE OF SHAPES

Yongchao Xu<sup>1,2</sup>, Thierry Géraud<sup>1,2</sup>

Laurent Najman<sup>2</sup>

<sup>1</sup> EPITA Research and Development Laboratory  
(LRDE)  
14-16, rue Voltaire  
FR-94276 Le Kremlin-Bicêtre, France

<sup>2</sup> Université Paris-Est,  
Laboratoire d'Informatique Gaspard-Monge (LIGM),  
A3SI, ESIEE Paris, Cité Descartes, BP 99  
FR-93162 Noisy-le-Grand, France

## ABSTRACT

Image segmentation can be defined as the detection of closed contours surrounding objects of interest. Given a family of closed curves obtained by some means, a difficulty is to extract the relevant ones. A classical approach is to define an energy minimization framework, where interesting contours correspond to local minima of this energy. Active contours, graph cuts or minimum ratio cuts are instances of such approaches. In this article, we propose a novel efficient ratio-cut estimator which is both context-based and can be interpreted as an active contour. As a first example of the effectiveness of our formulation, we consider the tree of shapes, which provides a family of level lines organized in a tree hierarchy through an inclusion relationship. Thanks to the tree structure, the estimator can be computed incrementally in an efficient fashion. Experimental results on synthetic and real images demonstrate the robustness and usefulness of our method.

**Index Terms**— Ratio-cut, Tree of shapes, Level lines, Active contours, Image segmentation.

## 1. INTRODUCTION

In natural images, significant contours are usually smooth and have a good contrast. Following the seminal work of Mumford and Shah [1], finding contours is often tackled thanks to an energy-based approach, as a compromise between some internal force (regularity) and some image-driven force (image contrast along the contours, data attachment, etc.) Among the most notable methods in this class, the snakes approach [2] finds a local optimum of the energy, while Chan-Vese approach [3] finds a global optimum.

It can be observed in practice that local optima are usually too local, in the sense that only information along the curve are taken into account. Conversely, a global optimum takes the whole image into account and some details can be found difficult to be included in the final result. Let us also mention that the search space for the optimal curve is quite large and depends on many parameters.

The main contribution of this paper is the design of a ratio-cut energy evaluator that takes into account some context along the curve, by looking at a couple of regions around it. This estimator aims at assessing the possibility that the curve under scrutiny is the contour of some object. It thus can be used to choose, in a set of closed curves, for the ones that best represent the objects of the scene. If the curves are organized in a tree, as it is the case for example in any hierarchical segmentation method [4], the search space is reduced, and it is then easier to find the objects: they are the minima of the estimator, when compared to their children and parent.

Recently, many authors (see for example [5]) claim that meaningful contours coincide with segments of the level lines of the image. The inclusion relationship of these level lines gives birth to the tree of shapes [6], a complete representation of the image that is invariant to changes of contrast. In this paper, we illustrate the soundness of our estimator on this kind of tree.

There exists several works that use the tree of shapes for image simplification and segmentation. In [7], the authors propose to remove any level lines that, when removed, decrease the Mumford-Shah energy. In [8], the authors remove the level lines that enclose a region similar to its parent w.r.t. an histogram-based distance. They then select interesting regions by identifying parts of the tree having an *homogeneous* histogram. Both methods require an information update after each removal of a non-relevant level line; such an update mechanism is a major bottleneck of those methods. In [9], the author proposes a segmentation algorithm that selects the perceptible level lines matching some criteria: number of T-junctions, compactness, and contrast.

Papers [10] and [11] are the closest ones to what we propose here. The authors define the *meaningfulness* of a given level line using the *a contrario* model. Only the smallest gradient along a level line and its length are used to estimate the *meaningfulness* of this line, based on the computation of a number of false alarms (we will call this method “NFA” in the following). This makes their estimator sensitive to noise. In [11], they improve the method by introducing a multi-scale approach, less sensitive to noise. In this paper, our estima-

tor is scale-invariant and designed from snake-like principles. We use the average of curvature along the curve as the internal force. A key contribution is the image force we propose, which integrates some contextual information. Eventually, we show that our estimator is robust both to noise and blur. Another advantage of our method is its efficiency for it has a quasi-linear time complexity.

The rest of this paper is organized as follows. Our proposed estimator is described in Section 2. How to use it on the tree of shapes is detailed in Section 3. In Section 4, we show some experimental results on synthetic and real images, and we conclude in Section 5.

## 2. CONTEXT-BASED ENERGY ESTIMATOR

The idea of snakes or active contours [2] is to evolve a curve under the influence of internal forces coming from the curve itself and of external forces computed from the input image or added by the user to impose some constraints. Those forces are mapped into the respective terms of an energy  $E^{local} = \alpha E_{int}^{snk} + E_{ext}^{snk} + \beta E_{con}$  whose minimization drives the curve. Given an input image  $u$  and a curve  $\partial\tau$  (contour of a region  $\tau$ , composed of pixel edges  $e$ ), those energy terms are:

$$E_{int}^{snk} = \sum_{e \in \partial\tau} |curv(u)(e)| \quad \text{and} \quad E_{ext}^{snk} = - \sum_{e \in \partial\tau} |\nabla u(e)|,$$

where  $curv$  and  $\nabla$  are respectively the curvature and the gradient of  $u$ . This energy is thus *local* to the curve. Other approaches such as [3] show that, thanks to a variational approach, the curve can minimize a *global* energy that can be for instance the cartoon segmentation model [1]. Given a region  $\mathcal{R}$ , let us consider the segmentation error:

$$V(u, \mathcal{R}) = \sum_{p \in \mathcal{R}} (u(p) - \bar{u}(\mathcal{R}))^2$$

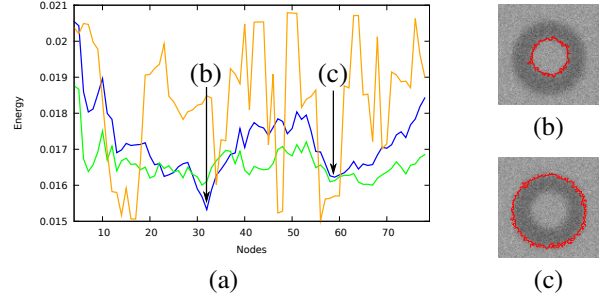
where  $\bar{u}(\mathcal{R})$  is the mean value of  $u$  over  $\mathcal{R}$ . When the image is segmented into the set of regions  $\{\mathcal{R}_i\}$ , the cartoon functional is:

$$E^{global}(u, \{\mathcal{R}_i\}) = \sum_i V(u, \mathcal{R}_i) + \nu L(\partial\mathcal{R}_i),$$

with  $L$  denoting the length of a contour.

Suppose now that we have decomposed the input image  $u$  into a set of regions  $\{\tau\}$ . We need an energy that can be computed for every regions. Such an energy can express two kinds of information: a local one, associated with the region contour  $\partial\tau$ , such as the smoothness term  $E_{int}^{snk}$ , and a *boundary* one such as the gradient term  $E_{ext}^{snk}$  of the snake energy. We replace that boundary energy by a *context* one, inspired by the error term of the cartoon functional. For that, given a curve  $\partial\tau$ , we define the regions  $\mathcal{R}_{in}^\varepsilon(\partial\tau)$  and  $\mathcal{R}_{out}^\varepsilon(\partial\tau)$  as the sets of points of maximal distance  $\varepsilon$  from  $\partial\tau$ , respectively inside and outside of the curve. Those two regions are illustrated by Figure 3 (b). The proposed external energy is:

$$E_{ext}(u, \partial\tau) = \frac{V(u, \mathcal{R}_{in}^\varepsilon(\partial\tau)) + V(u, \mathcal{R}_{out}^\varepsilon(\partial\tau))}{V(u, \mathcal{R}_{in}^\varepsilon(\partial\tau) \cup \mathcal{R}_{out}^\varepsilon(\partial\tau))}. \quad (1)$$



**Fig. 1.** (a): Energy estimator on a set of curves of an image, ordered by inclusion; snake energy is depicted in green, our proposed energy is in blue, and NFA is in orange. (b) and (c): the curves corresponding to significant energy minima.

This energy is low when the segmentation error is much lower for two classes than for one class, meaning that the curve is an object contour. Thanks to the numerator, it is also inversely proportional to the object contrast. Note that this contextual energy introduces some spatial information that is not naturally present in curve-based representation of an image. On a set of curves of an image, Figure 1 shows that replacing the gradient energy  $E_{ext}^{snk}$  by this new formulation dramatically improves the evidence of an object presence.

We also propose to modify the internal energy from the snakes approach: it is normalized to make it invariant to scale:

$$E_{int}(u, \partial\tau) = \sum_{e \in \partial\tau} |curv(u)(e)| / L(\partial\tau).$$

Now, all the curves  $\partial\tau$  are valued with the energy:

$$E = \alpha E_{int} + E_{ext} + \beta E_{con},$$

where the constraint energy aims at penalizing too small objects:

$$E_{con}(u, \partial\tau) = 1 / L(\partial\tau).$$

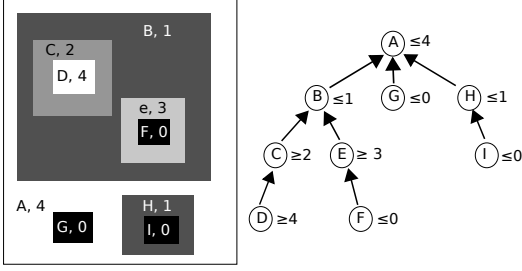
In our experiments we take  $\alpha = 60$ ,  $\beta = 2$ , and  $\varepsilon = 5$ .

## 3. APPLICATION ON THE TREE OF SHAPES

The method we propose is composed of three steps.

1. First the contents of the input image is described by a tree. Many tree-based representations of an image do exist (*e.g.* min/max tree, hierarchies, ...). In this paper, we use the tree of shapes [6], described in Section 3.1.
2. Second our context-based energy estimator is computed for each node of this tree.
3. Last we identify the significant local minima of the energy. Each node corresponding to such a minimum is a component whose contour is the one of an object.

We first describe the tree of shapes (Section 3.1), before giving some hints aiming at an efficient implementation (Section 3.2).



**Fig. 2.** An image (left) and its tree of shapes (right).

### 3.1. The Tree of Shapes

For any  $\lambda \in \mathbb{N}$ , the upper level sets  $\mathcal{X}_\lambda$  and lower level sets  $\mathcal{X}^\lambda$  of an image  $u : \mathbb{Z}^2 \rightarrow \mathbb{N}$  are defined by  $\mathcal{X}_\lambda(u) = \{p \in \mathbb{Z}^2 \mid u(p) \geq \lambda\}$  and  $\mathcal{X}^\lambda(u) = \{p \in \mathbb{Z}^2 \mid u(p) \leq \lambda\}$ . Both upper and lower level sets have a natural inclusion structure:  $\forall \lambda \leq \mu, \mathcal{X}_\lambda \supseteq \mathcal{X}_\mu$  and  $\mathcal{X}^\lambda \subseteq \mathcal{X}^\mu$ , which leads to two distinct and dual representations of an image, respectively the max-tree and the min-tree.

Another tree has been introduced in [6]. A *shape* is defined as a connected component of an upper or lower level set where its holes have been filled in. Thanks to the inclusion relationship of both kinds of level sets, the set of shapes gives a unique tree, called *tree of shapes*. This tree features an interesting property: it is invariant to contrast changes; put differently, it is a self-dual, non-redundant, and complete representation of an image. Furthermore, such a tree inherently embeds a morphological scale-space (the parent of a node/shape is a larger shape). An example on a simple image is depicted in Figure 2.

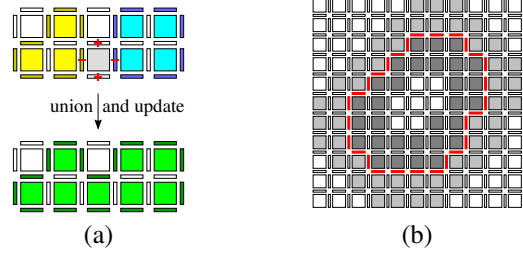
### 3.2. Making the Method Efficient

For all three steps of the method, we need to deal not only with image pixels but also with contours. We thus rely on a representation that handle both pixels *and* pixel edges. In other words, we materialize the elements what lie in-between pixels, as depicted in Figure 3.

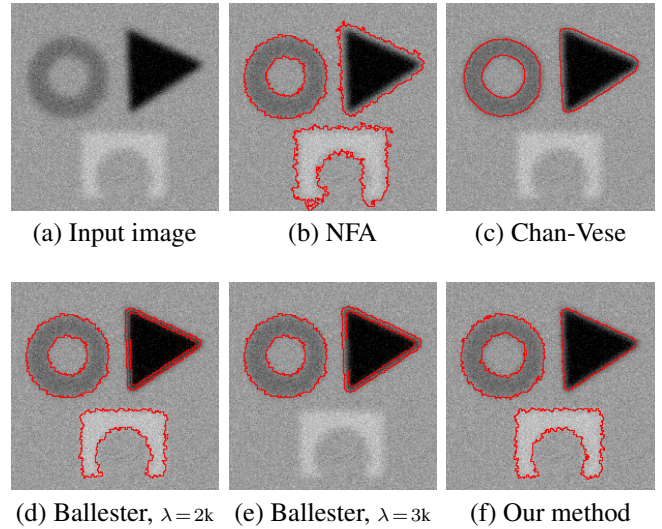
**Tree Computation.** To compute the tree of shapes, we use an algorithm (unpublished yet) very similar to the one described in [12]. A first pass sorts the pixels by a propagation process starting from the image boundary. Then a second pass, in reverse order, build the tree while performing the union-find algorithm. This algorithm has a quasi-linear time complexity when pixel values have a low quantization.

**Contours Information Computation.** During the union-find pass, contour information can be efficiently updated. As depicted in Figure 3 (a), when performing a union of two components (resp. yellow and blue) due to adding a pixel (gray), it is easy to know how to update the contour.

**Regions Information Computation.** In order to calculate the regions information efficiently, we approximate the inner



**Fig. 3.** (a): Updating contour information when a union between two components (yellow and blue) occurs thanks to a pixel (gray). (b): Regional context of a level line (red) is the inner (dark gray) and outer (light gray) regions.



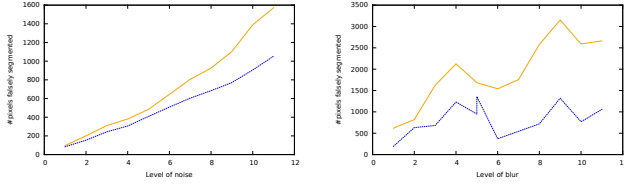
**Fig. 4.** Comparison with three other methods.

region and the outer region of each level line by only taking into account the pixels which are aligned perpendicularly to each edge of the level line. Note that some pixels may be counted several times. An example with  $\varepsilon = 2$  is given in Figure 3 (b).

**Energy Computation.** All energy terms can be computed incrementally during the union-find pass since they can be decomposed into linear (additive and subtractive) parts. The curvature, as a linear filter, is computed for every pixel edges in a preliminary step.

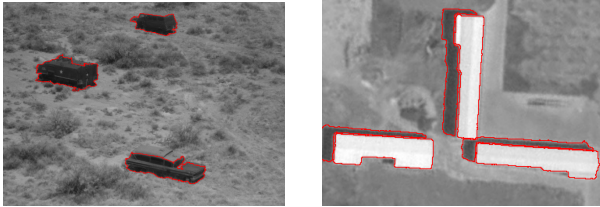
## 4. RESULTS AND COMPARISON

On Figure 1, one can see that the energy that we have defined (the blue curve) is a function that evolves rather smoothly upon the tree from a node to its parent (from left, a leaf, to right, the root node). Yet there exists many local minima and only a few of them effectively correspond to object contours. We thus filter the energy function in order to get rid of spurious minima [13]. The results obtained by our method on a synthetic image are shown in Figure 4(f) and they are com-



(a) Error due to increasing noise (b) and to increasing blur

**Fig. 5.** Evolution of number of falsely segmented pixels (ordinate) w.r.t. increasing noise and blur (abscissa) for the synthetic image in the Figure 4(a); NFA error evolution is depicted in orange and our method error in blue.



**Fig. 6.** Results on classical images.

pared to the ones of several related approaches.

The NFA approach [11] is sensitive to noise since the method relies on the minimum value of contrast along boundaries. This defect can be observed in Figure 1 where the NFA estimate (orange curve) does not evolve regularly enough. As a consequence, selecting relevant minima is hazardous and objects are not well located; see Figures 1 and 4(b). That confirms the importance of taking into account some regional information. A quantitative error comparison between NFA and our method is given in Figure 5 to depict the robustness to noise and blur of both methods.

Figure 4(c) is the result obtained by “active contour without edges” [3], with many circles as initial contours; the bottom object is missed since it is too similar to the background. That problem comes from the use of a global energy: such an object is not contrasted enough w.r.t. the rest of the image.

The approach of Ballester et al. [7] relying on Mumford-Shah functional gives pretty good results; see Figure 4(d). Unfortunately there is a false object, the triangle boundary, that does not disappear when increasing the regularization strength parameter  $\lambda$ ; see Figure 4(e). Instead of that, an actual object disappears.

Figure 6 presents the results of our method on two natural images. On the left image, the three vehicles are successfully identified but the roof of one car is missed since it is very similar to the background. On the satellite image, the buildings and their shadows are well segmented.

We have implemented the proposed method<sup>1</sup> using our C++ image processing library [14], available on the Internet as free software. Processing a  $512 \times 512$  pixels image takes less than 0.5 second on a regular PC station.

<sup>1</sup>Demo available on <http://olena.lrde.epita.fr/ICIP2012>

## 5. CONCLUSION

This paper presented a segmentation approach based on the set of level lines given by the tree of shapes. A major contribution of this paper is a new energy functional, well-suited to characterize object contours. We have shown that taking into account a regional context, so that the energy is neither too local nor global, improves both the relevance and robustness of object detection, as compared to three state-of-the-art approaches. Last, we end up with an efficient segmentation scheme with quasi-linear time complexity. A major perspective of our work is to demonstrate its usefulness with other tree-based representations, to extend it to color and, more generally, multi-valued images.

## 6. REFERENCES

- [1] D. Mumford and J. Shah, “Optimal approximations by piecewise smooth functions and associated variational problems,” *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
- [2] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes : Active contour models,” *IJCV*, vol. 1, pp. 321–331, 1987.
- [3] T.F. Chan and L.A. Vese, “Active contours without edges,” *IEEE Trans. on Image Proc.*, vol. 10, no. 2, pp. 266–277, 2001.
- [4] L. Najman, “On the equivalence between hierarchical segmentations and ultrametric watersheds,” *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 231–247, 2011.
- [5] V. Caselles, B. Coll, and J. Morel, “Topographic maps and local contrast changes in natural images,” *Intl. Journal of Computer Vision*, vol. 33, no. 1, pp. 5–27, 1999.
- [6] P. Monasse and F. Guichard, “Fast computation of a contrast-invariant image representation,” *IEEE Trans. on Image Processing*, vol. 9, no. 5, pp. 860–872, 2000.
- [7] C. Ballester, V. Caselles, L. Igual, and L. Garrido, “Level lines selection with variational models for segmentation and encoding,” *JMIV*, vol. 27, pp. 5–27, 2007.
- [8] J. Cardelino, G. Randall, M. Bertalmio, and V. Caselles, “Region based segmentation using the tree of shapes,” in *Proc. of IEEE ICIP*, 2006, pp. 2421–2424.
- [9] A. Pardo, “Semantic image segmentation using morphological tools,” in *Proc. of IEEE ICIP*, 2002, pp. 745–748.
- [10] A. Desolneux, L. Moisan, and J.M. Morel, “Edge detection by helmholtz principle,” *JMIV*, vol. 14, no. 3, pp. 271–284, 2001.
- [11] F. Cao, P. Musé, and F. Sur, “Extracting meaningful curves from images,” *JMIV*, vol. 22, pp. 159–181, 2005.
- [12] C. Berger, T. Géraud, R. Levillain, N. Widynski, A. Baillard, and E. Bertin, “Effective component tree computation with application to pattern recognition in astronomical imaging,” in *Proc. of IEEE ICIP*, 2007, vol. 4, pp. 41–44.
- [13] Y. Xu, T. Géraud, and L. Najman, “Morphological filtering in shape spaces: Applications using tree-based image representations,” 2012. <http://arxiv.org/abs/1204.4758>.
- [14] R. Levillain, T. Géraud, and L. Najman, “Why and how to design a generic and efficient image processing framework: The case of the Milena library,” in *Proc. of IEEE ICIP*, 2010, pp. 1941–1944. <http://olena.lrde.epita.fr>.