Impacts of contour saliency map transformations

Felipe de C. Belém*, Jean Cousty†, Zenilton K. G.do Patrocínio Jr.* and Silvio Jamil F. Guimarães*
*Audio-Visual Processing Laboratory, Department of Computer Science
Pontifical Catholic University of Minas Gerais
felipe.belem@sga.pucminas.br, {zenilton, sjamil}@pucminas.br
†Université Paris-Est, LIGM, ESIEE - UPEMLV - CNRS
jean.cousty@esiee.fr

***Abstract*—In recent years, many algorithms based on hierarchical segmentation were proposed. However, there has a need to represent their obtained results. The saliency map approach is an example of algorithm which allows the visualization of the hierarchy by defining a relevance to the contours found in the image. And, in order to visualize the segmentation result, a normalization is needed. Even though the classical approaches are satisfactory, there are two major problems: (i) the normalization does not prevent oversegmentation; (ii) it is not possible to alter the normalization to achieve better results. In this paper, we have studied a normalization by a sigmoid function, we analyzed its impacts as a filtering step, and as a final (visualization) one, and we also examined how the gradient impacts on the final result in accordance with our proposal. Experimental results show that, for both gradient functions, the pre-filtering step has better results than classical approaches.**

I. Introduction

Image segmentation is the process of grouping perceptually similar pixels into regions [1] and, one popular method is watershed algorithm. Beucher and Lantéjoul [2] presented the watershed transform in the context of image segmentations, and now it is used as fundamental step in several algorithms [3]. Consider the grayscale image as a topological surface in which the gray level of a pixel becomes the elevation of a point. The darker areas correspond to the valleys and basins and the lighter ones, to the peaks. The method consists on immersing the whole surface in a lake with holes in the minima. As the water fills, when lakes with different starting points would merge, a dam is built. The result is a surface partitioned by dams. However, this method has some drawbacks [4] including oversegmentation, which consists on the division of significant regions into small ones. This is caused, mainly, by the huge number of minima found in the image, leading to the increase number of dams built.

To cope with this problem, the authors in [5] studied the contour saliency map as a representation for better illustrate the "trully" contours in terms of a hierarchy of partitions. A hierarchy of segmentations could easily be infered by this hierarchy of partitions, and that hierarchy could be seen as a set of segmentations ordered by its level of details (from coarser to fine, or vice-versa), in which the coarser ones can be created by merging finer segmentations. So, the segmentations at finer levels are nested with respect to those at coarser levels [6].

A saliency map is a contour map, usually represented by a grayscale image, in which the gray level represents the strength of the contour (i.e., its level of disappearance in the hierarchy) [7]. The first notion of a saliency map was introduced in [5] for visualizing some hierarchies of watersheds. Then it was notably used in [8], [9] under the name of ultrametric contour maps [10]. However, to the best of our knowledge, there is no work that studied the impact of value transformations on saliency maps in order to improve the performance of these representations in terms of F-measure for Objects and Parts (F_{op}) [11]. Hierarchical segmentation has many applications such as occlusion boundary detection [12], image simplification [13]–[15], object detection [16], objects proposal [17] and visual saliency estimation [18].

Thus, the main goal of this work is to analyze the impact of different approaches on the saliency maps transformations by using different normalizations, such as linear, square root and sigmoid functions. Moreover, we have also studied the impact of filtering the gradient in order to reduce the number of minima, preventing the oversegmentation. Experimental results show that both normalization and gradient filtering have great impact on image segmentation metrics.

This paper is organized as follows. Section II presents more detailed notions of hierarchies and saliency maps, including some of the most recent works done in those areas. Using those definitions, we present the methodology used to obtain the segmentations, in Section III. Quantitative and qualitative results are presented in Section IV. Finally, in Section V, we draw some conclusions and point out possible research lines for future works.

II. Hierarchies and Saliency Maps

In this section we present and review the definitions of hierarchies, watershed hierarchies and saliency maps, that are used in this work.

A. Hierarchies

Let $\mathcal{G} = (V, E)$ be an edge-weighted graph, in which V is the vertex set, and E, the edge set. A *partition* (or a *segmentation*) $\mathbf{P}$ is a family of subsets of V such that the intersection of any two distinct elements of $\mathbf{P}$ is empty, and that the union of all elements of $\mathbf{P}$ is V. Each element of a partition is called a *region* and, given two partitions $\mathbf{P}_1$ and $\mathbf{P}_2$, we say that $\mathbf{P}_2$ is a *refinement* of $\mathbf{P}_1$, if every region of $\mathbf{P}_2$ is included in a region of $\mathbf{P}_1$.

A *hierarchy (of partitions)* $\mathcal{H} = (\mathbf{P}_0, ..., \mathbf{P}_n)$ is a sequence of partitions of V such that $\mathbf{P}_0 = \{V\}$, $\mathbf{P}_n = \{\{x\}|x \in V\}$, and P_i is a refinement of P_{i-1} $\forall i \in [1, n]$.

Given a hierarchy $\mathcal{H}$, the set of *regions of* $\mathcal{H}$, denoted by $\mathbf{R}_{\mathcal{H}}$ is the union of all partitions of $\mathcal{H}$. A partition $\mathbf{P}$ of V made of regions of $\mathcal{H}$ is called *cut of* $\mathcal{H}$, and it is horizontal if $\mathbf{P} = \mathbf{P}_i$ for some $i \in [1, n]$.

B. Watershed Hierarchies

The first proposal of watershed hierarchies can be found in [5], [19], [20], and it has been formalized in the context of minimum spanning forests (MSF) [7], [21]. Given an edge-weighted graph and subsets (markers) of the graph vertices indicating points of interest, a watershed can be obtained by performing a cut in the minimum spanning forest rooted in (exactly) one of the markers.

If the markers are ranked, it is possible to obtain a sequence of nested watershed, being the k^{th} MSF rooted in the k most important markers. Therefore, a hierarchy of watersheds is obtained. A usual choice to define a sequence of markers is to rank the minima according to extinction values [22]. The extinction value of a minima m for a given regional attribute is the smallest value λ_m such that m disappears when all components with attribute smaller than λ_m are removed. Dynamics, area and volume are common regional attributes [22], [23].

C. Saliency Maps

Let $\mathbf{P}$ be a partition of V. The *saliency map of* $\mathcal{H}$ is the map $\mathcal{S}_{\mathcal{H}}$ from E to n such that the weight of any edge u of $\mathcal{S}_{\mathcal{H}}$ is the maximum value λ for which u belongs to the cut of $\mathbf{P}_\lambda$ [10]. Informally, it consists on evaluating each point of the plane by the highest value h such that it appears in the boundaries of partition $\mathbf{P}_h$ [5].

Such partitions can be "stacked" to create a map that equivalently represents this hierarchy. Intuitively, such a map, called a (contour) saliency map, weights the cuts with their "level of dissapearance" in the hierarchy [7]. The low level (resp. upper level) of a hierarchy corresponds to weak (resp. strong) contours, so an over-segmentation (resp. under-segmentation) can be obtained by thresholding the saliency map with low (resp. high) value [24].

III. Saliency Maps Transformations

In this section, we present the steps to compute normalized saliency maps taking into account gradient filtering and normalization functions.

The first step consists in the extraction of the image gradient. The Euclidean Distance (denoted by EUC, from now on), is one of the most classical methods to calculate the image gradient, in the context of image analysis and segmentation. It consists in the calculation of the distance between pixels on a given color space (in this work, RGB space). New edge detection algorithms, like Structure Edges [25] (denoted by SE hereafter), were developed using supervised learning. Then, a hierarchy is built using the previously obtained gradient and considering area or volume as extinction values. Finally, a saliency map is computed taking into account the extinction values.

Since the human eye cannot discern a great range of color values, a normalization may help in order to facilitate the visualization. Thus, the saliency maps' normalization is the third step. Considering that the saliency map is represented by a grayscale image, the most common approaches for normalization are the linear and square root functions (denoted by LIN and SQRT, respectively). Their equations are defined as follows:

$$LIN(x) = \frac{x - min}{max - min} \tag{1}$$

$$SQRT(x) = \begin{cases} min, & \text{if } x \leq min \\ max, & \text{if } \sqrt{x} \geq max \\ \sqrt{x}, & \text{otherwise} \end{cases} \tag{2}$$

in which x is the original value, min and max are the minimum and maximum possible values, respectively.

In this work, we have studied the use of sigmoid function (denoted by SIGM, from now on) as a normalization step. It has a characteristic S-shaped curve dependent by two parameters: (i) middle point (x_0) and; (ii) steepness (k). If $x \in [0, 1]$, then the sigmoid equation can be defined as follows:

$$SIGM(x) = \frac{1}{1 + e^{-k(x - x_0)}} \tag{3}$$

IV. Experimental results

In this section, we present some quantitative and qualitative results of the proposed approach for computing normalized saliency maps for BSDS500 [9], [26] data set.

In this work, we have sistematically experimented: (i) the sigmoid, linear and square root functions; (ii) the impacts of gradient selection; (iii) the impacts of applying the sigmoid cost function in the gradient image. The idea here is to study the impact in the task of image segmentation taking into account hierarchies of watershed by using area and volume as extinction values.

In order to measure the segmentations obtained, we used the following metrics: (i) F-measure for Objects and Parts (F_{op}) [11] and; (ii) groundtruth covering. Being complementary, the usage of both permits a better analysis of the results obtained [11].

A. Experimental setup and limitations

We established that, for all tests made, whenever the sigmoid is used, the steepness and middle were empirically defined as 18 and 0.3. Also, for speeding purposes, we limited the number of regions in any partition to be, at most, 4000 regions.

B. Gradient function analysis

The difference between EUC and SE [25] gradient functions can be seen in Figure 1, where the images in EUC column presents lots of noisy edges (also sharpened ones), whereas

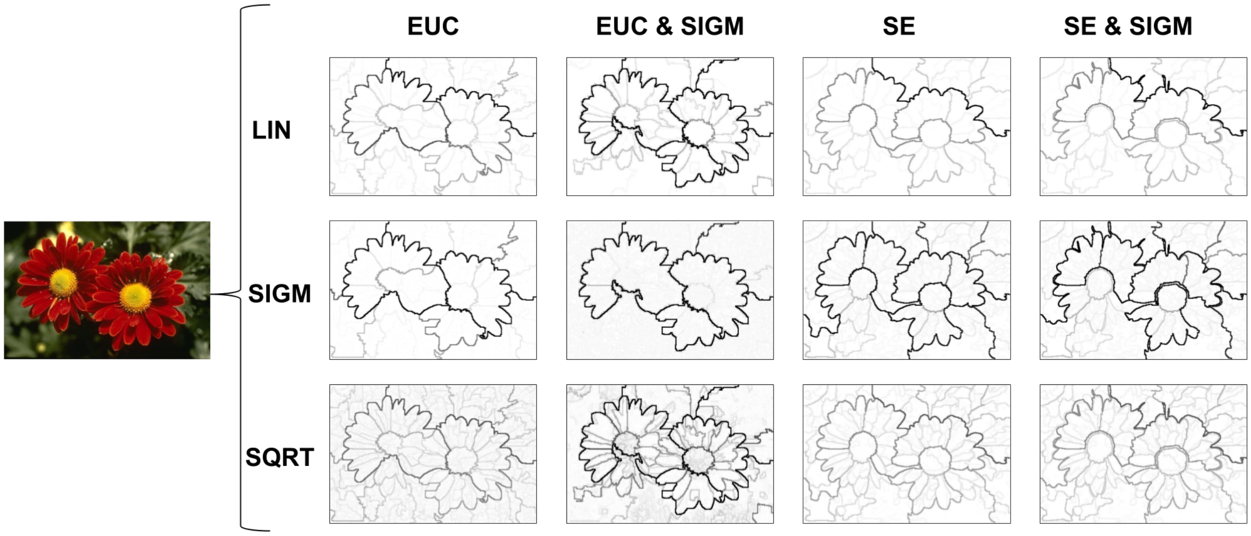

Fig. 1. Results obtained from the original (colored) image. Each column corresponds to a different gradient function, and each row corresponds to a different normalization value. All segmentations were obtained using area as extinction value.

the SE [25] results has more smooth edges and a significant non-relevant edge suppression.

It is possible to see that, in Figure 2, using a more efficient gradient function, the F_{op} [11] improved significantly, for every normalization. Since the EUC only calculates the distance between pixels, it does not verify the relevance of the contour. However, the SE [25] algorithm outlies and smoothes the borders by its relevance in the image, thus reducing the noise in the gradient and highliting the contours which are more likely to be relevant. This prevents the over and undersegmentation, respectively.

C. Normalization analysis

The linear scale is the normalization that, visually, best represents the structure of the hierarchy obtained, due to the contours contrast maintains linear in respect of the (hierarchy) height. However, there are some which are visually relevant that, during the normalization process, the resulting value is the same as irrelevant ones. In opposition to the previous one, the square root enhances mid-level values, resulting in a better perception of the nested components in the hierarchy. In the other hand, this process affects, also, non-relevant visual contours, leading to over segmentation in most cases.

The sigmoid function unifies the best of both by flattening the highest (resp. lowest) values, which does not impact visually, and by stretching specific mid-level ones. As the hierarchies differ from one to another, so does the parameters of the function. The bad selection of these can cause over or under segmentation in the final result.

One example of bad selection of the parameters is shown in Figure 1 in the column EUC & SIGM, and row SIGM. In comparison with the respective non-filtered result (EUC), it is possible to see that many relevant contours were suppressed in exchange for supressing the insignificant ones. However, if the gradient function is efficient, there are few noisy contours. Thus, the sigmoid function will filter a low quantity of noisy contours. This behaviour is illustrated in the figure's two leftmost columns, which has little difference between normalizations.

Analyzing Figure 3, it is possible to notice that the sigmoid, as normalization function, has achieved a better groundtruth covering performance and, being a function based on parameters, it is plausible to assume that an optimal configuration for each input may lead to even better results. It is interesting to notice that, although differ in their proposals, both LIN and SQRT obtained extremely close results in all the measures, which may suggest that they does not penalize over segmentation as well as it should.

With the presence of the pre-filtering step, most of the configurations were significantly better. That may be caused by the removal, or junction (if their difference is sufficiently low for suppression), of several basins. This consequence permits that all minima are equally deep (resp. shallow), creating a tendency of choice when using extinction values such as area and volume.

V. Conclusion and Further Works

In this paper, we evaluated the impacts of the contour saliency map transformations, more specifically the gradient and normalization functions. Also, we compared the usage of the sigmoid function as a normalization method, with two classical approaches.

Experimental results demonstrated that the sigmoid normalization is on pair with the classical approaches, and that the presence of a pre-filtering step has increased the quality of the segmentations. The combination of approaches, show best overall results in all measures, being on pair with some of state-of-the-art methods. Since the sigmoid functions is

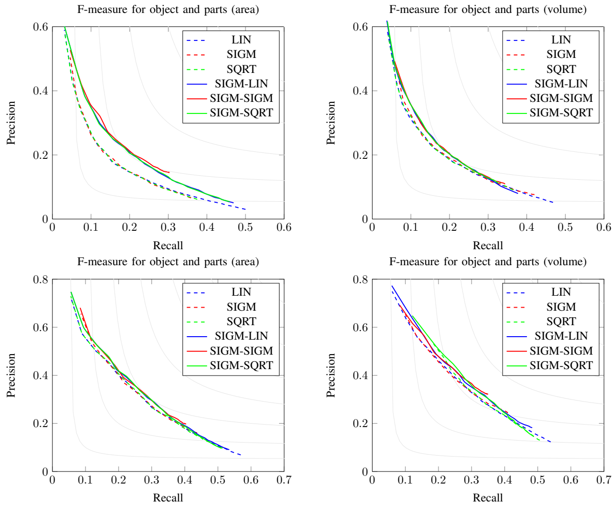

Fig. 2. Evaluation of Watershed by area and volume taking into account filtering of the Euclidean (first row) and Structured Edge (second row) gradients. Dashed lines represents results without filtering. The comparison is made by using F_{op}.

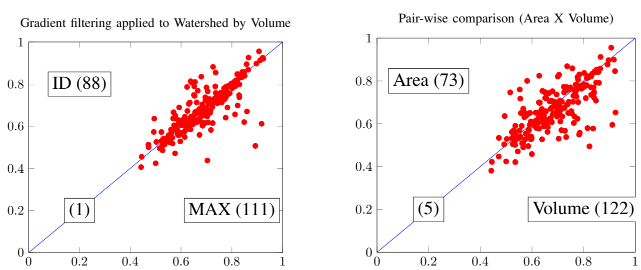

Fig. 3. Pair-wise comparison taking into account the best image GT-Cover. In the left, a comparison involving Watershed by Volume considering gradient filtering of the structure edge in which ID denotes a non-filtered gradient. In the right, a comparison for Watershed by Area and Volume. The number of images which have better results when compared to the other is illustrated in the figure.

dependant of a parameter tuning, it is possible to infer that, if a optimum configuration is known for any image, it may present even better results.

For future endeavours, we will study different state-of-the-art gradient functions, and different and new approaches to extend them to be used for any input graph.

Acknowledgment

The authors are grateful to FAPEMIG, CNPq, PUC Minas and CAPES for the financial support to this work.

References

[1] S. Guimarães, J. Cousty, Y. Kenmochi, and L. Najman, "A hierarchical image segmentation algorithm based on an observation scale," *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 116–125, 2012.

[2] S. Beucher and C. Lantuéjoul, "Use of watersheds in contour detection," 1979.

[3] L. Najman and M. Couprie, "Watershed algorithms and contrast preservation," in *International Conference on Discrete Geometry for Computer Imagery*. Springer, 2003, pp. 62–71.

[4] G. Bertrand, "On topological watersheds," *Journal of Mathematical Imaging and Vision*, vol. 22, no. 2-3, pp. 217–230, 2005.

[5] 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.

[6] 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," Tech. Rep., 2016.

[7] 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.

[8] 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.

[9] P. A. Arbeláez and L. D. Cohen, "A metric approach to vector-valued image segmentation," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 119–126, 2006.

[10] J. Cousty, L. Najman, Y. Kenmochi, and S. Guimarães, "Hierarchical segmentations with graphs: quasi-at zones, minimum spanning trees, and saliency maps," Tech. Rep., 2016.

[11] J. Pont-Tuset and F. Marques, "Measures and meta-measures for the supervised evaluation of image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2131–2138.

[12] 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.

[13] L. Guigues and H. Le Men, "Scale-sets image analysis," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2. IEEE, 2003, pp. II–45.

[14] 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.

[15] 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.

[16] 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.

[17] J. Pont-Tuset, P. Arbelaez, J. T. 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, no. 1, pp. 128–140, 2017.

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

[19] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Mathematical morphology and its applications to image processing*. Springer, 1994, pp. 69–76.

[20] F. Meyer, "The dynamics of minima and contours," in *Mathematical Morphology and its Applications to Image and Signal Processing*. Springer, 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] 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.

[23] M. Grimaud, "New measure of contrast: the dynamics," in *San Diego'92*. International Society for Optics and Photonics, 1992, pp. 292–305.

[24] Y. Xu, T. Géraud, and L. Najman, "Hierarchical image simplification and segmentation based on mumford–shah-salient level line selection," *Pattern Recognition Letters*, vol. 83, pp. 278–286, 2016.

[25] 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.

[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 416–423.