Superpixel Segmentation by Object-Based Iterative Spanning Forest

Felipe Belém[1(✉)], Silvio Jamil F. Guimarães[2], and Alexandre Xavier Falcão[1]

[1] University of Campinas, Campinas, SP 13083-852, Brazil
felipedkstro@hotmail.com, afalcao@ic.unicamp.br
[2] Computer Science Department, Pontifical Catholic University of Minas Gerais,
Belo Horizonte, MG 31980-110, Brazil
sjamil@pucminas.br

Abstract. Superpixel segmentation methods aim at representing image objects by the union of connected regions (superpixels). Such aim can be better approximated with a higher number of superpixels per object, which often leads to an unnecessary over-segmentation due to the absence of prior object information. In this work, we extend the *Iterative Spanning Forest* (ISF) framework to include object information and present a superpixel segmentation method based on object saliency detection. As ISF, the new framework, named *Object-based ISF* (OISF), relies on multiple executions of the *Image Foresting Transform* (IFT) algorithm for improved seed sets, such that each seed defines one connected superpixel as a spanning tree rooted at that seed. We describe an IFT-based method for object saliency detection and show that the corresponding saliency maps can improve seed estimation and connectivity function, increasing the superpixel resolution inside a given object. Experimental results on two medical image datasets demonstrate that the proposed OISF-based method outperforms the state-of-the-art in boundary adherence with higher number of superpixels inside the object.

Keywords: Superpixels · Object saliency map ·
Image Foresting Transform

1 Introduction

Image segmentation into connected regions (superpixels) has been actively investigated in order to represent image objects by the union of their superpixels [1,6,9,11,12]—a criterion that often leads to unnecessary over-segmentation of the image. For instance, content/structure-sensitive approaches may reduce the superpixel size (increase over-segmentation) in heterogeneous regions of the image, but the absence of object information makes them sensitive to the heterogeneity of the background [6,12]. Moreover, these methods cannot usually

guarantee a desired number of superpixels. In many applications, however, there is an object of interest and, for a fixed number of superpixels, it should be expected higher superpixel resolution inside that object than elsewhere, except for possible parts of the background with similar image properties. At the same time, for a reduced number of superpixels, the boundaries of the object should be preserved as much as possible (Fig. 1).

In this work, we extend a superpixel segmentation framework, named *Iterative Spanning Forest* (ISF) [11], to incorporate object information from an object saliency map. ISF-based methods use multiple executions of the *Image Foresting Transform* (IFT) algorithm [4] for improved seed sets, such that each seed defines one spanning tree as a connected superpixel. An ISF-based method involves the choice of four components: (i) a seed sampling strategy to obtain the first segmentation; (ii) an adjacency relation that defines the image graph in 2D or 3D (for superpixel- or supervoxel-based representation); (iii) a connectivity function that estimates how strongly connected are the pixels to the seed set; and (iv) a seed recomputation procedure for the subsequent execution of the IFT algorithm.

We first use the IFT framework to design a method for object saliency detection. For a given image and a set of training pixels (interior and exterior scribbles) on a given object, we train a pixel classifier to estimate an *object saliency map* from any new image containing that object. We then propose a method that exploits the saliency map to make seed sampling and connectivity function more specific for that object. The new framework is termed *Object-based* ISF (OISF) and the proposed OISF-based method is shown to increase boundary adherence with more superpixels inside the object than their ISF-based counterparts and state-of-the-art methods.

The next sections present the IFT framework and related definitions (Sect. 2), its applications to object saliency detection and superpixel segmentation (Sects. 3 and 4), the proposed OISF framework and its evaluation (Sects. 5 and 6), conclusion and future work (Sect. 7).

2   Image Foresting Transform

An image is a pair $(\mathcal{I}, I)$ such that $I(t)$ assigns a set of local image features (e.g., color) to every element $t \in \mathcal{I}$. We will address only 2D images, then those elements are pixels. For a given *adjacency relation* $\mathcal{A} \subset \mathcal{I} \times \mathcal{I}$ and set $\mathcal{N} \subseteq \mathcal{I}$, one can interpret $(\mathcal{N}, \mathcal{A}, I)$ as an image graph G weighted on the nodes. Let Π_G be the set of paths in the graph, a path $\pi_t \in \Pi_G$ be a sequence $\langle t_1, t_2, \ldots, t_n = t \rangle$ of nodes with terminus t, such that $(t_i, t_{i+1}) \in \mathcal{A}$, $i = 1, 2, \ldots, n - 1$ (being trivial when $\pi_t = \langle t \rangle$), and f be a *connectivity function* that assigns a value (e.g., a cost) to any path in Π_G. A path π_t is optimum when $f(\pi_t) \leq f(\tau_t)$ for any other path $\tau_t \in \Pi_G$ irrespective to its starting node. For the sufficient conditions in [2], Dijkstra's algorithm can solve the minimization problem $C(t) = \min_{\forall \pi_t \in \Pi_G} \{ f(\pi_t) \}$ by computing an *optimum-path forest* in the graph—i.e., a predecessor map P that assigns to every node $t \in \mathcal{N}$ its predecessor $P(t) \in \mathcal{N}$

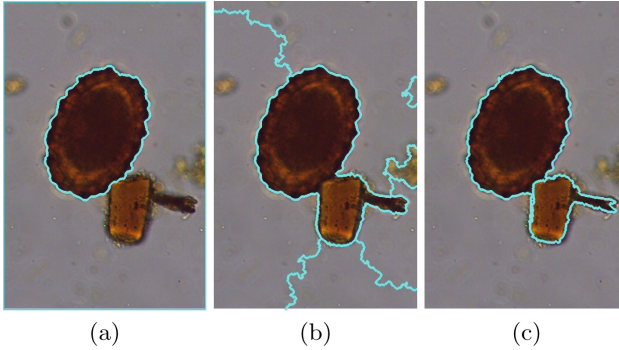(a)   (b)   (c)

Fig. 1. (a) Original image in which the contour indicates an object of interest. For only three superpixels, (b) the result of a content-sensitive approach based on entropy [11] and (c) the result of the proposed method based on object information.

in the optimum path π_t^* or a marker $P(t) = nil \notin \mathcal{N}$ when t is a *root* of the map. Even when those conditions are not satisfied, the algorithm can output a spanning forest with properties that are useful for several applications. This framework to the design of image operators based on optimum-path forest is called *Image Foresting Transform* (IFT) [4].

In this work we are interested in two of its applications: object saliency detection based on pixel classification [8]; and superpixel segmentation [11]. The next sections illustrate IFT-based image operators with examples of adjacency relation and connectivity function for those applications.

3   IFT-based Object Saliency Detection

A map O that assigns values $O(t)$, $t \in \mathcal{I}$, proportional to the similarity between t and a given object is said *object saliency map*. We create object saliency maps by training a pixel classifier [8] from user-drawn scribbles inside and outside a given object in one training image. Of course, one can build a pixel training set from scribbles drawn on several training images as well, whenever this is required by the application. The scribbles represent a set of training pixels whose color/texture properties may be mapped onto overlapping regions in the corresponding feature space. By clustering, we first select a small set (e.g., 500 pixels) of the most representative object and background pixels to train the classifier with minimum overlapping between regions of distinct classes in the feature space. Therefore, let $\mathcal{N}$ be such selected set of training pixels and $\mathcal{A}$ be a complete adjacency relation that connects any pair of pixels $(s, t) \in \mathcal{N} \times \mathcal{N}$. A seed set $\mathcal{S} \subset \mathcal{N}$ is defined with the closest pixels from distinct classes (object or background) in G according to the Euclidean norm $\|I(t), I(s)\|$ between their colors in the CIELab color space. The set $\mathcal{S}$ is usually obtained by computing a Minimum Spanning Tree (MST) in G and selecting nodes from distinct classes that share an arc in the MST [8]. Let f_o and f_b be path-cost functions such as

$$f_x(\langle t \rangle) = \begin{cases} 0 & \text{if } t \in \mathcal{S}_x \subset \mathcal{S}, \\ +\infty & \text{otherwise}, \end{cases} \qquad (1)$$

$$f_x(\pi_s \cdot \langle s, t \rangle) = \max\{f_x(\pi_s), \|I(t), I(s)\|\},$$

where $\mathcal{S}_x$ contains either object $(x = o)$ or background $(x = b)$ seeds, and $\pi_s \cdot \langle s, t \rangle$ indicates the extension of π_s by an arc $\langle s, t \rangle$ with the two joining instances of s merged into one. The IFT algorithm is executed for each path-cost function in order to obtain two minimum path-cost maps, which are combined into the final object saliency map O, such that $O(t) = \frac{C_b(t)}{C_o(t)+C_b(t)}$, where $C_x(t) = \min_{\forall \pi_t \in \Pi_G}\{f_x(\pi_t)\}$. For each node in $t \in \mathcal{N}$, $C_o(t)$ and $C_b(t)$ store the costs of the paths rooted at the most closely connected seeds in $\mathcal{S}_o$ and $\mathcal{S}_b$. Those seeds offer to t paths whose maximum arc weight $\|I(t), I(s)\|$ is minimum. For pixels t very similar to the object, it is expected that $C_b(t) \gg C_o(t) \implies O(t) \approx 1$.

4   Superpixel Segmentation by Iterative Spanning Forest

The *Iterative Spanning Forest* (ISF) framework consists of four components: (i) a seed sampling strategy; (ii) an adjacency relation; (iii) a connectivity function; and (iv) a seed recomputation procedure [11]. For a given choice of these components, one can design distinct superpixel segmentation methods. ISF executes the IFT algorithm multiple times for improved seed sets in order to obtain the final superpixel segmentation.

In 2D, the adjacency relation $\mathcal{A} \subset \mathcal{I} \times \mathcal{I}$ connects pairs of 4-neighboring pixels. The graph is defined as $G = (\mathcal{I}, \mathcal{A}, I)$. The connectivity function may be

$$f_1(\langle t \rangle) = \begin{cases} 0 & \text{if } t \in \mathcal{S}, \\ +\infty & \text{otherwise}, \end{cases} \qquad (2)$$

$$f_1(\pi_s \cdot \langle s, t \rangle) = f_1(\pi_s) + [\alpha \|I(r_s), I(t)\|]^\beta + \|t, s\|,$$

where $\mathcal{S}$ is a set of seed pixels, r_s is the starting pixel (root) of π_s, $\alpha \geq 0$, $\beta > 1$, and $\|t, s\| = 1$ since it represents the Euclidean norm between 4-neighboring pixels. The role of α is to provide user control over the superpixel compactness and regularity—lower is α, more compact and regular they are. The β parameter controls the boundary adherence—higher is β, higher is the adherence of superpixels to the boundaries of the objects, but this reduces their shape regularity and compactness. For an initial set $\mathcal{S} \subset \mathcal{I}$, the IFT algorithm aims at finding minimum-cost paths from $\mathcal{S}$ to the remaining pixels in $\mathcal{I} \backslash \mathcal{S}$. The connectivity function may not satisfy the conditions in [2], but each seed in $\mathcal{S}$ defines one spanning tree (connected superpixel) suitable for image representation. The seed recomputation procedure aims at improving the seed set $\mathcal{S}$ for the subsequent execution of the IFT algorithm using the same connectivity function. Among the components presented in [11], the authors concluded that the ones that use f_1, as defined in Eq. 2, and recomputes one seed inside each superpixel as the

closest pixel to its geometric center, were the most competitive. ISF uses a convergence criterion to select new seeds and so the spanning forest can efficiently be updated in a differential way [3].

Taking into account the seed sampling strategies in [11], *GRID* and *MIX* are the most competitive to estimate the initial set $\mathcal{S}$. GRID selects a given number of equally spaced pixels from $\mathcal{I}$ and then approximate them to the closest minimum in a gradient image. MIX seed sampling creates a two-level quad-tree, using the normalized Shannon entropy, as predicate, and performs GRID on the leaves of the tree. While GRID prioritizes a regular sampling over the image domain, MIX aims at increasing the number of seeds in heterogeneous regions, such as a content-sensitive approach, and at the same time preserving the regularity of the grid sampling.

5   Object-Based ISF for Superpixel Segmentation

In applications with a given object of interest (e.g., an organ in medical images), one can train a pixel classifier (e.g., the approach described in Sect. 3) to estimate the object saliency map O from any given image. We then propose the use of that map in ISF to increase the number of initial seeds in the image regions most similar to the object (brighter regions in the map). For a fixed number of superpixels, this should lead to higher superpixel resolution inside the object than elsewhere in comparison with other ISF-based methods. We call this approach *object-based seed sampling*. We also propose the use of an *object-based connectivity function* similar to the one proposed in [10] in order to increase the boundary adherence of the superpixels to the high-contrast regions of the saliency map. The new framework is then named *Object-based* ISF (OISF).

5.1   Object-Based Seed Sampling Strategy

A binary mask M with most object pixels is defined as $M(t) = 1$, if $O(t) \geq T$ (e.g., $T = 0.5$), or $M(t) = 0$ otherwise. The binary mask may consist of multiple components and the number of seeds in each component is proportional to its area. Our approach selects a percentage of seeds within those components and the remaining seeds in regions where $M(t) = 0$ to compose the initial set $\mathcal{S}$. This process uses *geodesic grid sampling—i.e., equally spaced seeds inside each component.*

5.2   Object-Based Connectivity Function

The authors in [10] proposed a new function f_2, derived from f_1, which takes into account the relevance of a presegmentation map (for segmentation resuming). Thus, for our proposal, f_2 can be rewritten as $f_2(\langle t \rangle) = f_1(\langle t \rangle)$ and

$$f_2(\pi_s \cdot \langle s, t \rangle) = f_2(\pi_s) + \|t, s\| + \tag{3}$$
$$\left[\alpha \|I(r_s), I(t)\| \gamma^{|O(r_s) - O(t)|} + \gamma |O(r_s) - O(t)| \right]^{\beta},$$

where $\gamma > 0$ controls the balance between boundary adherence to high-contrast regions of the image and saliency map. Figure 2 illustrates the impact of γ in the proposed OISF-based method, named *OISF-GRID* due to the geodesic grid sampling—i.e., higher is γ higher is the adherence to the object boundaries in the saliency map.

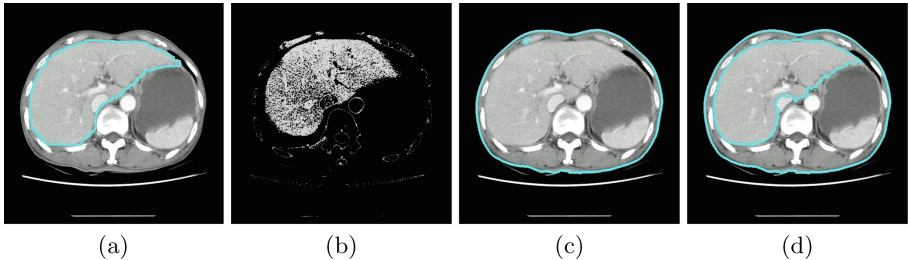|			
(a)	(b)	(c)	(d)

Fig. 2. (a) Original image with the contour indicating the object of interest. (b) The object saliency map using the classifier pre-trained on another image. Result for three superpixels only using OISF-GRID with (c) $\gamma = 1$ and (d) $\gamma = 10$.

6   Experimental Results

The experiments used two datasets: *Parasites*, with 77 images of *Schistosoma Mansoni* eggs, and *Liver*, with 40 CT-image slices of the liver, being the eggs and the liver their respective objects of interest. We fixed $\alpha = 0.5$ and $\beta = 12$, as suggested in [11], to prioritize boundary adherence over compactness. For γ, the best values for Liver and Parasites were $\gamma = 1.75$ and $\gamma = 1.5$, respectively, as obtained by grid search on $\approx 30\%$ of the images. The classifier used to create object saliency maps was trained from 500 pixels of a single image (Sect. 3).

Methods for superpixel segmentation are usually assessed by two boundary adherence measures: (i) boundary recall (BR) [1] (higher is better); and (ii) under-segmentation error (UE) [7] (lower is better). Since the size of the object's boundary is usually very small as compared to its size, these measures cannot capture the ability of a method to retain more superpixels inside the object than elsewhere. Except for a low number of superpixels, they can show when a method best preserves the object's boundary due to that property. Therefore, boundary adherence with higher superpixel resolution in a given object than elsewhere is measured by $wBR = BR \cdot P$ and $wUE = \frac{UE}{P}$, where P is the percentage of superpixels inside that object. We compare OISF-GRID with four ISF-based methods [11] (ISF-GRID-MEAN, ISF-GRID-ROOT, ISF-MIX-MEAN, ISF-MIX-ROOT) and two state-of-the-art approaches, the popular SLIC [1] and a more recent one, LSC [5], according to those weighted boundary adherence measures (see Fig. 3). The performance of OISF-GRID is by far the best, mainly because the penalization for irrelevant background superpixels.
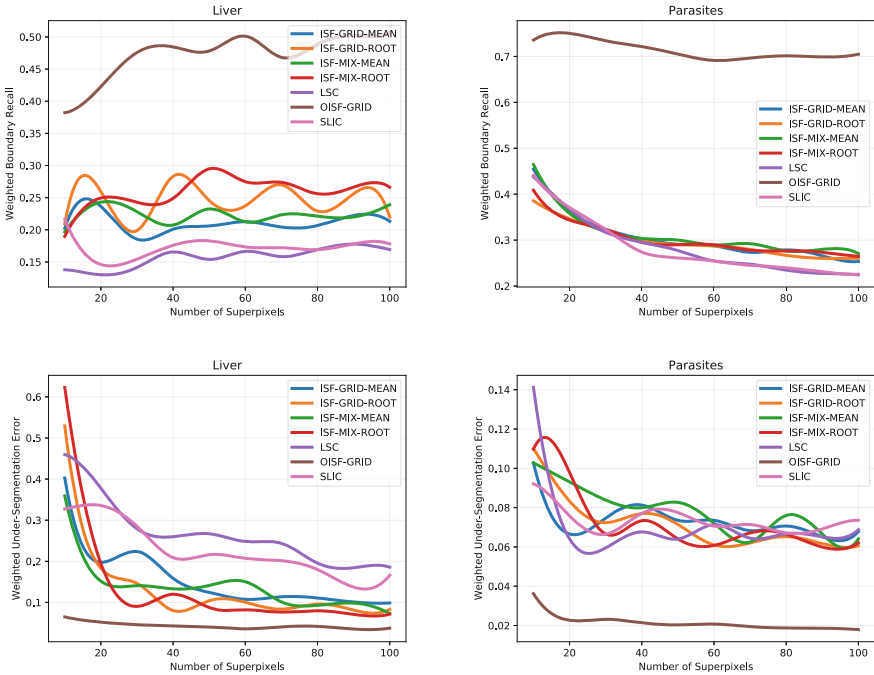
Fig. 3. OISF-GRID *versus* different superpixel-generation methods for varying number of superpixels.

Although the computation of the object saliency map is detached from the OISF-GRID algorithm, our proposal requires slightly higher processing time than ISF due to the geodesic grid sampling on each component of the map. However, the processing time of OISF is equivalent to the one of ISF in the remaining steps.

7   Conclusion

We presented the Object-based Iterative Spanning Forest framework (OISF) and an OISF-based method that considerably improves boundary adherence with higher number of superpixels inside a given object than elsewhere (thus, it reduces the quantity of irrelevant superpixels in the background). OISF incorporates object information from an object saliency map. We have shown an effective solution for saliency detection, but OISF can be used with other saliency detection methods. We intend now to investigate new OISF-based methods, evaluate them on 3D medical image datasets, and explore OISF in applications that require object delineation (i.e., semantic image segmentation).

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC super-pixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2274–2282 (2012). https://doi.org/10.1109/TPAMI.2012.120
2. Ciesielski, K.C., Falcão, A.X., Miranda, P.A.V.: Path-value functions for which Dijkstra's algorithm returns optimal mapping. J. Math. Imaging Vis. **60**, 1025–1036 (2018). https://doi.org/10.1007/s10851-018-0793-1
3. Condori, M.A.T., Cappabianco, F.A.M., Falcão, A.X., De Miranda, P.A.V.: Extending the differential image foresting transform to root-based path-cost functions with application to superpixel segmentation. In: Proceedings of 30th Conference on Graphics Pattern Images (SIBGRAPI), pp. 7–14, October 2017. https://doi.org/10.1109/SIBGRAPI.2017.8
4. Falcão, A.X., Stolfi, J., Lotufo, R.A.: The image foresting transform: theory, algorithms, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 19–29 (2004). https://doi.org/10.1109/TPAMI.2004.1261076
5. Li, Z., Chen, J.: Superpixel segmentation using linear spectral clustering. In: Proceedings of 28th Conference on Computer Vision and Pattern Recognition, pp. 1356–1363, June 2015. https://doi.org/10.1109/CVPR.2015.7298741
6. Liu, Y.J., Yu, M., Li, B.J., He, Y.: Intrinsic manifold SLIC: a simple and efficient method for computing content-sensitive superpixels. IEEE Trans. Pattern Anal. Mach. Intell. **40**(3), 653–666 (2018). https://doi.org/10.1109/TPAMI.2017.2686857
7. Neubert, P., Protzel, P.: Superpixel benchmark and comparison. In: Proceedings of Forum Bildverarbeitung, vol. 6, pp. 1–12 (2012)
8. Papa, J.P., Falcão, A.X., Suzuki, C.T.: Supervised pattern classification based on optimum-path forest. Int. J. Imaging Syst. Technol. **19**(2), 120–131 (2009). https://doi.org/10.1002/ima.20188
9. Stutz, D., Hermans, A., Leibe, B.: Superpixels: an evaluation of the state-of-the-art. Computer Vis. Image Underst. **166**, 1–27 (2018). https://doi.org/10.1016/j.cviu.2017.03.007
10. Tavares, A.C.M., Miranda, P.A.V., Spina, T.V., Falcão, A.X.: A supervoxel-based solution to resume segmentation for interactive correction by differential image-foresting transforms. In: Angulo, J., Velasco-Forero, S., Meyer, F. (eds.) ISMM 2017. LNCS, vol. 10225, pp. 107–118. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57240-6_9
11. Vargas-Muñoz, J.E., Chowdhury, A.S., Alexandre, E.B., Galvão, F.L., Miranda, P.A.V., Falcão, A.X.: An iterative spanning forest framework for superpixel segmentation. IEEE Trans. Image Process (to appear 2019)
12. Wang, P., Zeng, G., Gan, R., Wang, J., Zha, H.: Structure-sensitive superpixels via geodesic distance. Int. J. Comput. Vis. **103**, 1–21 (2013). https://doi.org/10.1007/s11263-012-0588-6