HAL
open science

Towards a Simple and Efficient Object-based Superpixel Delineation Framework

Felipe C Belém, Benjamin Perret, Jean Cousty, Silvio J F Guimarães, Alexandre X Falcão

To cite this version:

Felipe C Belém, Benjamin Perret, Jean Cousty, Silvio J F Guimarães, Alexandre X Falcão. Towards a Simple and Efficient Object-based Superpixel Delineation Framework. 2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Oct 2021, Gramado, Brazil. pp.346-353, 10.1109/SIBGRAPI54419.2021.00054 . hal-04341605

HAL Id: hal-04341605

https://hal.science/hal-04341605v1

Submitted on 13 Dec 2023

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.

Towards a Simple and Efficient Object-based Superpixel Delineation Framework

Felipe C. Belém[1], Benjamin Perret[2], Jean Cousty[2], Silvio J. F. Guimarães[3] and Alexandre X. Falcão[1]

[1] LIDS, Institute of Computing, University of Campinas, São Paulo, Brazil
[2] LIGM, Université Gustave Eiffel, CNRS, ESIEE Paris, Marne-la-Vallé, France
[3] IMScience, Pontifical Catholic University of Minas Gerais, Minas Gerais, Brazil
Emails: {felipe.belem,afalcao}@ic.unicamp.br, {benjamin.perret,jean.cousty}@esiee.fr, sjamil@pucminas.br

***Abstract*—Superpixel segmentation methods are widely used in computer vision applications due to their properties in border delineation. These methods do not usually take into account any prior object information. Although there are a few exceptions, such methods significantly rely on the quality of the object information provided and present high computational cost in most practical cases. Inspired by such approaches, we propose *Object-based Dynamic and Iterative Spanning Forest* (ODISF), a novel object-based superpixel segmentation framework to effectively exploit prior object information while being robust to the quality of that information. ODISF consists of three independent steps: (i) seed oversampling; (ii) dynamic path-based superpixel generation; and (iii) object-based seed removal. After (i), steps (ii) and (iii) are repeated until the desired number of superpixels is finally reached. Experimental results show that ODISF can surpass state-of-the-art methods according to several metrics, while being significantly faster than its object-based counterparts.**

I. Introduction

Superpixels are commonly defined as groups of connected pixels which share a common property (*e.g*, color, texture). Superpixel segmentation methods create an image partition by delineating relevant borders. Ideally, they should preserve the borders of an object of interest, such that the object can be represented by the union of its superpixels. These methods are often used as an intermediary step in several applications: (i) medical image analysis [1]–[3]; (ii) pedestrian segmentation [4]; and (iii) plant detection [5].

Several works [6], [7] elect a number of desirable properties for superpixel methods.

- Every pixel must be uniquely assigned to a single superpixel;
- Superpixels must be a connected set of pixels;
- The object boundaries should be overlapped by a superpixel border;
- Superpixels should be generated efficiently; and
- The user should be able to control the number of superpixels.

Moreover, these properties should be satisfied with as few superpixels as possible [8]. Although some authors list other properties, such as compacity, one may note that the aforementioned ones are a consensus amongst all methods.

A three-stage pipeline is an approach commonly seen in many superpixel segmentation methods [6], [11]–[13]: (a) initial seed sampling; (b) superpixel segmentation; and (c) seed recomputation, repeating steps (b) and (c) for a few iterations. As an example, *Iterative Spanning Forest* (ISF) [12] is a superpixel segmentation framework in which, benefiting from the properties of the *Image Foresting Transform* (IFT) [14] algorithm, all of its steps are independently defined. That is, modifications in one step do not require adjustments in the others. *Dynamic ISF* [13] is another example that follows the same pipeline. In DISF, the seeds are selected through oversampling and, through a few iterations, superpixels are delineated by the IFT algorithm as the seed set is reduced in step (c) based on superpixel properties. Although state-of-the-art methods present accurate object delineation, they do not usually consider prior object information. For instance, it is not possible to control the superpixel displacement in order to

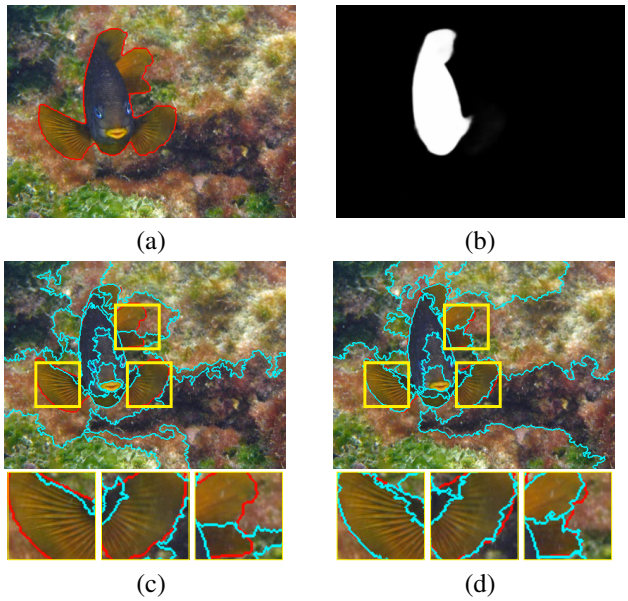


Fig. 1: Results of object-based superpixel segmentation methods considering 25 superpixels. (a) Original image with ground-truth borders overlayed in red; (b) Object saliency map [9]; Segmentation results (in cyan), drawn over (a), obtained by (c) OISF [10] and (d) our proposal.

improve object delineation in specific regions, often within or nearby the boundary of the objects of interest.

A recent category of superpixel methods permits the inclusion of object information during execution. Given such information, these algorithms may improve the delineation performance in relevant regions. The *Object-based ISF* [10], [15] method is a three-stage IFT-based superpixel approach that falls into this category. In OISF, object information is represented by an object saliency map. OISF variants consider the map's intensities as indication of probable object location and extension. As exemplified in Figure 1, such maps often present good estimation of the object location, but performs poorly in terms of object delineation (even when the saliency map is created by a deep neural network [9]). Given that, OISF methods are highly dependable on the object information quality, such that errors in the saliency map may critically affect its delineation performance. Furthermore, OISF variants often present high computational cost.

In this paper, we propose *Object-based DISF*, a novel object-based superpixel segmentation framework which exploits the major features of DISF while incorporating prior object information as represented by an object saliency map. Similarly to DISF, ODISF variants perform oversampling to guarantee that relevant seeds are part of the initial seed set and delineate superpixels by using the IFT algorithm. The major contribution of ODISF is the novel object-based seed removal criterion, which favors seeds whose superpixels are placed nearby probable object boundaries. By strategically incorporating the object information only in the seed removal step, ODISF exploits the good estimation of the object location and is more robust to eventual delineation errors in the saliency map. Therefore, such pipeline not only provides accurate object delineation with low computational cost, but also manages to overcome the high dependency on the saliency map quality seen in OISF (see Fig. 1).

This paper is organized as follows. First, we discuss recent works in superpixel segmentation in Section II. Then, in Section III, we detail the mathematical framework used for presenting DISF and ODISF in Section IV. The experimental setup and results are shown in Section V and, finally, we draw conclusions and possible future work in Section VI.

II. Related Works

In this section, we present an overview of the state-of-the-art superpixel segmentation methods. For a deeper discussion, one may refer to notable surveys [7], [16]–[18]. In Section II-A, we review methods which are completely unaware of any object information. Then, in Section II-B we discuss methods based on deep learning. Finally, we present algorithms that consider object information independently from their source (Section II-C).

A. Classic and Content-Sensitive Methods

We may broadly classify such methods in two groups: (i) clustering-based methods; and (ii) graph-based methods. The former comprises those that solve superpixel segmentation by pixel clustering (*e.g.*, K-means, DBSCAN, GMMs). *Simple Linear Iterative Clustering* (SLIC) [6], the most popular method, solves such task by an adaptive K-means approach, resulting in low computational cost and fair object delineation. Inspired by SLIC, *Linear Spectral Clustering* (LSC) [11] maps every pixel into a 10-dimensional space and, subsequently, applies K-means. LSC shows significant improvement over SLIC in delineation, with the expense of being slightly slower. Similarly to LSC, *Intrinsic Manifold SLIC* (IMSLIC) [19] runs K-means in a two-dimensional manifold, where area indicates the content density. The majority of these methods (except IMSLIC) cannot guarantee connected superpixels in a given desired number, being necessary to apply a post-processing step that compromises the number of desired superpixels. Finally, due to strict constraints, such as the restricted search scope in the adapted K-means, adjustments for effectively considering object information may become impractical.

The second group of methods solve superpixel segmentation as a graph partitioning problem. *Entropy Rate Superpixels* (ERS) [8] generates superpixels by removing edges based on the entropy of a random walk in the graph. Although it presents high object adherence, it also presents high computational cost. *Superpixel Hierarchy* (SH) [20] computes a hierarchy of superpixels with high boundary adherence and low computational cost using the Borůkva algorithm. Yet, such methods do not consider object information in their computation.

A major subset of graph-based algorithms segments the image through path concatenation using the *Image Foresting Transform* (IFT) [14] algorithm. Such methods are often fast and present accurate object delineation. In [12], the authors propose a three-stage superpixel segmentation framework named *Iterative Spanning Forest* (ISF), whose steps are independently defined. This flexibility favored the development of recent and more effective approaches. The *Recursive ISF* [21] is a hierarchical superpixel segmentation method that applies ISF at each layer of the hierarchy. *Dynamic ISF* (DISF) [13] generates superpixels dynamically through initial seed oversampling with subsequent removal along iterations based on a predefined criterion. Such methods, however, do not consider object information in their computation.

B. Deep Learning Methods

Recently, a significant number of studies propose deep learning solutions for superpixel segmentation. In [22], the authors present an algorithm, named *Deep-FLIC*, that considers deep features as input for an adapted version of the *Fast Linear Iterative Clustering* (FLIC) [23] algorithm. Conversely, the authors in [24] debate over the efficiency of incorporating deep features and propose a neural network for computing pixel affinities in the image, which are used as input for a superpixel segmentation method (*e.g.*, ERS). Finally, we recall the *Superpixel Sampling Network* (SSN) [25] and the *Superpixel with Fully Convolutional Network* (S-FCN) [26] as representatives of end-to-end trainable superpixel segmentation methods. Notably, deep learning methods are strongly

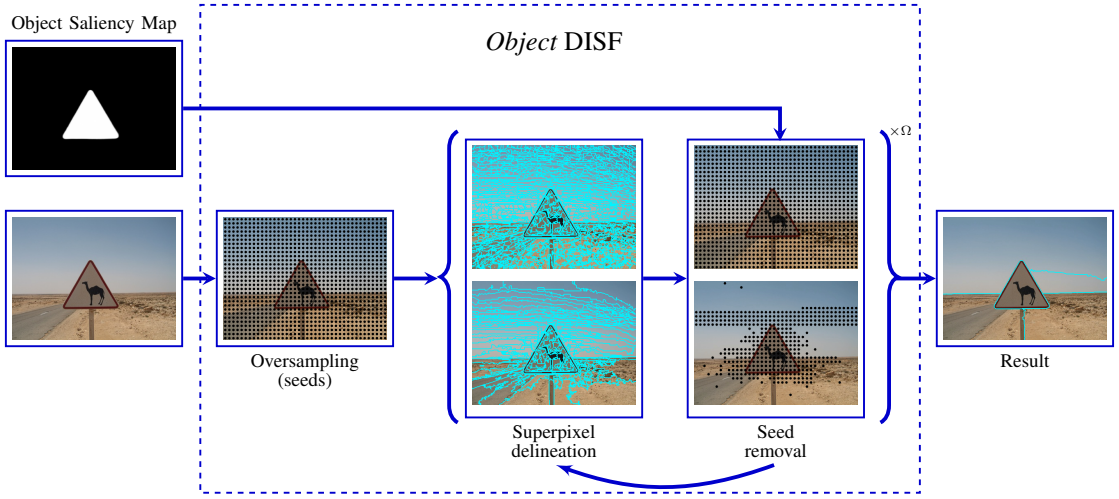


Fig. 2: Flowchart of the proposed Object-based DISF considering $N_0 = 1000$ and $N_f = 5$.

dependent on the amount of annotated training examples, which are usually scarce in scientific applications. Even if a large set is provided, one may argue that the profits of supervision lacks more evidence. Finally, for a different object of interest, the proposed model may not assure the reported performance even through fine-tuning.

C. Object-based Methods

Object-based algorithms generates superpixels by considering a prior information regarding the objects of interest. Similarly to ISF, the *Object-based ISF* (OISF) [10] is a three-stage object- and IFT-based superpixel segmentation framework. In OISF, the object information is provided through an object saliency map. This map is used in each of the three steps, permitting the user to control the superpixel morphology and displacement with respect to the borders of the map. As drawbacks, OISF variants are highly dependent of the quality of the saliency map, and their computational cost is high.

III. Theoretical Background

In this section, we briefly discuss the theoretical background. In Section III-A, we recall basic notions regarding images and graphs and, in Subsection III-B, we detail the *Image Foresting Transform* (IFT) [14], the core method of our approach.

A. Image Graph

Let an *image* I be a pair $\langle \mathcal{P}, \mathbf{F} \rangle$ in which $\mathcal{P} \subset \mathbb{Z}^2$ is the set of *picture elements* (*i.e.*, *pixels*) and $\mathbf{F}(p) \in \mathbb{R}^m$, for $m \in \mathbb{N}_{>0}$, consists in a particular and representative sequence of *features* (*e.g.* color) of a pixel p. If $m = 1$, I is a *grayscale* image; otherwise, I is a *colored* image (*e.g.*, CIELAB). Moreover, p can be represented a unique sequence of *(spatial) coordinates* $\langle x_p, y_p \rangle$. Finally, considering the previous definitions, we may define an *object saliency map* O as a grayscale image $\langle \mathcal{P}, \mathbf{O} \rangle$ in which $\mathbf{O}(p) \in [0, 1]$ maps every pixel p to a representative value that indicates its likelihood of belonging to an object of interest.

Given I, it is possible to build an *image graph* $G = \langle \mathcal{V}, \mathcal{E} \rangle$ such that $\mathcal{V} \subseteq \mathcal{P}$ is the set of *vertices* and $\mathcal{E} \subset \mathcal{V}^2$ is the set of *edges*. A classic approach for defining the edge set is through the *Euclidean (spatial) distance* between two distinct vertices u, v, with respect to a certain radius $r \in \mathbb{R}_{\geq 0}$. More specifically, $\mathcal{E} = \{\langle u, v \rangle \mid \|u - v\|_2 \leq r\}$. In this work, we use the 8-neighborhood (*i.e.*, $r = \sqrt{2}$). If $\langle u, v \rangle \in \mathcal{E}$, then u, v are *adjacents*. In this work, G is an *undirected* graph (*i.e.*, $\langle u, v \rangle \equiv \langle v, u \rangle$).

Consider a *path* $\rho = \langle v_i \rangle_{i=1}^{k}$ to be a sequence of distinct vertices such that, for $i < k$, v_i, v_{i+1} are adjacents. If $k = 1$, then ρ is a *trivial* path. We may explicitly exhibit the *origin* (or *root*) v_1 and the *terminus* v_k of ρ either by $\rho_{v_1 \rightsquigarrow v_k}$ or simply by ρ_{v_k}. Finally, the following notation $\rho_s \odot \langle s, t \rangle$ indicates a path resultant from a *concatenation* between a path ρ_s and an edge $\langle s, t \rangle$.

B. Image Foresting Transform

Several state-of-the-art algorithms use the *Image Foresting Transform* (IFT) [14] for generating superpixels [12], [13], [15], [21]. The IFT is a framework for the development of image operators based on connectivity and, in this work, we focus on its seed-restricted variant. For a given set $\mathcal{S} \subset \mathcal{V}$ of vertices (*i.e.*, *seeds*), the IFT finds optimum-paths in an non-decreasing order of cost from any $s \in \mathcal{S}$ to every $p \in \mathcal{V} \setminus \mathcal{S}$.

First, let Π be the set of all possible paths in G. Then, a *path-cost function* assigns a non-negative *path-cost value* $\mathbf{f}_*(\rho) \in \mathbb{R}$to any path $\rho \in \Pi$. The $\mathbf{f}_{max}$ is often chosen due to its effective performance in object delineation, and may be defined as follows (Eq. 1):

$$\mathbf{f}_{max}(\langle t \rangle) = \begin{cases} 0 & \text{if } t \in \mathcal{S} \\ +\infty & \text{otherwise} \end{cases} \\ \mathbf{f}_{max}(\rho_t \odot \langle t, s \rangle) = \max\{\mathbf{f}_{max}(\rho_t), \mathbf{w}_*(t,s)\} \tag{1}$$

in which $\mathbf{w}_*(t,s) \in \mathbb{R}$ defines a *cost* to an edge $\langle t, s \rangle \in \mathcal{E}$. An *optimum-path* ρ_t is a path in which, for any other $\tau_t \in \Pi$, $\mathbf{f}_*(\rho_t) \leq \mathbf{f}_*(\tau_t)$.

Let $\mathbf{P}$ be an acyclic map in which assigns a vertex $v \in \mathcal{V}$ to its *predecessor* $\mathbf{P}(v) = u \in \mathcal{V}$ in a unique path ρ_v or to a distinctive marker $\blacktriangle \notin \mathcal{V}$ when v is the root of ρ_v. $\mathbf{P}$ is said to be *optimum* if all of its paths are optimum. As one may note, we may assign every vertex v to its respective root $\mathbf{R}(v)$ defined in $\mathbf{P}$ through recursion. Therefore, for a given seed set $\mathcal{S}$ and a path-cost function $\mathbf{f}_*$, the IFT algorithm can output an optimum predecessor map $\mathbf{P}$ such that, for every seed $x \in \mathcal{S}$, it is defined an *optimum-path tree* $\mathcal{T}_x$ with paths that are more closely connected to x than to any other seed. In other words, the IFT minimizes a *path-cost map* $\mathbf{C}(v) = \min_{\rho_v \in \Pi}\{\mathbf{f}_*(\rho_v)\}$ and, consequently, builds $\mathbf{P}$ through path concatenation. In this work, every superpixel whose seed is $x \in \mathcal{S}$ is an optimum-path tree $\mathcal{T}_x$. Even if $\mathbf{f}_*$ is not *smooth* [27], the superpixel segmentation can be effective.

IV. Object-based DISF

In this section, we present the *Object-based Dynamic and Iterative Spanning Forest* (ODISF) method, illustrated in Figure 2, by making clear its differences with respect to the DISF approach [13]. In Section IV-A, we discuss two different strategies for seed oversampling and, subsequently in Section IV-B, we revisit the concept of dynamic edge-cost estimation for superpixel generation using the IFT framework. Finally, we detail our proposed object-based seed removal strategy for a given object saliency map in Section IV-C.

A. Seed Oversampling

For a given number $N_f > 0$ of desired superpixels, the first step of seed-based methods is to estimate an initial seed set $\mathcal{S}$. Differently from most methods [6], [11], [13], DISF and ODISF start off from seed oversampling such that $|\mathcal{S}| = N_0 \gg N_f$. This strategy significantly increases the probability of selecting, in the initial seed set, all relevant seeds for solving the problem.

Most methods [6], [11], [13] initially distribute seeds equidistantly in a grid pattern. Such strategy, hereafter named GRID, first estimates the expected superpixel size a for an image $I = \langle \mathcal{P}, \mathbf{F} \rangle$ as $a = |\mathcal{P}|/N_f$. Then, it establishes that seed pairs must be distanced by $\sqrt{a}$. Finally, in order to avoid seeds placed over borders, every seed $s \in \mathcal{S}$ is shifted to a position whose gradient is the lowest within its neighborhood.

In [28], the authors evaluate the impact of object-based seed sampling strategies for superpixel segmentation and conclude that a high concentration of seeds within the object leads to better results in delineation. However, these strategies are computationally expensive. On the other hand, the combination of seed oversampling and an accurate local criterion for seed removal along iterations may result into similar effectiveness with considerable efficiency gains. For example, one can preserve seeds near the borders of a given object saliency map in the seed removal step. We explore this strategy and, moreover, analyze if a random selection of seeds (hereafter named RND) is equivalent to GRID when using seed oversampling.Unlike GRID, RND does not require computing the image gradient for seed perturbation, and it is straightforward to implement for non-rectangular masks.

B. Superpixel Generation

Both DISF and ODISF use the IFT algorithm with path-cost function $\mathbf{f}_{max}$ and on-the-fly edge-cost estimation, as proposed in [29] for interactive object segmentation. Let $\mu_{\mathbf{F}}(\mathcal{T}_x) = \sum_{v \in \mathcal{T}_x} \mathbf{F}(v)/|\mathcal{T}_x|$ be the mean feature vector of a *growing* optimum-path tree rooted in x. The dynamic edge-cost function is defined as $\mathbf{w}(u,v) = \|\mu_{\mathbf{F}}(\mathcal{T}_x) - \mathbf{F}(v)\|_2$ for $x = \mathbf{R}(u)$ by the time the path ρ_u is optimum during the IFT algorithm.

Whenever the object saliency values in $\mathbf{O}$ are derived from $\mathbf{F}$, both should contain the same information about the boundaries of interest. In [15], the authors use a path-cost function with an object-based edge-cost estimation in which it is possible to control boundary adherence with respect to the borders in $\mathbf{O}$ and $\mathbf{F}$. As Figure 4 shows, for imperfect saliency maps, this strategy can negatively affect superpixel delineation. In such a case, the border information in $\mathbf{F}$ should be sufficient for accurate superpixel delineation. In our framework, we prefer to constrain the use of the object saliency map for seed removal (Section IV-C).

C. Object-based Seed Removal

In DISF, due to seed oversampling, the number of iterations is such that N_f seeds (superpixels) must result at the last iteration. At each iteration $i \in \mathbb{N}_{>0}$, $\mathbf{M}(i) = \max\left\{N_0 \exp^{-i}, N_f\right\}$ seeds are selected from $\mathcal{S}$ for the delineation step in iteration $i+1$, while the remaining are discarded. This process is repeated until $N_i = N_f$, resulting in Ω iterations. For most practical cases, $\Omega = 5$.

One approach for selecting $\mathbf{M}(i)$ seeds is to assign a *relevance value* $\mathbf{V}_*(s)$ to each $s \in \mathcal{S}$ based on the characteristics of its resulting superpixel $\mathcal{T}_s$. Two distinct trees $\mathcal{T}_x, \mathcal{T}_y$ are said *adjacent* if $\exists \langle u, v \rangle \in \mathcal{E}$ such that $u \in \mathcal{T}_x$ and $v \in \mathcal{T}_y$. Let $\mathcal{B}$ be the set of pairs $\langle \mathcal{T}_x, \mathcal{T}_y \rangle$ of adjacent trees within the forest. Then, we may define $\mathbf{V}_1(s)$ as a combination of size and contrast of $\mathcal{T}_s$ as presented in Eq. 2

$$\mathbf{V}_1(s) = \frac{|\mathcal{T}_s|}{|\mathcal{V}|} \min_{\forall \langle \mathcal{T}_s, \mathcal{T}_r \rangle \in \mathcal{B}} \{\|\mu_{\mathbf{F}}(\mathcal{T}_s) - \mu_{\mathbf{F}}(\mathcal{T}_r)\|_2\} \tag{2}$$

The major drawback of $\mathbf{V}_1$ is not distinguishing whether a superpixel is near an object border. Obviously, the size of a superpixel does not assist in such determination. As one may

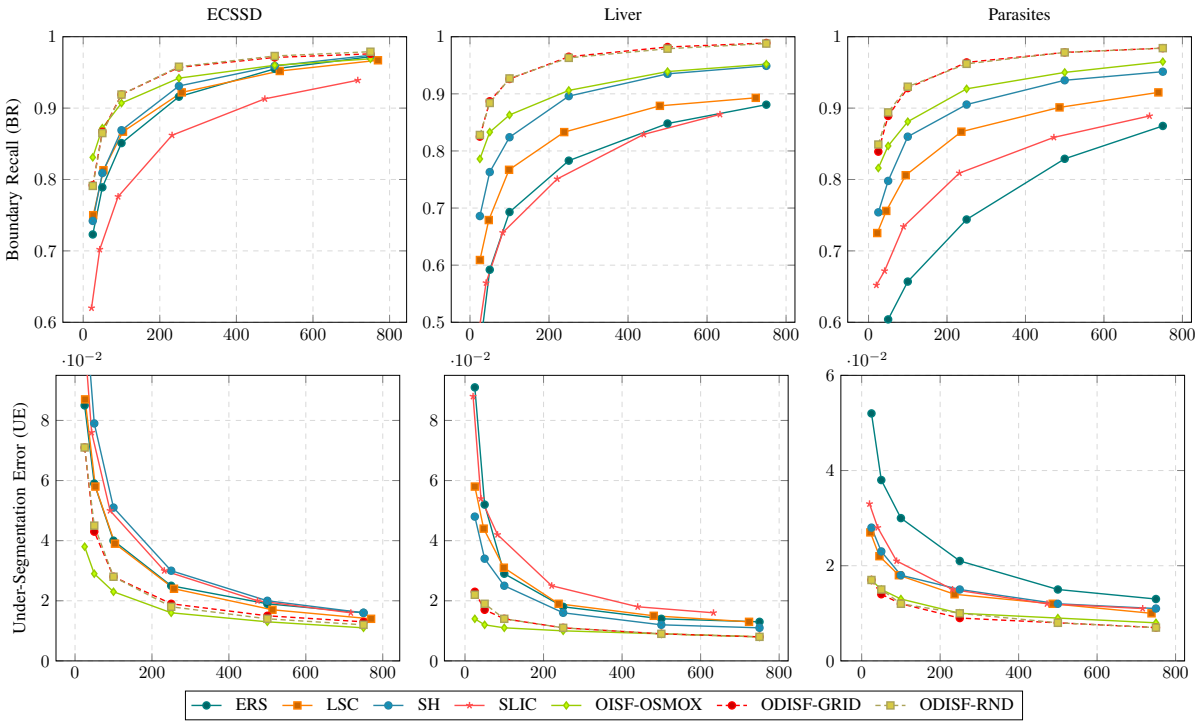


Fig. 3: Results obtained for ECSSD, Liver and Parasites, respectively. For all methods, the default configuration was set.

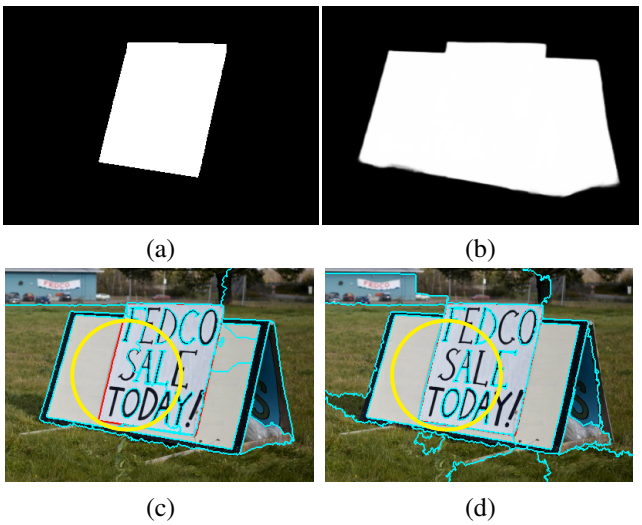


Fig. 4: Results of object-based superpixel segmentation methods considering 25 superpixels. (a) Ground-truth image; (b) Object saliency map [9]; Segmentation results (in cyan), drawn over (a), obtained by (c) OISF [10] and (d) our proposal.

note, the contrast information is also insufficient since low contrast regions are present within the whole image. Thus, inspired by [30], we propose an *object-based seed relevance* criterion $\mathbf{V}_2$ (Eq. 3)

$$\begin{aligned} \mathbf{G}(\mathcal{T}_s) &= \max_{\forall \langle \mathcal{T}_s, \mathcal{T}_r \rangle \in \mathcal{B}} \{ \| \mu_{\mathbf{O}}(\mathcal{T}_s) - \mu_{\mathbf{O}}(\mathcal{T}_r) \|_1 \} \\ \mathbf{V}_2(s) &= \mathbf{V}_1(s) \max \{ \mu_{\mathbf{O}}(\mathcal{T}_s), \mathbf{G}(\mathcal{T}_s) \} \end{aligned} \tag{3}$$

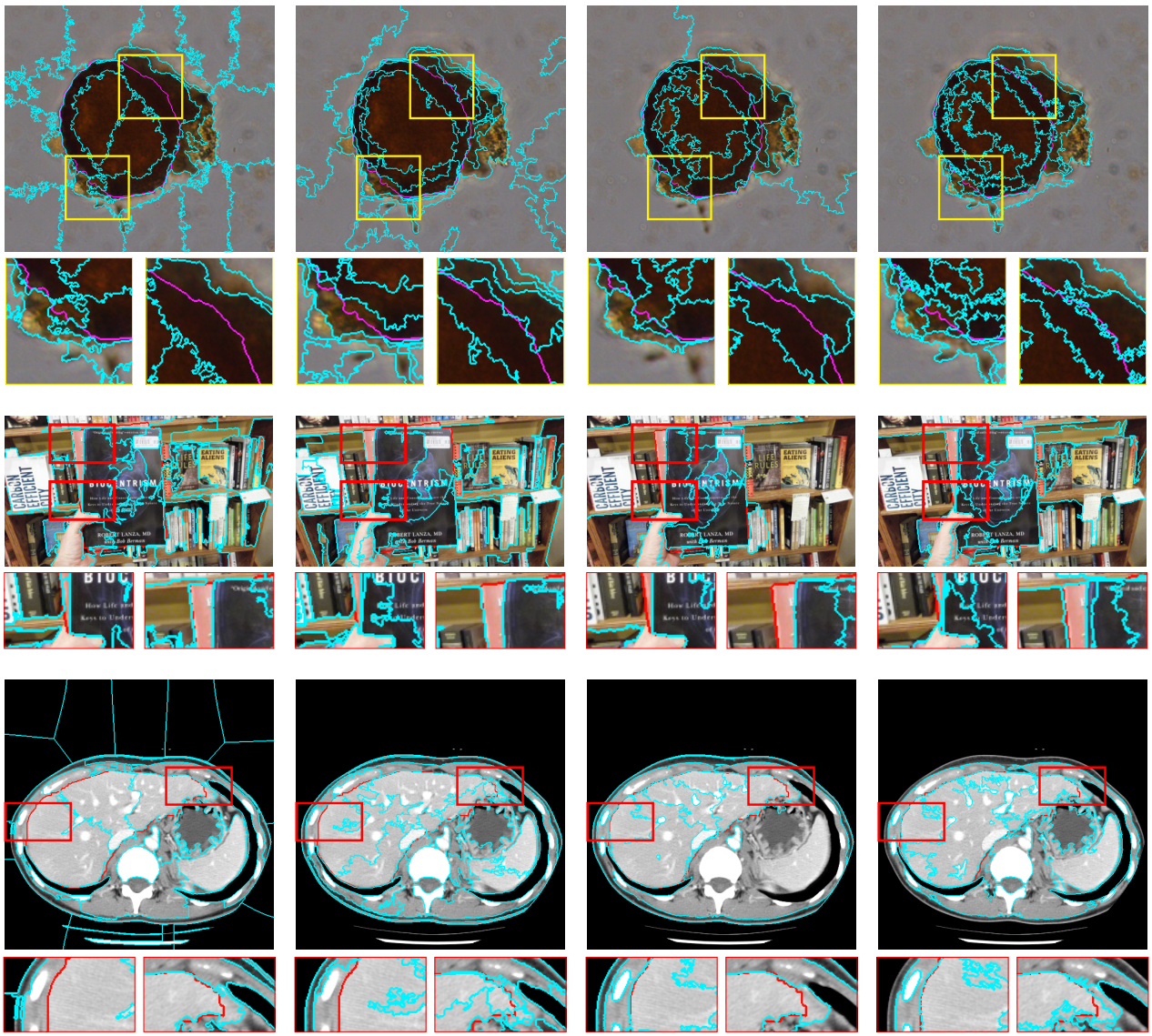
in which, analogously to $\mu_{\mathbf{F}}$, $\mu_{\mathbf{O}}(\mathcal{T}_s) = \sum_{v \in \mathcal{T}_s} \mathbf{O}(v)/|\mathcal{T}_s|$ is the mean saliency value of $\mathcal{T}_s$. In contrast to other object-based methods, which promotes a simple concentration of superpixels within (or outside) the objects of interest, our proposed function favors those nearby a probable object border, promoting seed competition in crucial regions, or within a region with high object certainty (*i.e.*, high saliency values).

V. Experimental Results

In this section, we present the experimental setup and the results obtained by our approach. In Section V-A, we present the baselines, the datasets and the evaluation metrics considered for benchmark. Subsequently, in Sections V-C and V-B, we present and illustrate the performance of all methods.

A. Experimental Setup

Most works evaluate their proposals in a contour-driven dataset [31] in which all relevant borders are desired. However, as one may note, our proposal aims the accurate delineation of the object of interest. Therefore, we chose three datasets from distinct domains and with different object properties for evaluating the robustness of ODISF. The *Extended Complex Saliency Scene Dataset* (ECSSD) [32] is a popular dataset for salient object detection and consists of 1000 natural images with diverse objects. The *Liver* [12] dataset is composed of 40 CT slices of the human liver, which imposes a great challenge for being a grayscale object with smooth borders. Similarly, the *Parasite* [15] dataset contains 72 images of colored objects (*i.e.*, helminth eggs) with smooth borders in which an impurity may be attached. For all datasets, we randomly defined 70%



(a) LSC (b) SH (c) OISF-OSMOX (d) ODISF-RND

Fig. 5: Segmentation results (in cyan), obtained for each method, overlaying the ground-truth (in red and magenta). A number $N_f = 25$ of superpixels was required, and the default parameter setting was used.

as test set. For generating object saliency maps, we considered a recent deep neural network approach [9][1], suitable for small datasets, whose detection accuracy is on par with many other deep-learning approaches. We trained such estimator in the remaining 30% of the dataset.

We selected five state-of-the-art methods in which the last one is, to the best of our knowledge, the only object-based method in literature: (i) SLIC [6][2]; (ii) ERS [8][3]; (iii) LSC [11][4]; (iv) SH [20][5]; and (v) OISF-OSMOX [10][6]. All methods were selected due to their performance in superpixel delineation and, for that, the recommended parameter configuration was set. For our approach, we set $N_0 = 8000$ (as in [13]) and evaluate ODISF with two options for seed oversampling: ODISF-GRID and ODISF-RND. The code of ODISF is available online [7]. The performances were assessed by two classic metrics. *Boundary Recall* (BR) [7] measures the percentage of object boundaries correctly overlapped by a superpixel border (*i.e.*, higher is better). On the other hand,

[1]https://github.com/xuebinqin/U-2-Net
[2]https://www.epfl.ch/labs/ivrl/research/slic-superpixels/
[3]https://github.com/mingyuliutw/EntropyRateSuperpixel
[4]https://jschenthu.weebly.com/projects.html
[5]https://github.com/semiquark1/boruvka-superpixel
[6]https://github.com/LIDS-UNICAMP/OISF
[7]https://github.com/LIDS-UNICAMP/ODISF

Under-Segmentation Error (UE) [18] measures the error of multiple ground-truth overlapping by superpixels (*i.e.*, lower is better). Since we do not aim for compact superpixels, and due to the high correlation between different superpixel metrics [7], we argue that the both selected are sufficient for a proper evaluation. We evaluated all methods considering $N_f \in [25, 750]$.

B. Quantitative Results

Figure 3 shows the results for all datasets. It is possible to see that ODISF surpasses all baselines with significant margin, especially in boundary adherence. Table I shows the average computational time of each method using a 64bit, Intel (R) Core(TM) i5-5200U PC with CPU speed of 2.20Ghz. Knowing that SLIC, LSC, and SH are $O(N)$ complex, considering N to be the number of pixels in the image, it is no surprise that they are the fastest baselines. When comparing to OISF, ODISF presents a significant speedup of 3.05 (for $N_f = 25$) in Parasites, with on pair performance in UE. It is important to emphasize that, although both ODISF and OISF are $O(N \log N)$, OISF-OSMOX exploits parallelism during sampling and computes differential IFTs (DIFTs) [33] for speed optimization. Therefore, since ODISF is an IFT-based method, our implementation is eligible for computing DIFTs and further improve its speed.

Similarly to [34], we found no statistical difference between ODISF-GRID and ODISF-RND for both metrics and, furthermore, both achieve similar performance to object-based seed sampling approaches [28]. Thus, we may conclude that the oversampling step can be significantly simplified without prejudicing the ODISF performance.

Method	Time(sec)		
	$N_f = 25$	$N_f = 250$	$N_f = 750$
SLIC	0.559±0.077	0.578±0.085	0.581±0.084
ERS	2.591±1.305	2.879±1.407	3.166±1.561
LSC	0.944±0.247	0.965±0.251	0.983±0.259
OISF-OSMOX	13.824±9.589	6.619±4.624	5.128±3.424
SH	0.991±0.067	0.990±0.064	0.974±0.070
ODISF-RND	4.540±2.625	3.612±1.949	2.945±1.639

TABLE I: Average processing time for superpixel segmentation in the Parasites dataset.

Interestingly, OISF, whose edge-cost estimation is object-based, presents performance inferior to ODISF, which considers the input image features only. Although there are other differences in their connectivity function, the neural network may generate imperfect object saliency maps and the use of this map restricted to the seed removal policy seems to make ODISF more robust than OISF. We argue that our object-based seed removal policy is effective to preserve relevant seeds and robust to imperfect saliency maps that at least locate the object of interest.

C. Qualitative Results

Figure 5 illustrates the segmentation results obtained by each method for the Parasite and ECSSD datasets. The strategy used by LSC yields effective results in natural images but, as one may note, it can lead to critical errors in different domains. On the other hand, SH shows a consistent performance irrespective of the image domain, justifying its position as one of the top superpixel methods. The results obtained by OISF-OSMOX illustrates its high dependency on the quality of the map by presenting high adherence to the borders of the map, including both correct and incorrect estimations. Finally, it is possible to see that ODISF segmentation manages to best approximate the object borders while it prevents severe superpixel leakings present in the results of the previous methods.

VI. Conclusion and Future Work

In this paper, we propose a novel object-based superpixel segmentation framework inspired in a recent superpixel method named *Dynamic and Iterative Spanning Forest* (DISF) [13]. Our proposal, *Object-based DISF* (ODISF), exploits the positive aspects of DISF, such as seed oversampling and dynamic edge-cost estimation, while it offers a more effective and robust seed removal criterion based on prior object information. Experimental results show that ODISF variants can surpass state-of-the-art algorithms in superpixel delineation by preventing segmentation errors, while being faster than its object-based counterpart [10], [15].

For future endeavors, we intent to study different curves for establishing the number of seeds at each iteration. Similarly to [30], we also desire to investigate the applicability of ODISF for interactive object segmentation based on user-drawn markers.

Acknowledgment

The authors thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – (PQ 303808/2018-7, 310075/2019-0), the Fundação de Amparo a Pesquisa do Estado de Minas Gerais – FAPEMIG – (Grants PPM-00006-18), the Fundação de Amparo a Pesquisa do Estado de São Paulo – FAPESP – (2014/12236-1) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant COFECUB 88887.191730/2018-00) for the financial support. Lastly, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance code 001.

References

[1] S. Martins, G. Ruppert, F. Reis, C. Yasuda, and A. Falcão, "A supervoxel-based approach for unsupervised abnormal asymmetry detection in MR images of the brain," in *16th International Symposium on Biomedical Imaging (ISBI)*, 2019, pp. 882–885.

[2] A. Sousa, S. Martins, A. Falcão, F. Reis, E. Bagatin, and K. Irion, "Altis: A fast and automatic lung and trachea ct-image segmentation method," *Medical Physics*, vol. 46, no. 11, pp. 4970–4982, 2019.

[3] J. Zhou, J. Ruan, C. Wu, G. Ye, Z. Zhu, J. Yue, and Y. Zhang, "Superpixel segmentation of breast cancer pathology images based on features extracted from the autoencoder," in *11th International Conference on Communication Software and Networks (ICCSN)*, 2019, pp. 366–370.

[4] Y. Yu, Y. Makihara, and Y. Yagi, "Pedestrian segmentation based on a spatio-temporally consistent graph-cut with optimal transport," *Transactions on Computer Vision and Applications*, vol. 11, no. 1, p. 10, 2019.

[5] S. Zhang, H. Wang, W. Huang, and Z. You, "Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG," *Optik*, vol. 157, pp. 866–872, 2018.

[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[7] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018.

[8] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *24th Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2097–2104.

[9] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested u-structure for salient object detection," *Pattern Recognition*, vol. 106, p. 107404, 2020.

[10] F. Belém, S. Guimarães, and A. Falcão, "Superpixel generation by the iterative spanning forest using object information," in *33rd Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2020, pp. 22–28, workshop of Thesis and Dissertations.

[11] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *28th Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1356–1363.

[12] J. Vargas-Muñoz, A. Chowdhury, E. Alexandre, F. Galvão, P. Miranda, and A. Falcão, "An iterative spanning forest framework for superpixel segmentation," *Transactions on Image Processing*, vol. 28, no. 7, pp. 3477–3489, 2019.

[13] F. Belém, S. Guimarães, and A. Falcão, "Superpixel segmentation using dynamic and iterative spanning forest," *Signal Processing Letters*, vol. 27, pp. 1440–1444, 2020.

[14] A. Falcão, J. Stolfi, and R. Lotufo, "The image foresting transform: Theory, algorithms, and applications," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 19–29, 2004.

[15] F. Belém, S. Guimarães, and A. Falcão, "Superpixel segmentation by object-based iterative spanning forest," in *23rd Iberoamerican Congress on Pattern Recognition*, 2018, pp. 334–341.

[16] A. Schick, M. Fischer, and R. Stiefelhagen, "An evaluation of the compactness of superpixels," *Pattern Recognition Letters*, vol. 43, pp. 71–80, 2014.

[17] L. Wan, X. Xu, Q. Zhao, and W. Feng, "Spherical superpixels: Benchmark and evaluation," in *14th Asian Conference on Computer Vision (ACCV)*, 2018, pp. 703–717.

[18] P. Neubert and P. Protzel, "Superpixel benchmark and comparison," in *Forum Bildverarbeitung*, vol. 6, 2012, pp. 1–12.

[19] Y. Liu, M. Yu, B. Li, and Y. He, "Intrinsic manifold SLIC: A simple and efficient method for computing content-sensitive superpixels," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 653–666, 2018.

[20] X. Wei, Q. Yang, Y. Gong, N. Ahuja, and M. Yang, "Superpixel hierarchy," *Transactions on Image Processing*, vol. 27, no. 10, pp. 4838–4849, 2018.

[21] F. Galvão, A. Falcão, and A. Chowdhury, "RISF: recursive iterative spanning forest for superpixel segmentation," in *31st Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018, pp. 408–415.

[22] M. Awaisu, L. Li, J. Peng, and J. Zhang, "Fast superpixel segmentation with deep features," in *36th Computer Graphics International Conference (CGI)*, 2019, pp. 410–416.

[23] J. Zhao, R. Bo, Q. Hou, M. Cheng, and P. Rosin, "FLIC: Fast linear iterative clustering with active search," *Computational Visual Media*, vol. 4, no. 4, pp. 333–348, 2018.

[24] W. Tu, M. Liu, V. Jampani, D. Sun, S. Chien, M. Yang, and J. Kautz, "Learning superpixels with segmentation-aware affinity loss," in *31st Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 568–576.

[25] V. Jampani, D. Sun, M. Liu, M. Yang, and J. Kautz, "Superpixel sampling networks," in *18th European Conference on Computer Vision (ECCV)*, 2018, pp. 352–368.

[26] F. Yang, Q. Sun, H. Jin, and Z. Zhou, "Superpixel segmentation with fully convolutional networks," in *33rd Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[27] K. Ciesielski, A. Falcão, and P. Miranda, "Path-value functions for which Dijkstra's algorithm returns optimal mapping," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 7, pp. 1025–1036, 2018.

[28] F. Belém, L. Melo, S. Guimarães, and A. Falcão, "The importance of object-based seed sampling for superpixel segmentation," in *32nd Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2019, pp. 108–115.

[29] J. Bragantini, S. Martins, C. Castelo-Fernandez, and A. Falcão, "Graph-based image segmentation using dynamic trees," in *23rd Iberoamerican Congress on Pattern Recognition*, 2018, pp. 470–478.

[30] I. Borlido, F. Belém, P. Miranda, A. Falcão, Z. Patrocínio, and S. Guimarães, "Towards interactive image segmentation by dynamic and iterative spanning forest," in *Discrete Geometry and Mathematical Morphology*, 2021, pp. 351–364.

[31] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.

[32] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended cssd," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 717–729, 2015.

[33] A. Falcão and F. Bergo, "Interactive volume segmentation with differential image foresting transforms," *Transactions on Medical Imaging*, vol. 23, no. 9, pp. 1100–1108, 2004.

[34] C. Jerônimo, F. Belém, S. Carneiro, Z. Patrocínio, L. Najman, A. Falcão, and S. Guimarães, "Graph-based supervoxel computation from iterative spanning forest," in *Discrete Geometry and Mathematical Morphology*, 2021, pp. 404–415.