Unsupervised Image Segmentation by Oriented Image Foresting Transform in Layered Graphs

Felipe A.S. Kleine*, Luiz F.D. Santos†, Fábio A.M. Cappabianco‡ and Paulo A.V. Miranda†
*Researcher at the IPT - Institute for Technological Research of the State of São Paulo, Brazil
†University of São Paulo, Institute of Mathematics and Statistics, São Paulo, SP, Brazil, Email: pmiranda@ime.usp.br
‡Instituto de Ciência e Tecnologia, São José dos Campos, SP, Brazil

***Abstract*—In this work, we address the problem of unsupervised image segmentation, subject to high-level constraints expected for the objects of interest. More specifically, we handle the segmentation of a hierarchy of objects with nested boundaries, each with its own expected boundary polarity constraint. To this end, this work successfully extends Hierarchical Layered Oriented Image Foresting Transform (HLOIFT), with the inclusion of nested object relations, to the unsupervised segmentation paradigm. On the other hand, this work can also be seen as an extension of Unsupervised OIFT (UOIFT) to include structural relationships of nested objects.**

The method is demonstrated in the segmentation of three datasets of colored images with superior performance compared to other existing techniques in graphs, requiring a smaller number of connected partitions to isolate the objects of interest in the images.

I. Introduction

Image segmentation is among the most tackled problems in computer vision, image processing, and image analysis. It is a fundamental procedure widely used in the area of autonomous vehicles and robotics for object detection and content recognition [1], [2], disease identification and etiology study in the field of medicine [3]–[5], crop area measurement for agriculture [6], [7], and many other applications [8]–[10].

There are several distinct possible image segmentation approaches, which depend on the goal application. Semantic segmentation consists of labeling all objects of the same class with the same color [11]. Instance segmentation, however, marks every instance of objects with a single different color [12]. Salient object segmentation focus on labeling objects which are more visible from a human visual perspective [13]. Another important taxonomy of image segmentation concerns the amount of supervision it requires. Unsupervised segmentation does not require any prior training [14], self-supervised segmentation uses unlabeled training data [15], supervised segmentation uses labeled data for training [16], and semi-supervised segmentation uses both labeled and unlabeled data for training [17].

Finally, there is an effort in segmenting objects based on known priors. These priors describe relationships among objects such are their relative pose, size [18], boundary polarity [19], [20] and hierarchy [21]. For some applications, objects may also assume shape restrictions [22].

In recent years, most of the community effort applies to general semantic or instance segmentation for natural images. Deep learning methods are by far the most suitable for these applications [23]. In the case of more specific areas such as medical imaging, methods include priors as input information to achieve better performance [24]. There is a large gap in the literature though with respect to the application of general image segmentation with priors. That is because generic priors are more abstract information, making it unfeasible to train networks for that purpose. Even using billions of training images, one can not train current methods to detect any set of objects with a specific boundary polarity, with a relative size and pose in an unsupervised way, and that is exactly the class of problem we tackle in this paper.

We propose an unsupervised segmentation method which accepts as priors the hierarchy of objects and their boundary polarity. In this way, it extends both the *Hierarchical Layered Oriented Image Foresting Transform* (HLOIFT) [21], with the inclusion of nested object relations, to the unsupervised segmentation paradigm and the *Unsupervised OIFT* (UOIFT) [20] by including structural relationships of nested objects.

The reminder of the paper is structured as follows: Section II describes the notations and definitions for our algorithm; Section III reviews previous works extended by this paper; in Section IV, we present our new segmentation method; and in Sections V and VI we show the experimental results and state our conclusions.

II. Notations and definitions

A partition of the finite set $\mathcal{I}$ is a set $\mathcal{P}$ of disjoint nonempty subsets of $\mathcal{I}$ whose union is $\mathcal{I}$ (that is, $\forall X, Y \in \mathcal{P}$, $X \cap Y = \emptyset$ if $X \neq Y$ and $\cup\{X \in \mathcal{P}\} = \mathcal{I}$). Any element of a partition $\mathcal{P}$ of $\mathcal{I}$ is called a region of $\mathcal{P}$.

We consider a weighted digraph G as a triple $\langle \mathcal{N}, \mathcal{A}, \omega \rangle$, where $\mathcal{N}$ is a nonempty set of vertices or nodes, $\mathcal{A}$ is a set of ordered pairs of distinct vertices called arcs or directed edges, and $\omega : \mathcal{A} \to \mathbb{R}$ represents the weights associated with the arcs. The digraph G is symmetric if for any of its arcs $\langle s, t \rangle \in \mathcal{A}$, the pair $\langle t, s \rangle$ is also an arc of G, but we can have $\omega(\langle s, t \rangle) \neq \omega(\langle t, s \rangle)$. The transpose G^T of G is the unique weighted digraph on the same set of vertices $\mathcal{N}$ with all arcs (and corresponding weights) reversed compared to the corresponding arcs in G. For a given graph $G = \langle \mathcal{N}, \mathcal{A}, \omega \rangle$, a path $\pi = \langle t_1, t_2, \ldots, t_n \rangle$ is a sequence of adjacent nodes (*i.e.*, $\langle t_i, t_{i+1} \rangle \in \mathcal{A}$, $i = 1, 2, \ldots, n-1$) with no repeated vertices ($t_i \neq t_j$ for $i \neq j$). A path $\pi_t = \langle t_1, t_2, \ldots, t_n = t \rangle$ is a path

with terminus at a node t. A path is *trivial* when $\pi_t = \langle t \rangle$. A path $\pi_t = \pi_s \cdot \langle s, t \rangle$ indicates the extension of a path π_s by an arc $\langle s, t \rangle$.

Given an unweighted and symmetric graph $G = \langle \mathcal{N}, \mathcal{A} \rangle$, a subset T of $\mathcal{N}$ is connected if, for any two vertices x and y of T, there exists a path from x to y in G that only passes through vertices in T. Given a graph $G = \langle \mathcal{N}, \mathcal{A} \rangle$, a partition of $\mathcal{N}$ is connected (in G) if all its regions are connected.

Let $\mathcal{I}$ be the image domain (that is, the set of pixels in $\mathbb{Z}^2$) and let $\mathcal{P}_\mathcal{I}$ be an initial connected partition (considering a 4-neighborhood graph) of $\mathcal{I}$, designed to merge neighboring pixels with similar intensity and color characteristics into a same region, called a superpixel [25]. The image can then be interpreted as a weighted digraph $G = \langle \mathcal{N}, \mathcal{A}, \omega \rangle$, whose nodes $\mathcal{N} = \mathcal{P}_\mathcal{I}$ are the superpixels and whose arcs are the ordered pairs $\langle s, t \rangle \in \mathcal{A}$ of neighboring superpixels, forming a *Region Adjacency Graph* (RAG).

A *connectivity function* $f : \Pi(G) \to \mathbb{R}$ computes a value $f(\pi_t)$ for any path π_t, usually based on arc weights, where $\Pi(G)$ indicates the set of all possible paths in a graph G. A path π_t is *optimum* if $f(\pi_t) \leq f(\tau_t)$ for any other path $\tau_t \in \Pi(G)$. Connectivity functions are usually described based on a path-extension rule. For instance, the max-arc connectivity function $f_{\max}$ is given by:

$$\begin{aligned} f_{\max}(\langle t \rangle) &= \begin{cases} -1 & \text{if } t \in \mathcal{S} \\ +\infty & \text{otherwise} \end{cases} \\ f_{\max}(\pi_s \cdot \langle s, t \rangle) &= \max\{f_{\max}(\pi_s), \omega(\langle s, t \rangle)\} \end{aligned} \quad (1)$$

where $\mathcal{S}$ is a seed set.

III. Background

Next, we present the two related methods that are relevant to the present work. In order to simplify the current exposition, the former will be presented for a general graph, while the latter will be presented in a layered graph, where each layer is a RAG of superpixels.

A. Oriented Image Foresting Transform (OIFT)

Image segmentation can be formulated as a graph partition problem subject to hard constraints. In the case of binary segmentation, we consider two non-empty disjoint seed sets $\mathcal{S}_0, \mathcal{S}_1 \subset \mathcal{N}$ indicating, respectively, background O_0 and object $O_1 = \mathcal{N} \setminus O_0$, such that $\mathcal{S}_1 \subset O_1$ and $\mathcal{S}_0 \subset O_0$. The object O_1 is identified with its *labeling* $X \colon \mathcal{N} \to \{0, 1\}$, so that $O_1 = \{v \in \mathcal{N} \colon X(v) = 1\}$.

The partial labeling, $X(t) = 1$ for all $t \in \mathcal{S}_1$ and $X(t) = 0$ for all $t \in \mathcal{S}_0$, given by the seeds, is propagated to all unlabeled nodes during the OIFT algorithm [19]. The resulting segmentation by OIFT gives, subject to the seed constraints, a global optimum solution by maximizing the graph-cut measure ε_{min} defined as

$$\varepsilon_{min}(X) = \min\{\omega(\langle s, t \rangle) \colon \langle s, t \rangle \in \mathcal{A} \ \&\ X(s) > X(t)\}. \quad (2)$$

The OIFT segmentation, indicated by X, can be build upon the *Image Foresting Transform* framework (IFT) [26] by considering a proper connectivity function in a connected and symmetric digraph G, as described in [19].

B. Hierarchical Layered Oriented Image Foresting Transform (HLOIFT)

Let $\mathcal{L} = \{1, \ldots, m\}$ denote an index set, where each element in $\mathcal{L}$ is associated with an object to be segmented and m is the number of objects. The HLOIFT graph associated with $\mathcal{L}$ and an image with superpixels $\mathcal{P}_\mathcal{I}$ will be defined on the set of nodes $\mathcal{N} = \mathcal{L} \times \mathcal{P}_\mathcal{I}$. The HLOIFT resulted segmentation of the image will be identified with a binary variable $X \colon \mathcal{N} \to \{0, 1\}$, where, for $i \in \mathcal{L}$, the ith object O_i and the background O_0 are defined, in superpixel resolution, respectively, as

$$O_i = \{t \in \mathcal{P}_\mathcal{I} \colon X(i, t) = 1\} \quad \text{and} \quad O_0 = \mathcal{P}_\mathcal{I} \setminus \textstyle\bigcup_{i \in \mathcal{L}} O_i. \quad (3)$$

Each object/background object O_i, $i \in \mathcal{L} \cup \{0\}$, will be identified with a corresponding set $\mathcal{S}_i \subset \mathcal{P}_\mathcal{I}$ of seeds, aiming for $\mathcal{S}_i \subseteq O_i$.

The hierarchy between the objects is understood as a prior knowledge on any pair $\langle O_i, O_j \rangle$ of objects we consider: either $O_i \cap O_j = \emptyset$ (exclusion relation), or one of them is properly contained in the other (inclusion relation). Here, we consider only the inclusion relation, represented as a function $h \colon \mathcal{L} \to \mathcal{L}$, so that $h(i) = j$ if, and only if, O_j is the smallest of the objects properly containing O_i. If $h(i) = j$, then we will refer to O_j as the *parent* of O_i. In this work, we consider a set of objects with nested boundaries, such that $h(i) = i + 1$, $i = 1, \ldots, m - 1$.

The first step of HLOIFT is to create a set of m layers, where each layer $\mathcal{H}_i$, $i \in \mathcal{L}$, is used to represent a single corresponding object O_i. A layer $\mathcal{H}_i = \langle \mathcal{N}_i, \mathcal{A}_i, \omega_i \rangle$ is a weighted digraph, where $\mathcal{N}_i = \{i\} \times \mathcal{P}_\mathcal{I}$ and each node $t = (i, v) \in \mathcal{N}_i$ corresponds to the image superpixel $p(t) = v$. Thus, the node set $\mathcal{N}$ of HLOIFT digraph is defined as $\mathcal{L} \times \mathcal{P}_\mathcal{I} = \bigcup_{i \in \mathcal{L}} \mathcal{N}_i$ and $p \colon \mathcal{N} \to \mathcal{P}_\mathcal{I}$ is the projection onto the second coordinate, while $\lambda \colon \mathcal{N} \to \mathcal{L}$ will denote the projection onto the first coordinate, that is, $\lambda(t) = i$ means that t belongs to the ith layer of the graph. We define the set of intra-layer arcs $\mathcal{A}_i$ on $\mathcal{H}_i$, $i = 1, \ldots, m$, as $\langle s, t \rangle \in \mathcal{A}_i$ if, and only if, $p(s)$ and $p(t)$ are neighboring superpixels in the RAG of superpixels. Regarding the weight function ω_i, it should highlight the desired boundaries for O_i as clearly as possible and we would like to incorporate in its definition the higher level priors whenever it is appropriate. In particular, to utilize the object-contour orientations, that is, the **boundary polarity priors**, for colored images, we use in our experiments ω_i as defined by the following formula:

$$\omega_i(\langle s, t \rangle) = \begin{cases} \|I(t) - I(s)\| \times (1 + \alpha_i) & \text{if } l_s > l_t \\ \|I(t) - I(s)\| \times (1 - \alpha_i) & \text{if } l_s < l_t \\ \|I(t) - I(s)\| & \text{otherwise} \end{cases} \quad (4)$$

where $I(s) = \langle l_s, a_s, b_s \rangle$ and $I(t) = \langle l_t, a_t, b_t \rangle$ are the mean colors of superpixels $p(s)$ and $p(t)$ in CIELAB color space, the symbol $\|\cdot\|$ denotes the vector norm, and α_i is a polarity parameter. In this setting, each object O_i has its own constant $\alpha_i \in [-1, 1]$, so that we can favor the segmentation of O_i with transitions from *bright to dark* pixels with $\alpha_i > 0$, or

the opposite orientation, with $\alpha_i < 0$. Note that $\alpha_i = 0$ can be used to indicate that O_i has no boundary polarity prior. In general, we have $\omega_i(\langle s,t\rangle) \neq \omega_i(\langle t,s\rangle)$ for $\alpha_i \neq 0$.

In the second step, HLOIFT generates a *hierarchical layered* weighted digraph $\mathcal{H}$ as the union of all layered graphs $\mathcal{H}_i$, $i = 1,\ldots,m$, with additional *inter-layer* arcs connecting only some of the distinct layers. Here, we consider the inclusion relation only. That is, if O_j is the parent of O_i (i.e., $h(i) = j$), then we define $\omega(t,s) = \infty$ and $\omega(s,t) = -\infty$, for all $s = (i,v) \in \mathcal{N}_i$ and $t = (j,u) \in \mathcal{N}_j$ such that u and v are neighboring superpixels or $u = v$. The reason for also using the neighboring superpixels is to guarantee that O_j will be bigger than O_i, in order to prevent $O_j = O_i$.

Finally, in the last step, a modified OIFT algorithm is applied over $\mathcal{H}$ to compute the segmentation map $X\colon \mathcal{N} \to \{0,1\}$. However, in the case where only the inclusion relation is considered, as done in this work, this algorithm becomes the regular OIFT (Section III-A) over the graph $\mathcal{H}$.

In HLOIFT, given a set of m objects satisfying the seed and inclusion constraints (that is, a valid solution), the energy of the object O_i in the $\mathcal{H}_i$ layer is given by:

$$e(O_i) = \min_{\langle s,t\rangle \in \mathcal{A}_i} \{\omega_i(\langle s,t\rangle) : p(s) \in O_i \;\&\; p(t) \notin O_i\}. \quad (5)$$

The final energy of the set of m objects is given by:

$$e(\langle O_1,\ldots,O_m\rangle) = \min_{i\in\mathcal{L}} \; e(O_i). \quad (6)$$

Note that for a valid solution, we have $e(\langle O_1,\ldots,O_m\rangle) = \varepsilon_{min}(X)$, as defined by Equation 2.

IV. Unsupervised HLOIFT

In order to perform an unsupervised segmentation of optimum energy on the layered graph of an image, which exploits the inclusion relation of HLOIFT, we need a way to automatically find the seeds. If we assume that the object of interest is fully included in the image domain, we can sample a superpixel at the border of the image and take it as a background seed (e.g., the first top/left superpixel in the image, as used in [20]).

Therefore, the problem boils down to finding an appropriate object seed for O_1, because in HLOIFT only the seeds of the innermost objects and background are actually required. Note that, as we are interested in the automatic selection of a set of objects resulting from a map X of maximum energy ε_{min} (Equation 2) in the graph $\mathcal{H}$, the specification of additional unnecessary seeds would only increase the number of constraints in the optimization problem, consequently reducing the energy obtained, therefore not being a valid option.

According to Lemma 1 from [27], it is known that for a given configuration of background seeds, the energy resulting from a segmentation for each possible individual object seed can be calculated by IFT [26] with the connectivity function $f_{\max}$ (Equation 1), that gives the highest arc weight value along the path, but considering its calculation in the transpose graph $\mathcal{H}^T$ and only from the background seeds.

However, as pointed out in [20], this approach has the drawback of indicating several nodes as having the same energy within each object, which actually correspond to equivalent seeds that are redundant, among other problems that make its use unfeasible for ranking the nodes to select the $k-1$ best ones, aiming at a partition of the image into k regions.

In [20], a solution is proposed for ranking the seeds based on a connectivity function that gives the last weight in the path as its cost in the transpose graph, in order to identify for each object a single node with energy given by the weight of its weakest outgoing arc. However, this solution given in [20] is not valid in the case of the layered graph $\mathcal{H}$ that is now being used, since the energy values that will be used for the ranking of seeds of O_1 correspond to the observed path cost values in Layer 1, that is, the costs of the vertices of $\mathcal{H}_1$. However, the energy must now reflect the weight of the weakest outgoing arc of the object boundaries of the various layers, and therefore a cost function of the type $f_{\max}$ is necessary so that these weights obtained in the upper layers are properly propagated to the first layer.

In order to solve the aforementioned problem, but maintaining the feasibility for the ranking of candidate seeds, we propose the usage of the following connectivity function:

$$f_\varepsilon(\langle t\rangle) = \begin{cases} -1 & \text{if } t = (i,v) \text{ and } v \in \mathcal{S}_0 \\ +\infty & \text{otherwise} \end{cases}$$

$$f_\varepsilon(\pi_s \cdot \langle s,t\rangle) = \begin{cases} \max\{f_\varepsilon(\pi_s), \omega(\langle s,t\rangle)\} & \text{if } s = (i,v) \text{ and } i > 1 \\ \omega(\langle s,t\rangle) & \text{otherwise} \end{cases} \quad (7)$$

Figure 1 presents an example of the calculation of function f_ε in the transpose graph $\mathcal{H}^T$ of a synthetic image. The arc weights used were computed by Equation 4, but changing $\|I(t) - I(s)\|$ to $|l_t - l_s|$, since it is a grayscale image. The example shows two pairs of objects with nested boundaries $\langle O_1, O_2\rangle$ and $\langle O_1', O_2'\rangle$, which are shown in Figure 1b, in yellow and pink colors, respectively, being O_1 and O_1' defined in the first layer, while O_2 and O_2' are defined in the second layer. The first pair $\langle O_1, O_2\rangle$ (on the left) has energy $e(\langle O_1, O_2\rangle) = \min\{6,9\} = 6$ (Equation 6), while the second pair $\langle O_1', O_2'\rangle$ (on the right) has a worse energy of $e(\langle O_1', O_2'\rangle) = \min\{12,3\} = 3$.

Note that the cost referring to the arc weight of value 3 of O_2' is successfully transferred from the second layer to the first layer by the maximum weight function used by f_ε (Equation 7), when the arc between layers is processed. However, after its arrival in the first layer, function f_ε starts to behave as a function of last weight, so that only a single node of O_1' receives the cost of 3. In the end, in layer 1, we end up with a single node of $\mathcal{H}_1$ having the maximum energy 6 and the second largest node of $\mathcal{H}_1$ having energy 3, so it is possible to rank the seeds in $\mathcal{H}_1$ for the selection of the best pair of nested boundaries (that is, the pair $\langle O_1, O_2\rangle$).

Following this seed ranking scheme by function f_ε, we can create a hierarchy of partitions according to the following proposed algorithm:

Algorithm 1. – UNSUPERVISED HLOIFT ALGORITHM (UHLOIFT)

INPUT: Hierarchical layered digraph $\mathcal{H} = \langle \mathcal{N}, \mathcal{A}, \omega \rangle$, a background reference superpixel r and the desired number of regions k.
OUTPUT: Graph partition of the first layer $\mathcal{H}_1$ into k regions.

1. *Compute the cost map $V^* : \mathcal{N} \to \mathbb{R}$ by IFT with f_ε (Eq. 7) and $\mathcal{S}_0 = \{r\}$ in the transpose graph $\mathcal{H}^T$.*
2. *Sort the nodes of $\mathcal{N}_1 \setminus \{(1, r)\}$ in a non-increasing order of costs in V^*, getting $\{t_1, t_2, \ldots, t_n\}$, such that $V^*(t_i) \geq V^*(t_{i+1})$, $i = 1, \ldots, n-1$, where $n = |\mathcal{N}_1| - 1$.*
3. *Set $\mathcal{S}_0 \leftarrow \{(1, r), \ldots, (m, r)\}$.*
4. ***For each** t_i, $i = 1, \ldots, k-1$, **do***
5. * Set $\mathcal{S}_1 \leftarrow \{t_i\}$.*
6. * Compute the labeling function X by OIFT with seed sets $\mathcal{S}_0$ e $\mathcal{S}_1$ in $\mathcal{H}$.*
7. * Set $\mathcal{C} \leftarrow \{\langle s, t \rangle \in \mathcal{A}_1 : X(s) \neq X(t)\}$.*
8. * Remove all arcs in $\mathcal{C}$ from $\mathcal{A}_1$ (i.e., $\mathcal{A}_1 \leftarrow \mathcal{A}_1 \setminus \mathcal{C}$).*
9. * Insert t_i in $\mathcal{S}_0$ (i.e., $\mathcal{S}_0 \leftarrow \mathcal{S}_0 \cup \{t_i\}$).*
10. *Returns the partition of layer $\mathcal{H}_1$ into k regions, by labeling its connected components.*

Algorithm 1 generates a hierarchical segmentation by successive binary divisions of layer $\mathcal{H}_1$, leading at the end to a segmentation with k partitions.

Although Figure 1 presents an example with $m = 2$, Algorithm 1 can be executed for an arbitrary number $m \geq 1$ of nested objects. Even though we are showing in the proposed output (Line 10) only the partition into k regions referring to the object in the first layer, it is important to note that the execution of OIFT in Line 6 generates the complete segmentation of the m objects in all layers.

The sorting of Line 2 can be done in the worst case in $O(n \log n)$ complexity, where $n = |\mathcal{N}_1| - 1$, but in practice as we only need the first $k - 1$ values (Line 4), when $k \ll |\mathcal{N}_1|$ then even faster solutions are possible. The IFT of Line 1 and the OIFT of Line 6 can both be calculated in $O(n \log n)$, with $n = |\mathcal{N}|$. Therefore, the complexity of Algorithm 1 is $O(kn \log n)$, with $n = |\mathcal{N}|$. In practice, however, the algorithm is quite efficient, due to the usage of a graph of superpixels, which considerably reduces the cardinality of the set $\mathcal{N}$.

V. EXPERIMENTAL RESULTS

In our experiments, we only consider methods that have some level of direct control over the number of generated connected regions, since our objective is to measure which methods can generate the objects of interest with the fewest connected regions in the image partition, by monitoring their accuracy for increasing values of k.

We conducted experiments, comparing UHLOIFT with other graph-based methods. In the following, MST denotes the clustering of RAG nodes, obtained by successive removals of edges of maximum weight from the minimum spanning tree, where $\omega(\langle s, t \rangle) = \|I(t) - I(s)\|$, which is related to the top-down version of nearest-neighbor (single-linkage) algorithm. MST has linearithmic complexity $O(n \log n)$ in the number of RAG nodes n. UOIFT denotes the Unsupervised OIFT from [20] computed over a digraph of superpixels, which allows the usage of the boundary polarity of objects through Equation 4. HFH denotes the method obtained by Hierarchizing Felzenszwalb-Huttenlocher segmentation from [28] as proposed by [29], with an area-filtering post-processing to eliminate small components with the superpixel size. EF+WS indicates the IFT-based watershed transform, after a volume extinction filter [30] set to preserve k leaves of the Min-tree, in order to consider only the most relevant catchment basins of a morphological gradient by a disk of radius 1. We used the code for the extinction filter available in the iamxt toolbox [31]. Its Min-tree can be computed in $O(N \times h + M)$, where N is the number of pixels, h the number of levels of the image and $M = \kappa \times N$, being κ the number of neighbors of each pixel [30]. We also included SICLE to segment the image into k regions, denoting the recent method to compute superpixels from [32], using its default configuration, without saliency estimation, from the code available online[1].

We performed quantitative experiments for colored images of 640×480 pixels for three different databases with 15, 70 and 61 images, respectively, which are available to the community[2]. Sample images of the three bases are shown in Figure 2. We computed the mean accuracy curves by Dice Coefficient between the ground truth and the best union of segmented regions leading to the object for all the methods and different values of k (Figure 3). We considered superpixels of size $10 \times 10 = 100$ pixels, for all RAG-based methods. UHLOIFT is the method that requires fewer partitions in order to get the desired segmented regions, demonstrating the usefulness of the inclusion relation of nested boundaries. For the first base, we use $m = 3$ with $\alpha_1 = \alpha_3 = 0.9$ and $\alpha_2 = -0.9$, while for the second base, we use $m = 3$ with $\alpha_1 = \alpha_3 = -0.9$ and $\alpha_2 = 0.9$, and for the third, we use $m = 2$ with $\alpha_1 = -0.9$ and $\alpha_2 = 0.9$. UOIFT is usually the second best method, with its polarity given by α_1. An example of segmentation is shown in Figure 4.

Regarding the computational time, for a 640×480 image, to compute superpixels of size 10×10 by IFT-SLIC [25] takes 621.8 ms and the final clustering into 5 regions by UHLOIFT with three layers ($m = 3$) in the RAG takes only 16.83 ms, in an Intel Core i5-10210U CPU @ 1.60GHz×8.

VI. CONCLUSION

In this work, the HLOIFT method with inclusion relations was successfully extended to the unsupervised paradigm. As attested in the experiments, with the simple specification of the expected high-level constraints of the desired nested objects, the method can be quickly employed in new applications without the need to have a previously built training base.

As future work, we intend to compute UHLOIFT even more efficiently, by using the differential OIFT algorithm [18]. We also intend to extend the proposed method to include the exclusion relation between sibling objects, and to investigate

[1] https://github.com/LIDS-UNICAMP/SICLE
[2] http://www.vision.ime.usp.br/~pmiranda/downloads.html

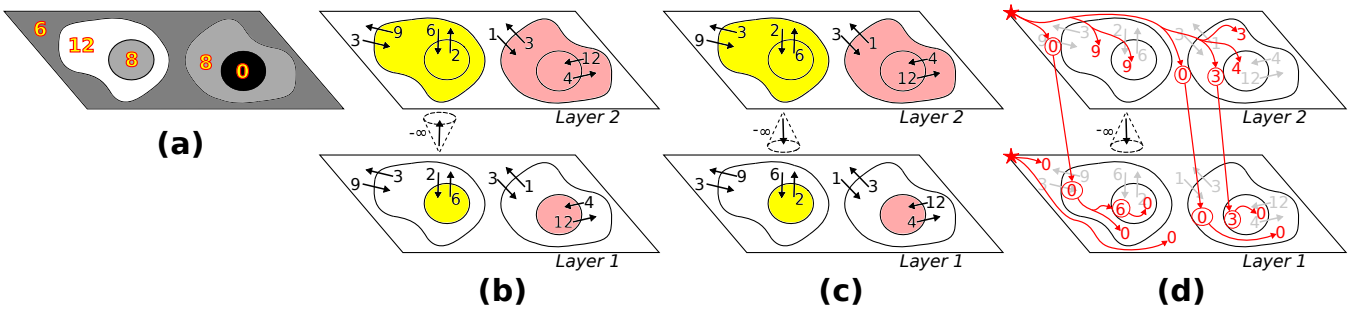


Fig. 1. UHLOIFT computing example: (a) Synthetic image with intensity values indicated within each homogeneous region. (b) The hierarchical layered digraph $\mathcal{H}$ with two layers using $\alpha_1 = -0.5$ and $\alpha_2 = 0.5$. (c) The transpose graph $\mathcal{H}^T$. (d) The computed energy map for seed ranking in $\mathcal{N}_1$ using f_ε.

Fig. 2. Sample images with 640×480 pixels. (a-b) Stop signs written in Portuguese to segment the four letters of the word "PARE". (c-d) QR codes to segment the center squares of the three position markers. (e-f) State flags of São Paulo to segment the geographic silhouette of Brazil inside the white circle in their upper left corner.

other ways of combining the energies from the various layers, in addition to the solution already proposed by Equation 6.

Acknowledgment

Thanks to Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – (Grant 407242/2021-0, 313087/2021-0, 166631/2018-3), CAPES (88887.136422/2017-00), FAPESP (2014/12236-1, 2014/50937-1) and IPT (Institute for Technological Research).

References

[1] R. Zeng, Y. Wen, W. Zhao, and Y.-J. Liu, "View planning in robot active vision: A survey of systems, algorithms, and applications," *Computational Visual Media*, vol. 6, pp. 225–245, 2020.

[2] Y. D. Yasuda, F. A. Cappabianco, L. E. G. Martins, and J. A. Gripp, "Automated visual inspection of aircraft exterior using deep learning," in *Anais Estendidos do XXXIV Conference on Graphics, Patterns and Images*. SBC, 2021, pp. 173–176.

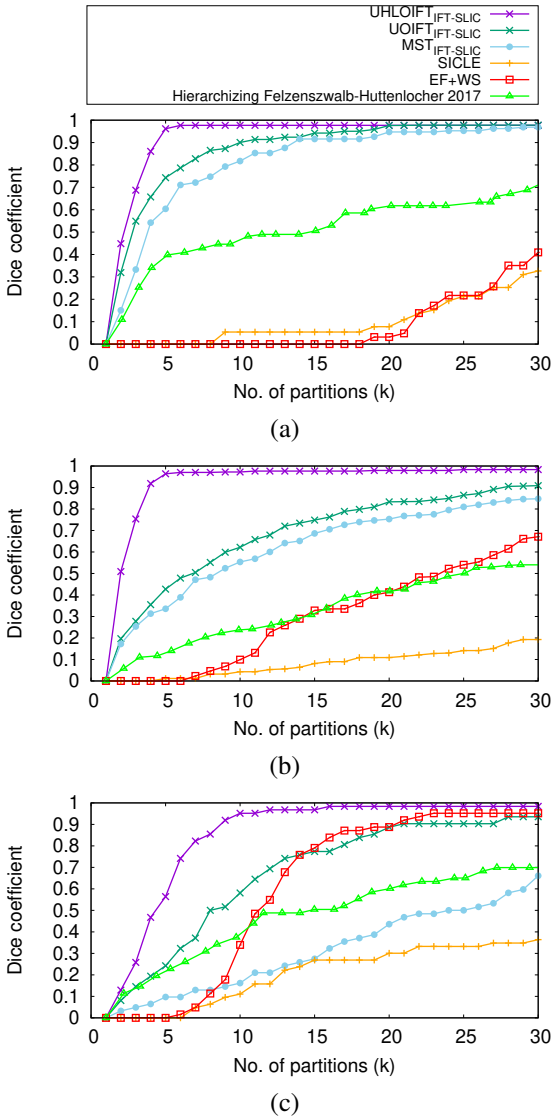
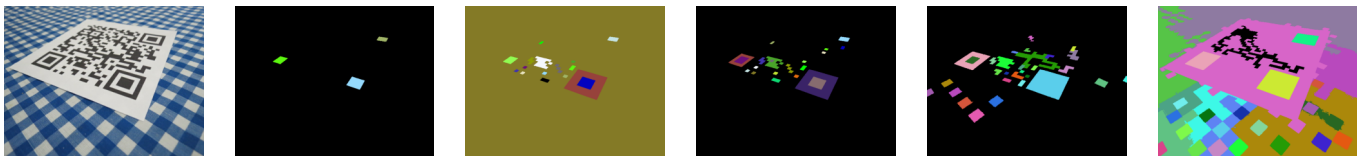


Fig. 3. The mean accuracy curves by Dice of the best union of produced connected regions for different values of k and methods, to segment: (a) The four letters of the stop sign written in Portuguese (Figures 2a-b). (b) The center squares of the three position markers of a QR code (Figures 2c-d). (c) The geographic silhouette of Brazil inside the white circle in the upper left corner of the state flag of São Paulo (Figures 2e-f).



(a) Input image (b) UHLOIFT$_{k=4}$ (c) UOIFT$_{k=17}$ (d) MST$_{k=22}$ (e) HFH$_{k=30}$ (f) SICLE$_{k=30}$

Fig. 4. (a) QR code image with 640×480 pixels to segment the central square of the three position markers. (b) The proposed result by UHLOIFT can solve the problem with $k = 4$. (c) UOIFT [20] with polarity favoring transitions from dark to bright pixels requires $k = 17$. (d) The segmentation by a single-linkage algorithm using the MST of the RAG requires $k = 22$. (e-f) The results for $k = 30$ by HFH and SICLE, respectively, fail to segment all three central black squares.

[3] E. Nigri, N. Ziviani, F. Cappabianco, A. Antunes, and A. Veloso, "Explainable deep cnns for mri-based diagnosis of alzheimer's disease," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[4] T. Magadza and S. Viriri, "Deep learning for brain tumor segmentation: a survey of state-of-the-art," *Journal of Imaging*, vol. 7, no. 2, p. 19, 2021.

[5] C.-L. Phuah, Y. Chen, J. F. Strain, N. Yechoor, O. J. Laurido-Soto, B. M. Ances *et al.*, "Association of data-driven white matter hyperintensity spatial signatures with distinct cerebral small vessel disease etiologies," *Neurology*, vol. 99, no. 23, pp. e2535–e2547, 2022.

[6] W. Cai, Z. Wei, Y. Song, M. Li, and X. Yang, "Residual-capsule networks with threshold convolution for segmentation of wheat plantation rows in uav images," *Multimedia Tools and Applications*, vol. 80, pp. 32 131–32 147, 2021.

[7] X. Jin, J. Che, and Y. Chen, "Weed identification using deep learning and image processing in vegetable plantation," *IEEE Access*, vol. 9, pp. 10 940–10 950, 2021.

[8] T. L. Barreto, R. A. Rosa, C. Wimmer, J. R. Moreira, L. S. Bins, F. A. M. Cappabianco, and J. Almeida, "Classification of detected changes from multitemporal high-res xband sar images: intensity and texture descriptors from superpixels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5436–5448, 2016.

[9] X. Yu, X. Ye, and Q. Gao, "Pipeline image segmentation algorithm and heat loss calculation based on gene-regulated apoptosis mechanism," *International Journal of Pressure Vessels and Piping*, vol. 172, pp. 329–336, 2019.

[10] A. Sampath, P. Bijapur, A. Karanam, V. Umadevi, and M. Parathodiyil, "Estimation of rooftop solar energy generation using satellite image segmentation," in *2019 IEEE 9th International Conference on Advanced Computing (IACC)*. IEEE, 2019, pp. 38–44.

[11] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020.

[12] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *International journal of multimedia information retrieval*, vol. 9, no. 3, pp. 171–189, 2020.

[13] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *Computational visual media*, vol. 5, pp. 117–150, 2019.

[14] K. Raza and N. K. Singh, "A tour of unsupervised deep learning for medical image analysis," *Current Medical Imaging*, vol. 17, no. 9, pp. 1059–1077, 2021.

[15] V. Rani, S. T. Nabi, M. Kumar, A. Mittal, and K. Kumar, "Self-supervised learning: A succinct review," *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2761–2775, 2023.

[16] L. L. Vercio, K. Amador, J. J. Bannister, S. Crites, A. Gutierrez, M. E. MacDonald, J. Moore, P. Mouches, D. Rajashekar, S. Schimert *et al.*, "Supervised machine learning tools: a tutorial for clinicians," *Journal of Neural Engineering*, vol. 17, no. 6, p. 062001, 2020.

[17] M. Zhang, Y. Zhou, J. Zhao, Y. Man, B. Liu, and R. Yao, "A survey of semi-and weakly supervised semantic segmentation of images," *Artificial Intelligence Review*, vol. 53, pp. 4259–4288, 2020.

[18] M. A. T. Condori and P. A. V. Miranda, "Differential oriented image foresting transform segmentation by seed competition," in *Discrete Geometry and Mathematical Morphology*, É. Baudrier, B. Naegel, A. Krähenbühl, and M. Tajine, Eds. Cham: Springer International Publishing, 2022, pp. 300–311.

[19] P. Miranda and L. Mansilla, "Oriented image foresting transform segmentation by seed competition," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 389–398, Jan 2014.

[20] H. H. Bejar, S. J. Ferzoli Guimaraes, and P. A. Miranda, "Efficient hierarchical graph partitioning for image segmentation by optimum oriented cuts," *Pattern Recognition Letters*, vol. 131, pp. 185–192, 2020.

[21] L. M. Leon, K. C. Ciesielski, and P. A. Miranda, "Efficient hierarchical multi-object segmentation in layered graphs," *Mathematical Morphology - Theory and Applications*, vol. 5, no. 1, pp. 21–42, 2021.

[22] C. de Moraes Braz, P. A. Miranda, K. C. Ciesielski, and F. A. Cappabianco, "Optimum cuts in graphs by general fuzzy connectedness with local band constraints," *Journal of Mathematical Imaging and Vision*, vol. 62, pp. 659–672, 2020.

[23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[24] P.-H. Conze, G. Andrade-Miranda, V. K. Singh, V. Jaouen, and D. Visvikis, "Current and emerging trends in medical image segmentation with deep learning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2023.

[25] E. B. Alexandre, A. S. Chowdhury, A. X. Falcão, and P. A. V. Miranda, "IFT-SLIC: A general framework for superpixel generation based on simple linear iterative clustering and image foresting transform," in *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug 2015, pp. 337–344.

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

[27] H. H. Bejar and P. A. Miranda, "Oriented relative fuzzy connectedness: Theory, algorithms, and its applications in hybrid image segmentation methods," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 21, Jul 2015.

[28] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep 2004.

[29] S. Guimarães, Y. Kenmochi, J. Cousty, Z. P. Jr., and L. Najman, "Hierarchizing graph-based image segmentation algorithms relying on region dissimilarity," *Mathematical Morphology - Theory and Applications*, vol. 2, no. 1, pp. 55–75, 2017.

[30] A. G. Silva and R. d. A. Lotufo, "Efficient computation of new extinction values from extended component tree," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 79–90, Jan. 2011.

[31] R. Souza, L. Rittner, R. Machado, and R. Lotufo, "iamxt: Max-tree toolbox for image processing and analysis," *SoftwareX*, vol. 6, pp. 81 – 84, 2017.

[32] F. Belém, I. Borlido, L. João, B. Perret, J. Cousty, S. J. F. Guimarães, and A. Falcão, "Fast and effective superpixel segmentation using accurate saliency estimation," in *Discrete Geometry and Mathematical Morphology*, É. Baudrier, B. Naegel, A. Krähenbühl, and M. Tajine, Eds. Cham: Springer International Publishing, 2022, pp. 261–273.