Probabilistic Image Segmentation with Closedness Constraints

Bjoern Andres, Jörg H. Kappes, Thorsten Beier,
Ullrich Köthe and Fred A. Hamprecht

HCI, University of Heidelberg,
Speyerer Str. 6, 69115 Heidelberg, Germany

Abstract

We propose a novel graphical model for probabilistic image segmentation that contributes both to aspects of perceptual grouping in connection with image segmentation, and to globally optimal inference with higher-order graphical models. We represent image partitions in terms of cellular complexes in order to make the duality between connected regions and their contours explicit. This allows us to formulate a graphical model with higher-order factors that represent the requirement that all contours must be closed. The model induces a probability measure on the space of all partitions, concentrated on perceptually meaningful segmentations. We give a complete polyhedral characterization of the resulting global inference problem in terms of the multicut polytope and efficiently compute global optima by a cutting plane method. Competitive results for the Berkeley segmentation benchmark confirm the consistency of our approach.

1 Introduction

We study the image partitioning problem, where the task is to decompose an image into a previously unknown number of segments that are somehow homogeneous but do *not* belong to a predefined set of categories such as {ground, car, sky}. The most popular representation for this kind of problem is in terms of pixel labels: segments are then defined as connected components of pixels with the same label. However, while a labeling uniquely defines a segmentation, the converse is not true. This aggravates the inference problem, as discussed in Section 3.

Alternatively, one may represent image partitioning as an edge labeling problem. Here, a region adjacency graph of pixels or superpixels can be constructed, and each edge in the graph can be labeled as "active" (1) or "dormant" (0). Each maximal set of nodes that is connected by edges of type 0 corresponds to one segment. An important advantage is that arbitrary partitionings can thus be represented using only binary labels. On the downside, not every binary labeling results in closed contours. Such inconsistencies could be addressed by a heuristic postprocessing that closes gaps in
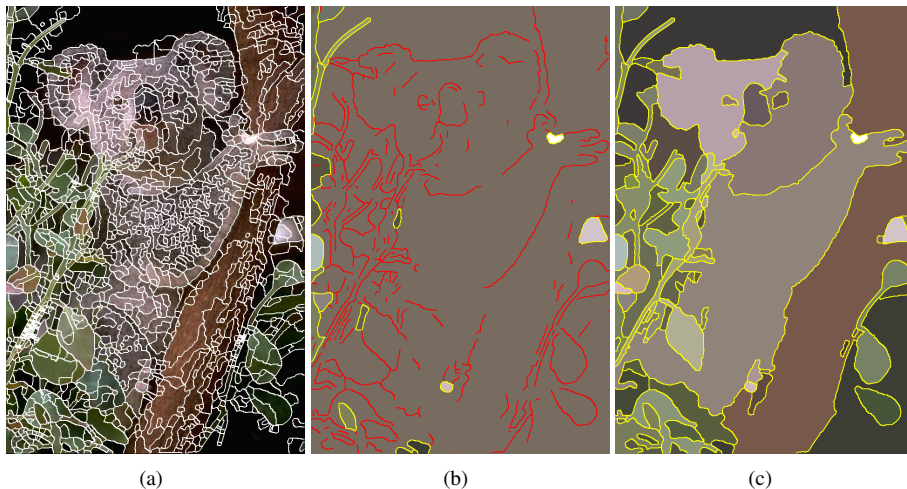
| (a) | (b) | (c) |

Figure 1: (**a**) Oversegmentation of an image. The curves that separate superpixels are shown in white. (**b**) Given local boundary and Gestalt features that characterize each curve, we make a prediction if a curve should be kept ("active") or discarded ("dormant"). Connected component analysis then yields a segmentation that is, however, of poor quality because the contours of most real objects have gaps. Each gap arises from a curve that was falsely discarded. Many "active" curves (shown in red) then end up in the interior of connected components and are hence inconsistent with the segmentation. (**c**) The introduction of topological constraints yields improved and consistent closed-contour segmentations.

contours, or eliminates dangling boundaries within a segment. While empirically useful, such methods are hard to characterize theoretically and hence difficult to improve systematically.

To address these challenges, we start from an oversegmentation whose regions become the nodes of an adjacency graph. A binary random variable is associated with each edge, stating whether the corresponding curve should become active as part of a segment boundary, or remain dormant. A probabilistic graphical model is proposed that associates a probability with each realization of these random variables. Importantly, we formulate a prior that assigns zero probability to all those configurations that correspond to inconsistent edge labelings. This corresponds to an exponential number of constraints in the integer linear programming (ILP) problem to which the inference problem can be cast. However, we can find violated constraints in polynomial time (Section 3.2) and add these iteratively, thus solving the inference problem to global optimality using branch-and-cut. Overall, this amounts to a practical solution of the multicut problem, despite its NP hardness.

Summarizing, our main contributions are

- A statistically sound formulation of the partitioning problem that explicitly includes a closedness constraint and achieves state-of-the-art performance on the

2

Berkeley segmentation dataset (BSD) [15].

- An explicit objective function that measures the fit of a partitioning to an image, rather than a mere procedural recipe.

- Empirical proof that globally optimal (maximum a posteriori) solutions can be found even in the face of non-local closed-contour constraints.

- A probabilistic model whose underlying statistical assumptions are made explicit and which can be parameterized without manual tweaking based on statistical learning from training data.

2 Related Work

Image segmentation has successfully been formulated as an optimal graph partitioning problem, e.g. in normalized cuts [25]. While the normalized cut framework settles for an iterative bisectioning of the graph by solving a relaxation of the normalized cut objective, we solve an unrelaxed multicut objective to global optimality. We formulate this objective in terms of a higher-order probabilistic graphical model over edge labelings.

Graphical models over edge labelings have been proposed in [17, 28]. These models are restricted to the special case in which curves cannot intersect. Each segment is thus adjacent to at most one other segment in a global optimum. If it is known a priori that only two segments exist, this assumption is mild and the optimization problem can be solved efficiently [17]. We drop this assumption for the general case in which the number of segments is not known a priori. We still guarantee the closedness of curves globally through higher-order constraints and find global optima.

In [28], the "multi class segmentation" problem is addressed in which object categories are given and the goal is to assign one category to each pixel. We, in contrast, address the multicut problem where the goal is to partition the image, based merely on a notion of similarity. Model assumptions are violated in the solutions of [28] (Fig. 14) found by Loopy Belief Propagation which indicates that the problem has not been solved to global optimality.

The advantages of superpixels and features derived from these have been expounded by [23] and others.

On the theoretical side, our model is strongly related to the partitioning or multicut problem in combinatorial optimization, cf. [6, 9, 7]. While the calculation of minimal cuts has become a standard technique in computer vision [12, 24], its generalization to the multi-class case (multicuts) has so far been deemed to be impractical for computer vision applications. However, the multicut problem can be formulated as a linear program over the multicut polytope [11]. Unfortunately, the number of facets defining this polytope is exponential in general. Therefore, cutting plane methods are used to iteratively tighten an outer approximation of the multicut polytope [10, 3].

Pioneering applications in computer vision that exploit these techniques in the primal (node) domain are [26] and [18]. In contrast to these, we consider the dual (edge)
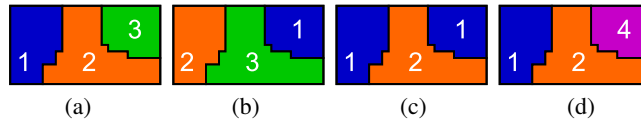
Figure 2: Unequal node labelings that represent the same segmentation (a through d) can be obtained by permuting labels (b) and by using one or several labels for more than one segment (c). The number of choices increases if there are more labels than necessary to represent all segmentations (d).

domain and apply branch-and-cut algorithms [5] in order to guarantee optimal integer solutions.

3 Probabilistic Models of Graph Segmentation

We represent an image as a graph $G = (V, E)$ whose nodes correspond to pixels or image regions / superpixels. We set out to find a probabilistic model on the set of all possible partitionings $\mathcal{S}_G$ of that finite graph G into connected subsets of nodes. That is, we want to find a probability mass function p that assigns a probability $p(S) \in [0, 1]$ to every possible segmentation $S \in \mathcal{S}_G$. For all but the smallest graphs, the set $\mathcal{S}_G$ is too big to explore exhaustively, requiring an implicit definition. While such a definition in terms of node labelings is challenging (Section 3.1) we use binary edge labelings together with additional constraints that guarantee closed contours.

3.1 Graph Segmentation by Node Labeling

One way to define the desired probability mass function implicitly is to define a graphical model in terms of a node labeling. Assuming a set L of labels, each of the $|V|$ discrete variables in the graphical model can take any of $|L|$ states. The segmentation induced by a node labeling is then defined as the partition of V into maximal connected subsets of nodes that have the same label.

The set of all node labelings is greater than the set of all segmentations represented by these labelings (see Fig. 2). This implies that the optimizer has to work in a search space that has degenerate optima [11], and that certainly is (much) larger than is theoretically necessary. Two approaches to tackle this problem are by introduction of a label bias [8] and through a Dirichlet process prior [19]. While the first still has to operate on a huge state space, the latter requires sampling methods for optimization. We present a third way via the dual problem in which edges are labeled.

3.2 Graph Segmentation by Edge Labeling

Indeed, we represent any segmentation of a finite graph (V, E) by an *edge labeling* $y \in \{0, 1\}^{|E|}$ that indicates for each edge $e \in E$ whether its incident nodes belong to the same segment (curve dormant, $y_e = 0$) or not (curve active, $y_e = 1$). A probability mass function $p : \{0, 1\}^{|E|} \to [0, 1]$ on the set of all possible edge labelings can now
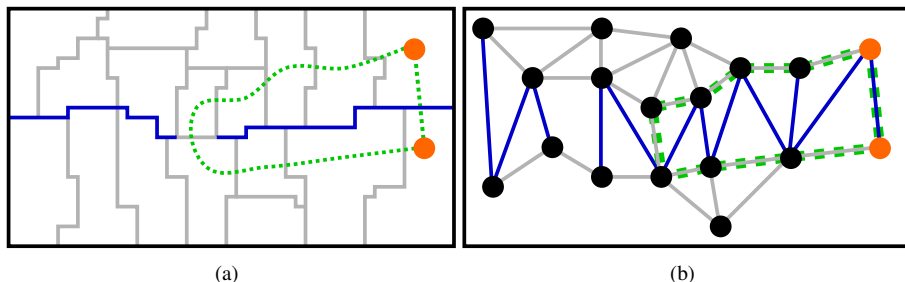
4

Figure 3: A set of superpixels (**a**) and its corresponding adjacency graph (**b**). The current configuration of curves / edges labeled as active (blue) or dormant (gray) is inconsistent: the closed path depicted in green does not meet the requirements from Definition 1 or Lemma 1. The section between the orange dots indicates that the respective superpixels should belong to different segments; while the remainder of the path claims the contrary.

be defined in terms of a graphical model with $|E|$ binary variables. This representation has received little attention so far, with the notable exception of [17, 28].

The state space of all $2^{|E|}$ possible edge labelings is still too large because not every edge labeling is consistent. The notion of consistency is clarified in Fig. 3 and the following

Definition 1: Given a finite graph $G = (V, E)$, an edge labeling $y \in \{0, 1\}^{|E|}$ is termed *consistent* if, for all closed paths $(v_1, \ldots, v_n = v_1)$ in G, either none or more than one edge is labeled as active (1), i.e. $\sum_{j=1}^{n-1} y_{\{v_j, v_{j+1}\}} \neq 1$.

The number of these paths can be exponential. Therefore in practice, to determine which inequalities are violated, if any, it is more convenient to look for paths that short-circuit two nodes that lie on opposite sides of an activated edge. Such short-circuiting paths are the subject of

Lemma 1: An edge labeling $y \in \{0, 1\}^{|E|}$ of a graph $G = (V, E)$ is consistent if and only if, for all edges $\{v, v'\} \in E$ with $y_{\{v, v'\}} = 1$ and all paths $(v_1, \ldots, v_n)$ from $v_1 = v$ to $v_n = v'$: $\sum_{j=1}^{n-1} y_{\{v_j, v_{j+1}\}} > 0$.

See [6] or the appendix for a short proof of the lemma.

Thus, while the number of inequalities can be exponentially large, violated constraints can be found in time $O(|V| + |E|)$ and added iteratively until the solution is feasible. One simple algorithm for this purpose labels the connected components of the graph $G_0 = (V, \{e \in E | y_e = 0\})$ in time $O(|V| + |E|)$ and then tests for each edge $\{v, v'\} \in E$ that is labeled 1 whether v and v' belong to the same component, in time $O(1)$. If this is the case, the edge labeling is inconsistent, and any path in G_0 from v to v' yields a violated inequality, including a shortest path that can be found in time $O(|V| + |E|)$ using breadth-first-search.

This insight is crucial for the implementation of our model which is designed to allow only consistent edge labelings. Each consistent edge labeling relates bijectively
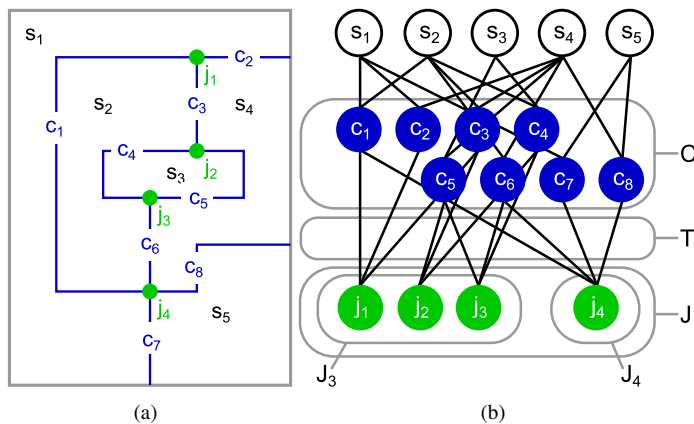
Figure 4: The segmentation of a pixel grid (**a**) partitions the continuous image plane into regions $s_1, \ldots, s_5$, *curves* $c_1, \ldots c_8$ that bound these regions, and junctions (points) $j_1, \ldots, j_4$ that bound these curves. The topology of this segmentation is captured in a cellular complex [13] that relates each region to its bounding curves and each curve to its bounding junctions (**b**).

to a segmentation. This bijection is the subject of the multicut problem in optimization [7]: the set of consistent edge labelings corresponds to the vertices of the multicut polytope [9, 6]. Inconsistent edge labelings lie outside the multicut polytope. In this sense, any heuristic that generates a segmentation from an inconsistent labeling can be seen as some kind of (and typically suboptimal) mapping onto the multicut polytope. As shown in [7], the multicut problem can be formulated as an integer linear program which is usually solved with cutting plane methods [10, 5].

4 A Probabilistic Higher-Order Graphical Model for Image Segmentation

4.1 Representation, Terminology and Notation

So far, we have discussed graph segmentation. Image segmentation is a special case, where the initial graph is an adjacency graph [20] whose nodes are either individual pixels or superpixels (connected subsets of pixels).

The initial segmentation partitions the continuous image plane into (i) regions, (ii) curves that bound regions, and (iii) junctions where several curves meet (Fig. 4).

It is important to note that the topology of these sets can have a richer structure than the region adjacency graph reveals. For example, in Fig. 4, the two curves c_3 and c_6 bound the same regions s_2 and s_4. However, in the adjacency graph, these regions are connected by a single edge. From the viewpoint of topology, c_3 and c_6 are distinct and need to be handled separately. We use a topological grid [4] to represent all regions,

curves and junctions, and a cellular complex [13] to capture their topology, see Fig. 4.

The topology of curves and junctions can be expressed as a bipartite graph (C, J, T), Fig. 4, in which C is the set of curves and J is the set of junctions. A relation $(c, j) \in T$ indicates that the curve $c \in C$ is bounded by the junction $j \in J$. Due to the regularity and discreteness of the nearest-neighbor Cartesian pixel grid, junctions can only bound either three or four curves. For future use in our graphical model, we call these corresponding sets of junctions J_3 and J_4. All junctions that delimit a curve and vice versa are referenced by $\mathcal{N}(c) = \{j \in J | (c, j) \in T\}$ and $\mathcal{N}(j) = \{c \in C | (c, j) \in T\}$.

Summarizing, then, in this representation a segmentation is defined indirectly through a given configuration of curves that are either switched on (active) or off (dormant).

4.2   The Probabilistic Model

We now come to the core of our modeling effort. Based on the arguments from Section 3, we assign a binary random variable to each curve of an initial segmentation which determines if that curve is active or dormant. We further propose a probability mass function $p : \{0, 1\}^{|C|} \rightarrow [0, 1]$ that assigns a probability to every conceivable configuration of active and dormant curves. It assigns zero probability to all inconsistent curve labelings, thus guaranteeing that each admissible solution has a one-to-one correspondence to a closed-contour segmentation. We define p in terms of a graphical model as a probability mass function conditioned on local features of junctions and curves, as well as on the topology of the segmentation.

Qualitatively, we learn unary potentials that look to the underlying image for evidence of "boundariness". If there is strong local evidence for a boundary, these potentials encourage a curve becoming active. The third and fourth order junction potentials allow to express Gestalt laws such as good continuation [23, 28]. The combination of all potentials in a single model trades off the potentially conflicting local beliefs encouraged by the different potentials.

The potentials depend on features $f_c^{(1)}$ of curves $c \in C$ and features $f_j^{(n)}$ of junctions $j \in J_{n=3,4}$. The curve features are standard descriptors of the color distribution and filter responses across curves and adjacent regions (cf. [23] and supplementary material). The junction features are the angles between incident curves (cf. supplementary material). The collection of all features extracted from an image is abbreviated as $F := (f^{(1)}, f^{(3)}, f^{(4)})$.

We introduce random variables over the states y of the model, over the features F and over the topology T^1 and denote these by $\mathcal{Y}$, $\mathcal{F}$ and $\mathcal{T}$, respectively.

We now make a series of conditional independence assumptions that are all detailed in the appendix. While having to make such assumptions is always undesirable, at least being able to state them explicitly is to be preferred. The first assumption is that the features and topology are statistically independent, $\mathcal{F} \perp\!\!\!\perp \mathcal{T}$, thus making for the

---
[1]Recall that the topology of junctions and curves is described by the bipartite graph (C, J, T).

factorization

$$p(y|F,T) = \frac{p(y,F|T)}{p(F|T)} = \frac{p(F|y,T)p(y|T)}{p(F|T)} \tag{1}$$

$$= \frac{p(F|y)p(y|T)}{p(F)} = \frac{p(F|y)}{p(F)} \frac{p(T|y)p(y)}{p(T)}$$

$$\propto p(F|y)p(T|y)p(y) \ . \tag{2}$$

In the last line, we have discarded the denominator since we are only interested in that configuration y which has highest probability, and not the probability itself. We now address the modeling of each of the three remaining factors in turn.

The Curve Prior $p(y)$. We assume that the prior for labeling curves as dormant (0) or active (1) is identical for all curves. We here introduce our only design parameter $\beta \in (0,1)$,

$$p(y_c) \overset{!}{=} \begin{cases} 1-\beta & \text{if } y_c = 0 \\ \beta & \text{if } y_c = 1 \end{cases} \tag{3}$$

that states if, without looking at an image, we would prefer to keep curves active (resulting in a fine-grained segmentation) or dormant (resulting in a coarse segmentation). This crucial parameter thus trades off boundary detection precision vs. recall (Fig. 6).

The Likelihood of a Topology $p(T|y)$ given a configuration y is set to nil if y is inconsistent for the topology T, and to a constant otherwise. Learning the true likelihood of topologies for a given configuration from data is certainly interesting but very challenging and beyond the scope of this work. In order to avoid over-fitting, we assume a uniform distribution over all consistent topologies.

The Likelihood of the Features $p(F|y)$. Given the conditional independence assumptions stated in the Appendix, the likelihood $p(F|y)$ factorizes according to

$$p(F|y) = p(f^{(1)}, f^{(3)}, f^{(4)}|y) \tag{4}$$

$$= p(f^{(1)}|y)p(f^{(3)}|y)p(f^{(4)}|y)$$

$$= \prod_{c \in C} p(f_c^{(1)}|y_c) \prod_{d \in \{3,4\}} \prod_{j \in J_d} p(f_j^{(d)}|y_{\mathcal{N}(j)}) \ .$$

We propose to learn the approximate probability $\hat{p}(y_c|f_c^{(1)})$ from class-balanced training data. We then have $\hat{p}(y_c) = 0.5$ and thus

$$p(f_c^{(1)}|y_c) \overset{!}{=} \hat{p}(f_c^{(1)}|y_c) \propto \hat{p}(y_c|f_c^{(1)}) \, \hat{p}(f_c^{(1)}) \ . \tag{5}$$

For each junction $j \in J_3$ of three curves $\{c_1, c_2, c_3\} \in \mathcal{N}(j)$, we also learn the likelihood $\hat{p}(y_{\mathcal{N}(j)}|f_j^{(3)}) = \hat{p}(y_{c_1}, y_{c_2}, y_{c_3}|f_j^{(3)})$ from training data[2] under the assumption that $\hat{p}(y_{\mathcal{N}(j)})$ is constant. Plugging in this estimate, we have

$$p(f_j^{(3)}|y_{\mathcal{N}(j)}) \overset{!}{\propto} \hat{p}(y_{\mathcal{N}(j)}|f_j^{(3)}) \, \hat{p}(f_j^{(3)}) \ . \tag{6}$$

---

[2] Among the $|\{0,1\}^3|=8$ possibilities of removing or preserving the three curves incident to a junction, the three assignments $(0,0,1)$, $(0,1,0)$ and $(1,0,0)$ are inconsistent and thus not represented by any samples in the training data. We assume a uniform distribution of these three assignments to obtain a likelihood term that is agnostic with regard to topology.

Junctions $j \in J_4$ with four incident curves are rare in practice, and a reliable estimate of the likelihood of all possible assignments is hard to obtain from limited training data. In order to avoid over-fitting, we assume a uniform distribution $p(f_j^{(4)}|y_{\mathcal{N}(j)})$.

The Full Model: A Conditional Random Field. In summary, then, the proposed probabilistic model for segmentation is the Conditional Random Field

$$p(y|F,T) \propto \prod_{c \in C} \hat{p}(y_c|f_c^{(1)}) \prod_{j \in J_3} \hat{p}(y_{\mathcal{N}(j)}|f_j^{(3)}) \, p(T|y) \, p(y).$$

The optimization problem of finding a segmentation with maximum posterior probability

$$\underset{y \in \{0,1\}^{|C|}}{\arg\max} \underbrace{\prod_{c \in C} \hat{p}(y_c|f_c^{(1)}) \, p(y)}_{\exp\left(-g^{(1)}(y)\right)} \underbrace{\prod_{j \in J_3} \hat{p}(y_{\mathcal{N}(j)}|f_j^{(3)})}_{\exp\left(-g^{(3)}(y)\right)} \cdot \underbrace{p(T|y)}_{\exp\left(-g^{(T)}(y)\right)}$$

is equivalent to the high-dimensional binary minimization problem

$$\underset{y \in \{0,1\}^{|C|}}{\arg\min} \; g^{(1)}(y) + g^{(3)}(y) + g^{(T)}(y) \tag{7}$$

The term $g^{(1)}(y)$ includes local evidence whether a curve should be removed or preserved, the Gestalt term $g^{(3)}(y)$ supports the smoothness of segment contours, and the topological term $g^{(T)}(y)$ enforces their closedness. The effect of these terms is illustrated in Fig. 1. Using only $g^{(1)}(y) + g^{(3)}(y)$ is not optimal because the consistency of the edge labels cannot be guaranteed, Fig. 1(b). The segmentation using the full model is depicted in Fig. 1(c).

5 Optimization

The combination of all three terms can be formulated as an Integer Linear Program[3] in which $g^{(T)}(y)$ is encoded as the set of constraints in eq. (9) whereas $g^{(1)}(y)$ and $g^{(3)}(y)$ are encoded in the weights $w \in \mathbb{R}^{2 \cdot |C| + 8 \cdot |J|}$. The labeling y is represented by an overcomplete indicator vector μ.

$$\min_{\mu} \sum_{c \in C} \sum_{a \in \{0,1\}} w_{c,a} \mu_{c,a} + \sum_{j \in J} \sum_{a \in \{0,1\}^3} w_{j,a} \mu_{j,a} \tag{8}$$

$$\text{s.t. } \mu \in \{0,1\}^{2 \cdot |C| + 8 \cdot |J|}$$

$$\forall c \in C: \; \mu_{c,0} + \mu_{c,1} = 1$$

$$\forall j \in J_3, b \in \{0,1\}, k \in \{1,2,3\}:$$

$$\sum_{a \in \{0,1\}^3, a_k = b} \mu_{j,a} = \mu_{\mathcal{N}(j)_k, b}$$

$$\forall \text{ cycles } (c_1, \ldots, c_n): \sum_{i=2}^{n} \mu_{c_i, 1} \geq \mu_{c_1, 1}. \tag{9}$$

---

[3]This over-complete representation [27] can be simplified; it is used to support readability.
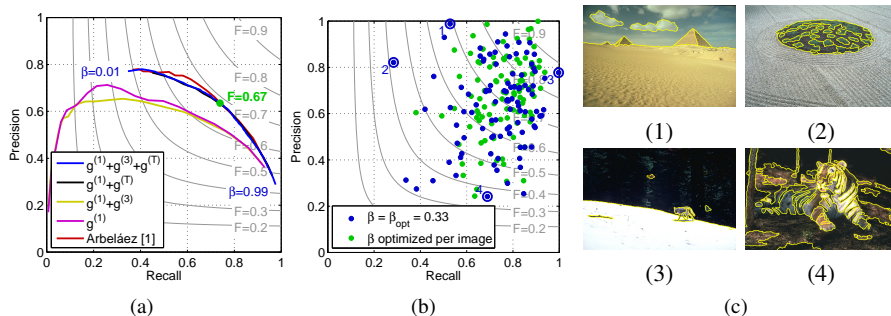
Figure 5: **(a)** Average boundary detection precision and recall (over all 100 images in the BSD test set) of closed contours obtained by [1] (red) and by the proposed method (blue: full model, other colors: simplified models, eq. 7), for different β. **(b)** Quality of segmentations obtained by the proposed method, for each BSD test image, in a fixed-parameter setting (blue), and in a setting where β is optimized for each image (green). **(c)** Those images for which our segmentations have maximum precision (1), minimum recall (2), maximum recall (3) and minimum precision (4).

Without eq. (9), the number of constraints is only polynomial and a commercial ILP solver[4] can be applied. Including the topological constraints (9), however, is not straightforward because their number can be exponential. Luckily, thanks to Lemma 1, we can find violated topological constraints in polynomial time and add these iteratively so as to solve the *full* problem to global optimality by a branch-and-cut approach [5]. In our experiments, a few hundred topological constraints are sufficient to solve the full ILP.

6 Experiments and Benchmark Results

Model Description. To apply the proposed model to the color images of the BSD, we first need to learn the likelihood functions from Section 4.2. To that end, we start from a watershed segmentation[5] of the training images (Fig. 1(a)), and use a simple tool that displays training images and curves between segments to annotate in total 8000 curves each as active or dormant. The curve likelihood $\hat{p}(y_e|f_e^{(1)})$ is then learned by a Random Forest, using these labels and a set of features of each curve and its adjacent segments (cf. supplementary material). For the junction likelihood $p(y_{e_1}, y_{e_2}, y_{e_3}|f_j^{(3)})$, the features $f_j^{(3)}$ consist of the angles between those curves e_1, e_2, e_3 that are incident to the junction j. Their distribution is learned by a Gaussian mixture model while respecting the sum constraint (all angles add up to 2π). The learned model is applied to the watershed segmentation of the BSD test images, using all bias settings from $\beta = 0.01$

---

[4]Here: IBM CPLEX 12.1

[5]Depending on the image, the number of curves between segments varies between 439 (minimum) and 10970 (maximum). The median over all images is 4276. It can be seen from Fig. 6, for high β, that the initial watershed segmentations are in fact over-segmentations with many excessive curves.

to $\beta = 0.99$ (cf. Fig. 6). For each image and each boundary bias β, we obtain one segmentation.

Evaluation on BSD. The BSD [15] is the standard benchmark for assessing these segmentations. It compares the closed contours of the estimated segments to those of human-made segmentations in terms of precision and recall (and F-Score). It has also been used to evaluate boundary detectors which need not produce closed contours. As is best practice [2], we compare segmentations to the BSD ground truth also in terms of the Variation of Information (VI) [16] and Rand index (RI) [21] which measure the discrepancy of partitions. Results are shown for two settings, one in which β is chosen optimally for the entire BSD test set, and one in which β is chosen optimally per image, see Table 1 and Fig. 5(b). It is apparent from this figure that closedness constraints are necessary. Perhaps surprisingly, the Gestalt terms seem to add little information in addition to the boundary terms.

At the time of writing, the quality of the partitioning as measured by the F-Score in the setting where the same (optimal) parameterization of algorithms is used for all images is on a par with [1] and second to no other algorithm that produces closed contours. But note that pure boundary detectors that need not produce closed contours [14, 22] still have a higher F-Score, that [2] achieves a higher RI and lower VI, and that [1] performs better in a different setting where parameters of the algorithms are optimized separately for each image.

Closedness constraints arguably alleviate the risk of under-segmentation: If two objects in an image which should be separated by a segmentation meet in n pairs of adjacent superpixels, n independent decisions to merge these pairs, with an average risk $p \in (0, 1)$ of false mergers, lead to an exponential average risk $1 - (1 - p)^n$ of falsely merging the objects. A remarkable achievement of e.g. [1, 2] is that p is low enough on the BSD to achieve top performance despite this risk. Closedness constraints enforce a consistent decision for all pairs and thus avoid the exponential risk.

Runtime. The runtime for the construction of the model, including the watershed segmentation, feature extraction and random forest prediction, is about 100 s in the median, cf. Fig. 7. The runtime spent on the optimization is small in the median (4 s) and no longer than 400 s for the most complex image of the BSD test set (Fig 7). It takes between one (for two out of 100 test images) and 18 cutting plane iterations (four in the median) until all constraints are satisfied and thus, the global optimum has been found. The warm-start mechanism of CPlex is used.

The runtimes we observe are interesting even for inter-active applications since most operations can be parallelized.
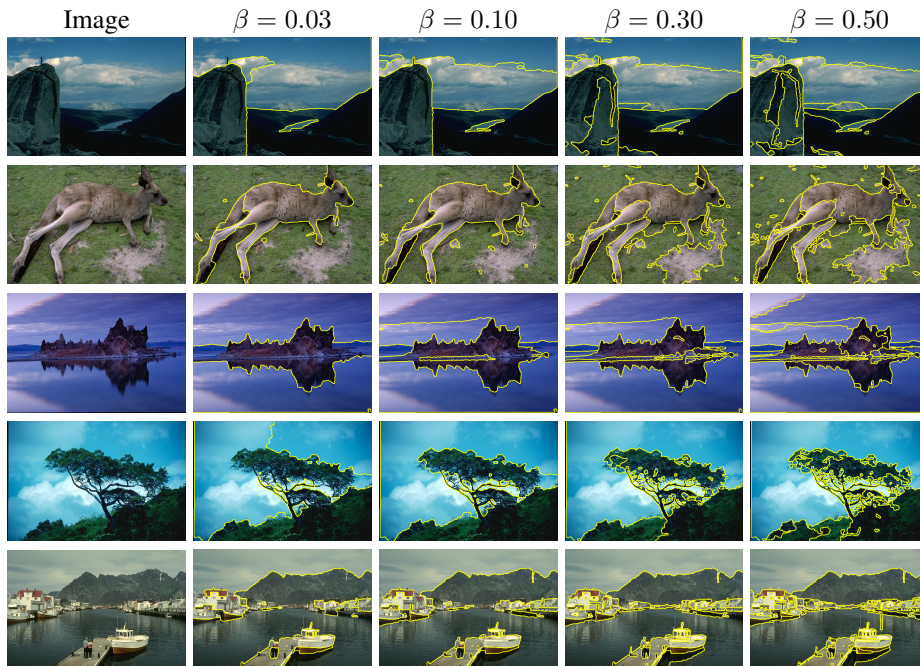
Figure 6: Different segmentations of the same image can be obtained by adjusting the prior probability $\beta \in (0, 1)$ of preserving curves between regions. The closed contours of all regions are depicted in yellow. The perceptually optimal β differs w.r.t. the image, and for the same image, several settings of β can correspond to meaningful but different segmentations.

		F (Prec, Rec) [15]	VI [16]	RI [21]
Parameters fixed	[1]	**0.67** (0.66, 0.69)	1.74	0.78
	[2]	**0.59**	1.65	0.81
	Ours	**0.67** (0.64, 0.74)	1.88	0.78
Parameters optimized per image	[1]	**0.71** (0.72, 0.72)	1.53	0.83
	[2]	**0.65**	1.47	0.85
	Ours	**0.70** (0.68, 0.76)	1.68	0.83

Table 1: Segmentation quality. A good algorithm has high F-measure, low index of variation (VI), and high Rand index (RI).
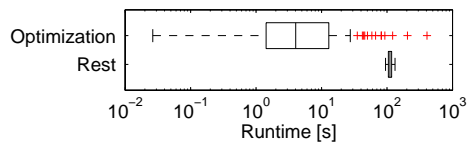

Figure 7: The absolute runtime for segmenting a BSD color image is about 100 s on average and about 500 s in the worst case.

7 Conclusion

We have proposed a new probabilistic graphical model for image segmentation. Introducing the topology of a cellular complex as a random variable has allowed us to exclude inconsistent edge labelings from the state space of this model. The likelihood of topology given an edge labeling is represented by a set of linear inequalities that enforce closed contours. In conjunction with appropriate conditional independence assumptions, the overall prior probability for merging adjacent regions is the only free parameter of this model. For any setting of this parameter, the edge labeling with maximum a posteriori (MAP) probability has closed contours. The MAP inference problem is a multicut problem and is amenable to a practicable solution, as we show, using a branch-and-cut algorithm. The quality of the resulting segmentation is comparable with the state of the art in closed-contour segmentation on the BSD benchmark. First experiments indicate that the model scales to 3D segmentation problems. We are currently working on interactive extensions where constraints can be added based on user input.

Appendix

Proof of Lemma 1

Proof. If an inequality in Definition 1 is violated, there exists an $n \in \mathbb{N}$, a closed path $(v_1, \ldots, v_n)$, and a $j \in \mathbb{N}$ such that all edges of the path except $\{v_j, v_{j+1}\}$ are labeled zero. Thus, there exists a path from v_j to v_{j+1} along which all edges are labeled zero, and $y_{\{v_j, v_{j+1}\}} = 1$. Hence, at least one inequality in Lemma 1 is violated.

Conversely, if an inequality in Lemma 1 is violated, there exists an $n \in \mathbb{N}$ and a path $(v_1, \ldots, v_n)$ along which all edges are labeled zero, and there exists the edge $\{v_1, v_n\}$ that is labeled one. Thus, the closed path $(v_1, \ldots, v_n, v_1)$ violates an inequality in Definition 1. $\square$

Conditional Independence Assumptions

$$\mathcal{F} \perp\!\!\!\perp \mathcal{T}$$

$$\forall c, c' \in C | c \neq c' : \quad \mathcal{Y}_c \perp\!\!\!\perp \mathcal{Y}_{c'}$$

$$\forall c, c' \in C | c \neq c' : \quad \mathcal{Y}_c \not\perp\!\!\!\perp \mathcal{Y}_{c'} | \mathcal{T}$$

$$\forall d, d' \in \{1, 3, 4\} | d \neq d' : \quad \mathcal{F}^{(d)} \perp\!\!\!\perp \mathcal{F}^{(d')} | \mathcal{Y}$$

$$\forall c, c' \in C | c \neq c' : \quad \mathcal{F}_c^{(1)} \perp\!\!\!\perp \mathcal{F}_{c'}^{(1)} | \mathcal{Y}$$

$$\forall c \in C : \quad \mathcal{F}_c^{(1)} \perp\!\!\!\perp \mathcal{Y}_{C \setminus \{c\}} | \mathcal{Y}_c$$

$$\forall d \in \{3, 4\} \forall j, j' \in J_d | j \neq j' : \quad \mathcal{F}_j^{(d)} \perp\!\!\!\perp \mathcal{F}_{j'}^{(d)} | \mathcal{Y}$$

$$\forall d \in \{3, 4\} \forall j \in J_d : \quad \mathcal{F}_j^{(d)} \perp\!\!\!\perp \mathcal{Y}_{C \setminus \mathcal{N}(j)} | \mathcal{Y}_{\mathcal{N}(j)}$$

References

[1] P. Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *CVPRW*, 2006.

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *TPAMI*, 33:898–916, 2011.

[3] F. Barahona, M. Grötschel, M. Jünger, and G. Reinelt. An application of combinatorial optimization to statistical physics and circuit layout design. *Oper. Res.*, 36:493–513, May 1988.

[4] C. R. Brice and C. L. Fennema. Scene analysis using regions. *Artificial Intelligence*, 1:205–226, 1970.

[5] A. Caprara and M. Fischetti. Branch and cut algorithms. In M. Dell'Amico, F. Maffioli, and S. Martello, editors, *Annotated Bibliographies in Combinatorial Optimization*, chapter 4. John Wiley, Chichester, 1997.

[6] S. Chopra and M. R. Rao. The partition problem. *Math. Program.*, 59:87–115, 1993.

[7] M.-C. Costa, L. Letocart, and F. Roupin. Minimal multicut and maximal integer multiflow: A survey. *European J. of Operational Research*, 162(1):55–69, April 2005.

[8] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *IJCV*, 2010.

[9] M. Deza, M. Grötschel, and M. Laurent. Complete descriptions of small multicut polytopes. In P. Gritzmann, Kabadi, and B. Sturmfels, editors, *DIMACS*, volume 4 of *Series in Discrete Mathematics and Theoretical Computer Science*, pages 221–252. AMS, 1991.

[10] M. Grötschel and Y. Wakabayashi. A cutting plane algorithm for a clustering problem. *Math. Program.*, 45:59–96, August 1989.

[11] J. H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schnörr. Globally optimal image partitioning by multicuts. In *EMMCVPR*, 2011.

[12] V. Kolmogorov and R. Zabih. What energy functions can be minimizedvia graph cuts? *TPAMI*, 26:147–159, 2004.

[13] V. A. Kovalevsky. Finite topology as applied to image analysis. *Computer Vision, Graphics, and Image Processing*, 46(2):141–161, 1989.

[14] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik. Using contours to detect and localize junctions in natural images. In *CVPR*, 2008.

[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001.

[16] M. Meila. Comparing clusterings by the variation of information. In B. Schölkopf and M. K. Warmuth, editors, *Learning Theory and Kernel Machines*, volume 2777 of *Lecture Notes in Computer Science*, pages 173–187. Springer Berlin / Heidelberg, 2003.

[17] E. N. Mortensen and J. Jia. Real-time semi-automatic segmentation using a bayesian network. In *CVPR*, 2006.

[18] S. Nowozin and C. Lampert. Global connectivity potentials for random field models. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 818 –825, june 2009.

[19] P. Orbanz and J. Buhmann. Nonparametric Bayesian image segmentation. *IJCV*, 77:25–45, 2008.

[20] T. Pavlidis. *Structural Pattern Recognition*, volume 1 of *Electrophysics*. Springer, 1977.

[21] W. M. Rand. Objective criteria for the evaluation of clustering methods. *J. of the American Statistical Association*, 66(336):846–850, 1971.

[22] X. Ren. Multi-scale improves boundary detection in natural images. In *ECCV*, 2008.

[23] X. Ren and J. Malik. Learning a classification model for segmentation. In *ICCV*, 2003.

[24] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007.

[25] J. Shi and J. Malik. Normalized cuts and image segmentation. *TPAMI*, 22:888–905, August 2000.

[26] D. Sontag and T. Jaakkola. New outer bounds on the marginal polytope. In *NIPS*, 2008.

[27] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[28] L. Zhang and Q. Ji. Image segmentation with a unified graphical model. *TPAMI*, 32:1406–1425, 2010.